Artificial Intelligence and Machine Learning

David C. Wyld,
Dhinaharan Nagamalai (Eds

Computer Science & Information Technology

- 11th International Conference on Advanced Computer Science and Information Technology (ICAIT 2022), July 23~24, 2022, Toronto, Canada
- 3rd International Conference on Cloud, Big Data and IoT (CBIoT 2022)
- 14th International Conference on Wireless & Mobile Network (WiMo 2022)
- 11th International Conference on Cryptography and Information Security(CRYPIS 2022)
- 11th International Conference on Digital Image Processing and Vision (ICDIPV 2022)
- 3rd International Conference on Artificial Intelligence and Machine Learning (CAIML 2022)
- 3rd International conference on Natural Language Computing Advances (NLCA 2022)

Published By

AIRCC Publishing Corporation

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Dhinaharan Nagamalai (Eds),
Wireilla Net Solutions, Australia
E-mail: dhinthia@yahoo.com

ISSN: 2231 - 5403
ISBN: 978-1-925953-71-8
DOI: 10.5121/csit.2022.121201 - 10.5121/csit.2022.121223

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

11th International Conference on Advanced Computer Science and Information Technology (ICAIT 2022), July 23~24, 2022, Toronto, Canada, 3rd International Conference on Cloud, Big Data and IoT (CBIoT 2022), 14th International Conference on Wireless & Mobile Network (WiMo 2022), 11th International Conference on Cryptography and Information Security (CRYPIS 2022), 11th International Conference on Digital Image Processing and Vision (ICDIPV 2022), 3rd International Conference on Artificial Intelligence and Machine Learning (CAIML 2022), 3rd International conference on Natural Language Computing Advances (NLCA 2022) was collocated with 3rd International Conference on Artificial Intelligence and Machine Learning (CAIML 2022). The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The ICAIT 2022, CBIoT 2022, WiMo 2022, CRYPIS 2022, ICDIPV 2022, CAIML 2022 and NLCA 2022. Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically.

In closing, ICAIT 2022, CBIoT 2022, WiMo 2022, CRYPIS 2022, ICDIPV 2022, CAIML 2022 and NLCA 2022 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the ICAIT 2022, CBIoT 2022, WiMo 2022, CRYPIS 2022, ICDIPV 2022, CAIML 2022 and NLCA 2022

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld,
Dhinaharan Nagamalai (Eds)

General Chair

Organization

David C. Wyld, Southeastern Louisiana University, USA
Dhinaharan Nagamalai (Eds) Wireilla Net Solutions, Australia

Program Committee Members

Abdalhossein Rezai, University of Science and Culture, Iran
Abdel-Badeeh M. Salem, Ain Shams University, Egypt
Abdelhadi Assir, Hassan 1st University, Morocco
Abdelhak Merizig, Mohamed Khider University of Biskra, Algeria
Abdellatif I. Moustafa, Umm AL-Qura University, Saudi Arabia
Abdelouafi Ikidid, Cadi Ayyad University, Morocco
Abderrahim Siam, University of Khenchela, Algeria
Abderrahmane Ez-Zahout, Mohammed V University, Morocco
Abdessamad Belangour, University Hassan II Casablanca, Morocco
Abdul Rehman, Kyungpook National University, South Korea
Abdus Satter, University of Dhaka, Bangladesh
Abhay Kumar Agarwal, Kamla Nehru Institute of Technology, India
Abhilash, Cyrielle Castle, India
Addisson Salazar, Universitat Politècnica de València, Spain
Adil Bouhouch, Mohammed V University, Morocco
Ahmad A. Saifan, Yarmouk university, Jordan
Ahmed Bentajer, ENSA of Tetouan, Morocco
Ahmed Farouk AbdelGawad, Zagazig University, Egypt
Ahmed Haitham Najim, Sfax Tunisia-Imam Al-Adham University, Iraq
Ajay Anil Gurjar, Sipna College of Engineering and Technology, India
Ajmer Singh, DCRUST, India
Akhil Gupta, Lovely Professional University, India
Ali A. Shukur, Belarusian State University, Belarus
Ali Abdrhman Mohammed Ukasha, Sebha University, Libya
Ali Hussein Wheeb, University of Baghdad, Iraq
Aliasghar Tarkhan, University of Washington, USA
Alireza Valipour Baboli, University Technical and Vocational, Iran
Altay Guvenir, Bilkent University, Turkey
Amal Azeroual, Mohammed V University, Morocco
Amando P. Singun Jr, University of Technology and Applied Sciences, Oman
Amar Ramdane Cherif, University Paris Saclay, France
Amit Agarwal, Wells Fargo, India
Amit Kumar, Cadence Design Systems, USA
Anastasios Doulamis, National Technical University of Athens, Greece
Angelina Tzacheva, University of North Carolina, USA
Anita Dixit, SDM College of Engineering and Technology, India
Anita Yadav, Harcourt Butler Technical University, India
Anjan Krishnamurthy, BMS Institute of Technology and Management, India
Anouar Abtoy, Abdelmalek Essaadi University, Morocco
Antinisca Di Marco, University of L'Aquila, Italy
Aref Wazwaz, Dhofar University, Oman
Aridj Mohamed, Hassiba Benbouali University Chlef, Algeria
Arti Jain, Jaypee Institute of Information Technology (JIIT), India
Asif Khan, Integral University, India

Mohammed Bouhorma,	Abdelmalek Essaadi University, Morocco
Mohd Soperi Mohd Zahid,	Universiti Teknologi PETRONAS, Malaysia
Monji Zaidi,	King Khalid,university, KSA
Morris Riedel,	University of Iceland, Iceland
Mostafa S. Shadloo,	INSA Rouen Normandie, France
Moulay Youssef Ichahane,	Chouaib Doukkali University, Morocco
Mounir Zrigui,	University of Monastir, Tunisia
Mridula Prakash,	L&T Technology Services (LTTS), India
Muhammad Shafiq,	Guangzhou University, China
Munshi Md Shafwat Yazdan,	Idaho State University, USA
Mu-Song Chen,	Da-Yeh University, Taiwan
Mussa Turdalyuly,	Satbayev University, Kazakhstan
Mu-Yen Chen,	National Cheng Kung University, Taiwan
MV Ramana Murthy,	Osmania University, India
Nadia Abd-Alsabour,	Cairo University, Egypt
Nadine Akkari,	Lebanese University, Lebanon
Nameer N. El-Emam,	Philadelphia University, Jordan
Ngoc Hong Tran,	Vietnamese-German University, Vietnam
Nicolas Durand,	Aix-Marseille University, France
Nikola Ivkovic,	University of Zagreb, Croatia
Nikola Ivković,	University of Zagreb, Croatia
Nikolai Prokopyev,	Kazan Federal University, Russia
Nishant Doshi,	Pandit Deendayal Energy University, India
Noor Mowafeq Allayla,	University of Mosul, Iraq
Nur Eiliyah Wong,	Senior Lecturer/ Researcher, Malaysia
Oday Ali Hassen,	University Technical Malaysia Melaka, Malaysia
Okba Kazar,	University of Biskra, Algeria
Okwonu,	Universiti Utara Malaysia, Malaysia
Olufade Onifade,	University of Ibadan, Nigeria
Omar Khadir,	Hassan II University of Casablanca, Morocco
Omid Mahdi Ebadati,	Kharazmi University, Tehran
Osman Toker,	Yildiz Technical University, Turkey
Parameshachari B D,	GSSSIETW, India
Pascal Lorenz,	University of Haute Alsace, France
Patrick Fiati,	Cape Coast Technical University, Ghana
Peiying Zhang,	China University of Petroleum, China
Ping Zhang,	Anhui Polytechnic University, China
Pranita Mahajan,	Sies graduate school of technology, India
Prasan Kumar Sahoo,	Chang Gung University, Taiwan
Prasang Gupta,	Emerging Technologies, PwC US, India
Prem Kumar Singh,	Gandhi Institute of Technology and Management, India
Prudhvi Parne,	Bank of Hope and University of Louisiana, USA
Przemyslaw Falkowski-Gilski,	Gdansk University of Technology, Poland
Qi Zhang,	Shandong University, China
Quang Hưng Do,	University of Transport Technology, Vietnam
Rachid Zagrouba,	Imam Abdulrahman Bin Faisal University, Saudi Arabia
Radha Raman Chandan,	Banaras Hindu University, India
Radu Vasiu,	Politehnica University of Timisoara, Romania
Rahul Mulajkar,	Jaihind COE,Pune, India
Ramadan Elaiess,	University of Benghazi, Libya
Ramgopal Kashyap,	Amity University Chhattisgarh, India
Rami Raba,	Al- Azhar University, Palestine

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Artificial Intelligence Community (AIC)

Soft Computing Community (SCC)

Digital Signal & Image Processing Community (DSIPC)

11th International Conference on Advanced Computer Science and Information Technology (ICAIT 2022)

3rd International Conference on Cloud, Big Data and IoT (CBIoT 2022)

14th International Conference on Wireless & Mobile Network (WiMo 2022)

11th International Conference on Cryptography and Information Security (CRYPIS 2022)

11th International Conference on Digital Image Processing and Vision (ICDIPV 2022)

3rd International Conference on Artificial Intelligence and Machine Learning (CAIML 2022)

3rd International conference on Natural Language Computing Advances (NLCA 2022)

PERFORMANCE AND EFFICIENCY ASSESSMENT OF DRONE IN SEARCH AND RESCUE OPERATION

Tauheed Khan Mohd, Vuong Nguyen, Trang Hoang, P. M. Zeyede, and Beamlak Abdisa

Augustana College

Rock Island, Illinois, 61201, USA

Abstract. With the development of technology, human beings have successfully predicted and prevented the damage caused by natural disasters. However, due to climate change, society has witnessed the rising actions of forest fire, earthquake, tsunami, etc., and there are many which people cannot prevent, and the level of dangerous situations are increasing rapidly for the Search and Rescue (SAR) operation. Not to mention, more and more people are turning their attention and hobbies to exploring wilderness where they might get lost or, worse, get injured. For that reason, to raise the chance of survival for the victims and reduce the risk for the search team, the use of Unmanned aerial vehicles (UAV) has been proposed. The plan is the headquarters will deploy a fleet of drones to get into the areas where human cannot enter easily and then report the situations as well as the condition of the victims with images and videos. In most research papers, it seems very promising; however, there is still much work that needs to be done. In this paper, some of the features which included for future researches are which algorithm is the most optimal, what standard structure should be used for the drones so it can complete the missions under any kind of circumstances, and how to set up a communication line that guaranty the effective to reduce the level of miscommunication.

Keywords: UAV, AI, drone, search, rescue, efficiency

1 Introduction

The term search and rescue (SAR) has been a challenging activity, especially recently when people are witnessing many natural disasters such as volcanic eruptions, earthquakes, and hurricanes. In addition, accidents are sometimes caused by humans' carelessness and irresponsibility. In search and rescue scenarios, time is often the most critical factor as the life of victims are at risk [1]. For that reason, it is crucial that operators need to work on the SAR operation to find the most optimal solution. The idea is to use drones with artificial intelligence (AI) to help with the searching procedure. With a fleet of drones, people can scan a considerable area in

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 01-16, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121201

a short amount of time, reducing the amount of cost and generating a safer working environment. The plan works in most articles; however, the proposal has encountered some issues. Inspectors need to figure out the best way to do this within the battery life of the drone fleet for them to have the best performance. Another barrier is communication between the fleet and the control command center since the rescue area can be significantly distant [2]. The purpose of this paper is to gather information about the effective use of drone fleets and modify some of its features to maximize its productivity in search and rescue operations. The paper has four main sections: Introduction, where it talks about the benefit of the drones for the search and search operation, and then the related work section, which describes the works in the past as well as what kind of data had been collected. After a review of what has been done by the previous researchers, this paper will propose some of the ideas for future work in order to suppose this proposal. Last but not least, it is the conclusion.

1.1 The use of drones for search and rescue operation

The use of drones has dramatically exceeded the expectations set for them. Drones have proved now and again that they can perform tasks with much greater efficiency and speed rather than ourselves. Furthermore, it has been an enormous relief that executives do not have to put humans in harm's way to complete taxing tasks. The development of drones is still in its infancy, so engineers still have more to gain from working on them. Hence, their use in search and rescue missions will make human missions obsolete, and doing so will significantly decrease collateral damage and save the lives of both victims and rescuers. One of the prominent reasons drones are better suited for SAR missions is that they can work 24/7 non-stop on any terrain or environment. [3] The versatile designs that exist as well as the designs that operators can yield from another development guarantee the manufacture of powerful drones that can decrease the number of lives lost in these missions in an immense amount, not to mention the fact that the consistent upgrades in their analogs, drones are becoming easier to use by the day. Moreover, various innovations are promoting the growth of the drone industry, a plethora of innovations are frequently being released to meet humankind's needs; Bluetooth low energy technology is one example. [4] Additionally, for every search and rescue operation, the critical factor is time. The rescuer needs to be responsive to the victim's area as quickly as possible to lower the level of danger and give necessary assistance if needed. Moreover, because of that, drones will be an optimal solution as it is easy for them to get into the areas where it might take days for on foot rescuer to get in and report the situation to the headquarter with images.

1.2 The use of AI in deploying drone for search and rescue operation

The discussion of using drones for search and rescue operations is getting more attention due to the development of technology and the efficiency of drones. The drone has significant benefits in terms of time as well as resources. Therefore, engineers must minimize or eliminate traditional search and rescue operations and focus on developing new methods. The use of drones in operation is not about sending a fleet to the located victim's position, but it is more about communication. Each drone needs to collect and exchange information so that the search and rescue operation can reach its best performance. The drones should follow a specific algorithm so that each of them will have a different mission as well as be able to calculate the fastest road to the victim's locations. For example, it will be unwise for two drones to appear in the exact location and have the same task. Each of them is a piece of a puzzle, and together with the right algorithm, they can gather enough information to finish the mission, which is to find the victims. In addition, it will not be economically efficient if supervisors need a whole team to operate the drone fleet. Operators do not need a group of people to decide where each drone should go while questioning if an area is covered yet. For such reasons and to decrease human errors, AI is crucial for drones. With artificial intelligence, every drone will have its position and function. Therefore, controllers will not have to worry about data duplication or miscommunication. Administrators can minimize the human resources since they might only need one to two people to monitor the process and transmit the necessary information to the ground teams.

2 Related work

There have been many works related to this topic. Most of the works focus on specific natural areas with a wide range of areas such as sea or forest. Furthermore, it is effortless to understand since humankind has been witnessing a considerable amount of missing planes recently. The works below are some examples in which drones' application is used in search and rescue operations.

2.1 Experiment

Now let's take a look and see in what kind of situations UAV brings the best result. Doing routine testing and lucid experiments are essential if people wish to have a better understanding of what needs to be developed more and what should be maintained [5]. Typically, the area where operators need to use drones is enormous. In the article "Autonomous Drones for Assisting Rescue Services within the context of Natural Disasters," it is reported that authorities are using UAVs to rescue the people who suffer from natural disasters. The plan is to use drones to get places where humans cannot get access, and then they will scan and send the image

back to the base. From that, searchers can make a 3D image like in figure 1 and figure 2 and locate where the victims are, and make an optimal plan to rescue. Also, in this article, the authors have pointed out three specific sub-tasks for the UAV to perform: Detect people, evaluate the group's composition, and estimate the direction position of the group and its velocity. [6] [7]

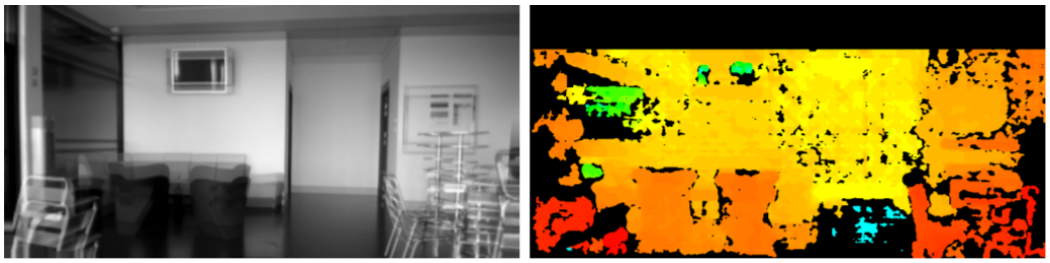


Fig. 1. Dense3D reconstruction The overlaid rectified images before and after the height change visualize the precision of the estimated camera motion (left). Therefore, any standard implementation for distance reconstruction, e. g. [12], may be used without modification (right).

Fig. 2. Sparse3D indoor (left figure) and outdoor (right figure) reconstruction based on a few dozens of points. Blue/purple lines show optical flow vectors consistent/ conflicting with the camera's motion. The points in color represent their longitudinal distance – red indicates 1 meter or less, cyan for 10 meters or more. A larger green circle marks the flight direction targeted by the drone, which is computed according to the furthest possible distance with no potential obstacles

Moreover, in the article "Intelligent Drone Swarm for Search and Rescue Operation at Sea," the use of ships and helicopters has wasted a tremendous amount of time and money; therefore, they have proposed a solution which is using the drone. The plan is to launch a fleet of UAVs to search for different ranges of area. This idea will increase the chance of survival for the victims since operators will only need to use the same amount of time (maybe less), but they can look for a wider area. Studies propose optimization techniques for SAR operations which present

a search heuristic that minimizes the time to find a single stationary entity on a given area. [8] [9] [10]

In addition, another article [11] has proposed the plan to search the missing people in large swaths of underpopulated wilderness. This is brought up because the number of people hiking to explore nature has increased rapidly. The idea is still using drones to search for missing people, but this time, they will attach a heat camera so it will navigate the lost people by their heat signature as shown in Fig 3.

Fig. 3. A drone enabling remote collaboration between an outdoor user and a remote (indoor) user. Such a system could be used to support remote collaboration in wilderness SAR.

As mentioned above, it is a huge challenge when people get lost or injured in a wilderness as the terrains can make it extremely difficult for the Search and Rescue operator to get to their position on time. Therefore, to minimize the time and the risks, Yunus Karaca and his team have proposed an idea where they use the drone to search and reach the victims [12]. They conduct an experience in a mountain environment using a simulation model (as shown in Figure 4). They focus on comparing the amount of time for two operations which are the Classical Line Search Technique (CLT) and the Drone-snowmobile Technique (DST), in three categories with $p < 0.001$ for a,b and c: First human contact, Total Searched Area and Searched Area for a minute. The CLT was more about human resources since

the search was by foot to reach the victims. Meanwhile, for the DST, the search was performed by drone as shown in Fig 5, and the victim reached by snowmobile.

Fig. 4. The simulation where an unconscious victim in snow-covered ground was enacted 10 times for each group using a 180 cm shop window mannequin to represent the accident victim [12].

The term "search and rescue operation" refers to situations with life and death, and the working conditions are hazardous. However, at this moment, communities are suffering from COVID-19 and are witnessing millions of people die because of it. Even though this situation is not expected for search and rescue operation, it is a life-threatening circumstance. Researchers have found some of the practical applications of AI and drones to identify people with COVID-19 symptoms and immediately put them in quarantine if necessary. Scientists are on the path to finding the cure; however, before they can do that, it is essential to minimize the consequence by keeping distance and wearing masks. These are the most effective ways, but it is not enough. A plan using AI and drones has been suggested [13]. First, drones use AI to detect the people suspended from COVID-19; then, AI will judge based on their X-RAY scan of their lungs, their blood test, the place they have traveled, and the temperature. Figure 6 below are the proposed framework

Fig. 5. The model of drone used in this experiment was DJI Phantom 3 Pro

in the article [4] by Mr. Maghded and his team. The first layer is responsible for FPR reading data from the sensor. It will read the captured CT scan images of the lung through using the smartphone camera, getting the inertial sensors measurements during 30-second sit-to-stand, as well as recording cough voice samples through microphone voice measurements during a series of cough, and finally, it will scan for the temperature sensor measurements during fingerprint touching on the smartphone screen. The reading of these samples is based on the symptoms from the COVID-19 patients. The second layer is structured to construct the on board smartphone sensors, including reading intervals, image size, timer resolution, etc. The third layer provides the calculated symptoms level, separately, and is then stored as a record input to the next layer. The last layer is to apply ML techniques to predict the COVID-19. The ML techniques could be used according to the nature of the recorded data. In order to increase the reliability of the result, the recorded data and the result will exchange in the cloud. If executives are managed to get more people using this framework, then the data-set will be larger, and the result will be more trustful.

Another way to prevent the spread of COVID-19 is to decontaminate the area where people with Corona viruses are permanent residents and their visited location. Many countries are using the chemical for the purpose of sanitizing, but there

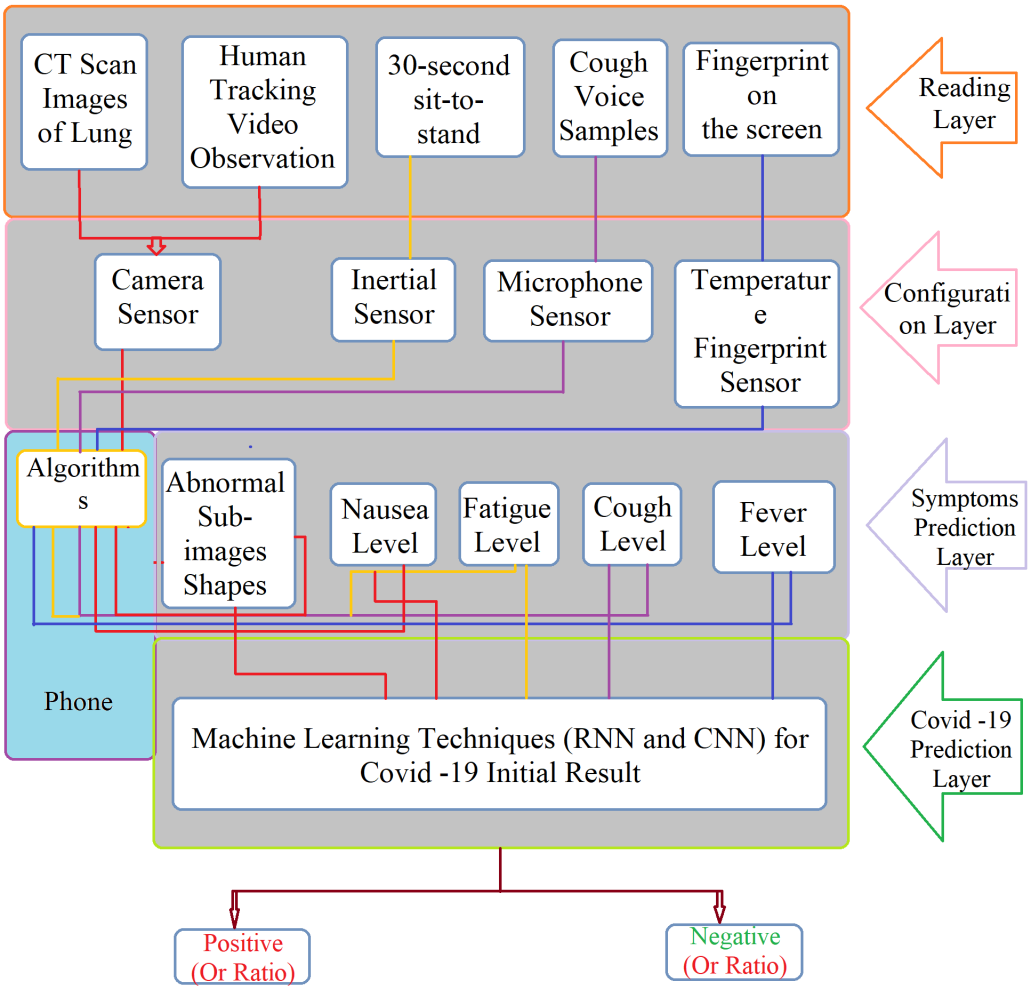


Fig. 6. Framework for AI-based Screening method

is a problem with using human resources since it takes too much time. Therefore, the idea is to use the drone to spray each place, which will cost half of the time. In the article [14], the author has shown that for a big country like India with a mass area of 3.28 million sq.km, it will take more than one hour for manual spraying compared to 15 minutes using drone (as seen in Figure 7).

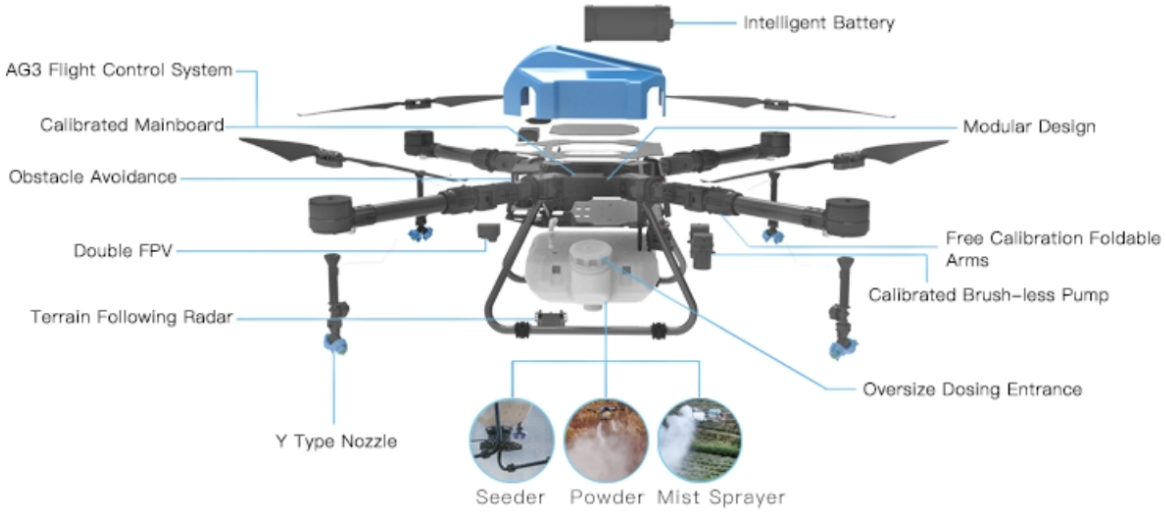


Fig. 7. Corona Killer Drones Disinfecting in India

Above are some examples which research made it very clear to understand and very persuasive about using the drone and AI to serve search and rescue purposes. There are more, but the most common thing is that they use drones and AI to minimize the time of the operation and increase collaboration and coordination while performing SAR tasks. [15]

2.2 Results of the experiments

- It is noticeable that the chance to be found for the people who are in danger using drones is much higher since the drone will be able to cover a much larger area, compared to the traditional way as in table 1 and table 2.
- Unmanned aerial vehicles will ensure the coverage of a large area in less time, reducing the impact of the limited battery capacity. As one can see in table 1 and table 2, this is the data which was found in the article [12] which demonstrate the effectiveness of using the drone for search and rescue operation. It is easy to notice a vast difference between traditional way versus drone search in three

categories. As a result, it takes less time for drones to search and find the missing person than traditional solutions.

Table 1. Characteristics of classic search in the mountain rescue scenarios

Operation No.	CLT		
	First Human Contact (min)	Total Searched Area (m^2)	Searched Area for a minute (m^2/min)
1	39	66408	1702.8
2	53.1	78209	1475.6
3	67.1	88664	1323.3
4	95	120891	1272.5
5	50.2	85861	1717.2
6	95.2	104479	1099.8
7	54	98385	1821.9
8	61.1	77378	1268.5
9	59.1	87980	1503.7
10	56.1	99375	1774.6
Median (25-75%)	57.3^a (52.3 – 74.0)	88322.0^b (78001.3 – 100651.0)	1489.7^c (1271.5 – 1731.6)

– Drones will give operators a better look at the scenario. Instead of just seeing trees and the outside context, controllers will be able to have a closer look and know what is happening inside. From that, operators can make a better judgment about the situation and take effective action [16].

– Using Drone will reduce the possibility of a rescuer getting injured or dying while on duty. Due to the lack of adequate information about the victim's situation, rescuers might cause more harm to the victim and themselves [1]. During the operation, the rescuer will have to remove barriers; however, they are not fully aware of what might happen if they clear the way. Therefore, it is possible for them to injure themselves and the victims. For that particular reason, it will be safer to use drones to locate along with estimating the most optimal pathway to reach the victims. Drones can provide a clear outlook for the rescuers so that they can perform with an adequate grasp of the victim's situation while maintaining their safety.

Table 2. Characteristics of drone search in the mountain rescue scenarios

Operation No.	DST		
	First Human Contact (min)	Total Searched Area (m^2)	Searched Area for a minute (m^2/min)
1	7.7	168395	28065.8
2	8.2	217624	33225.1
3	8.5	239602	35080.8
4	11.2	310981	32734.8
5	5.6	192224	49162.1
6	13.1	346268	30294.7
7	7.4	144480	25302.9
8	4.2	138945	54488.2
9	9.7	266722	33340.3
10	12.9	313525	27968.3
Median (25-75%)	8.4^a (6.9 − 11.6)	228613.0^b (162416.0 − 311617.0)	32979.9^c (28041.4 − 8601.1)

2.3 Limitations

- Battery Life: These drones can only work in such an amount of time before they need to go back to base and charge again. It is noticeable the biggest weakness of drones as it might affect the effectiveness of the task.
- Hardware and software issues have to be addressed: which algorithmic architectures to adopt? This question contains the answer to many problems. It is imperative to decide which algorithm in drones will be used since with the suitable algorithm, operators can find the best path (which mean with the shortest flying time but still able to cover the entire area) for the drone, and they might even solve the battery issues Which embedded system configuration is the most suitable one? The drones built, which are currently common use, are easily attacked by outside factors, and there are no protection layers for essential parts such as the wings.
- Camera and sensor: what kind of camera has the best weight and is able to give users the best visual of the situation no matter what the scenario is [17]. This is very important as operators need to remember that the place where they send the drone fleet to scout is very dangerous. There are some factors that can interfere with the view, such as heat, smoke or fog, etc. which will be challenging for them to detect the location of the victim if operators want to

locate them using heat signature and terrain that will become quite challenging for the controller to go in and collect information. The current use cameras are working correctly but only in a transparent environment. If the view is blocked, then the drone fleet is useless [18] [19].

- Communication: The communication capacity due to the limit of LoS (Limited Line of Sight) from the base on the ground to UAV, engineers need to consider that the wireless communication will not be strong enough to support the bandwidth requirements after few kilometers [20].
- Last but not least, how can a drone help to appease people in critical conditions or to provide helpful information which might help them stay alive? This is also an important task because the victims need to keep calm to make good calls, but can they do that when the only help they can get is from a drone? The answer to this is still unclear, but it is straightforward that the victims will be comforted if they see an actual human being. Therefore, making a drone able to appease people and give them helpful guidance so that they can survive is extremely necessary.

3 Proposal for future works

This proposal is promising as it will reduce the time and the risks for both the victims and the rescuer. However, there are still many works that need to be done.

3.1 AI Algorithm for the drone fleet

What algorithm should engineers use? It is easy to notice that there has been many articles and experiments about AI algorithm for search and rescue operation, and each of them has their strength as well as weakness. Based on the research from the existing articles, this paper found out that their algorithm might only work best in their stimulation situation since that algorithm is created to serve their need. People need to do more research so analysts can find out which algorithm will fit all the scenarios. This is challenging work, but researchers believe that it is necessary. It will affect the effectiveness of the mission if operators have to consider which algorithm should be used every time they encounter it. When the situation is life-threatening, time is critical. Because of that, executives need an answer or algorithm which will immediately respond no matter what the environments are. Moreover, the algorithm should also assess the drone's condition as when a drone is having technical problems such as low battery or malfunction of the camera and it needs to come up with a solution. The idea of deploying drones for search and rescue operations will become more practical if scientists manage to come up with an optimal algorithm.

3.2 Equipment and architecture

- Firstly, what kind of battery will provide the best lifespan? It is economically unwise to replace the drones instead of finding the appropriate battery that can enable the drone to fully function. In two or three hours drones run out of battery and they need to be charged. It is hard to predict the time for an operation to end. Every operation needs eyes on the sky as long as possible to observe activities and changes that might endanger the victim or the rescuer.
- Secondly, it is challenging to decide which material to use for the drones. Operators encounter several unpredictable variables during the operation. The drone's building material should be light enough so that the weight may not affect the drone, and it should be strong enough to enable the drones to complete their missions. In addition, the material should be water-proof so that water may not affect some of its functionalities.
- Thirdly, what should be the standard structure for the drones? There have been many models for drones released each year. Even though some of them are made for search and rescue purposes, they are not optimal since the operator has to consider the situation before those drones can be deployed. The idea is to have the drone model ready for any circumstance. In the article by S. P. Yeong [21], the author has pointed out some of the advantages and disadvantages for a different type of drone built-in table 3. Each drone has its owns unique pros and cons, and in order for UAVs to be ready for search and rescue operation, the researchers need to come up with a model that can incorporate all the necessary strengths and limit the weakness so that the drones will reach their maximum potential.
- Fourthly, what kind of communication device is suitable to apply? If the communication line is not stable, the operators will only receive piece of information, making it very confusing while endangering the whole operation. Therefore, it is essential to insert the right equipment to ensure the communication line may not get interrupted.
- Last but not least, it is also necessary to pay attention to the quality of the images and videos. The primary purpose of using drones is to collect information about the victim's situation by searching a large area and locating the missing person in a short amount of time. For this reason, everyone can not ignore the fact that drones must send images and videos of a high quality so that the headquarter fully understand what they are dealing with. Furthermore, researchers raise the following question regarding image and video quality: what kind of camera should be used? The camera for this procedure needs to meet these criteria: always assure the images and videos are clear to be seen and understood and, working properly under any situation, and be able to withstand certain hits from outside impact if necessary.

Table 3. Advantages and Disadvantages of different type of drone build

Drone Type	Advantages	Disadvantages
Fixed-wing	• Long range • Endurance	• Require an amount of space for horizontal take-off • Less maneuverability compared to VTOL (Vertical Take-Off and Landing)
Tilt-wing	• Has the advantages of both fixed-wing and VTOL	• Expensive • Complicated technology
Unmanned Helicopter	• VTOL • Maneuverability • High payloads possible	• Expensive • Require high level of maintenance
Multi-copter	• Expansive • Not difficult to launch • Light weight	• Limited payloads • Easy to be influence by wind

4 Conclusions

In conclusion, the proposal of using drones to serve search and rescue missions is very promising. There are endless possibilities with what drones can achieve. Still, there are many things that need to be done as the incorporation of drones with artificial intelligence is currently not widely used for search and rescue operations. This paper has pointed out some of the limitations and works that are needed to be done in the future based on other research papers and their results. Their works and this study intend to shorten the gap between the problem and solution. Since this concept is still a proposal, it is facing many difficulties in order to make it more reliable and make sure it works properly. In the future, in order to avoid doing the same experiments and collecting the same database, the research papers need to approach in different ways. From the articles for this paper and the result conducted from related works, there exist some similarities, such as the simulations that they are using are too specific, which might not be reusable for other contexts. Furthermore, authors are paying lots of attention in comparing the time between the traditional and drone methods while little do they remember that they also need to develop other aspects such as how should the drone notice the characteristic of humans or how should they respond to the headquarters when they found the victims? This is a promising idea; however, we should not rely on the development

of modern technology; future researchers need to reinforce the formal data as well as make it reliable and at the same time, consider about every other aspect of this idea as mention in the section above so that the plan of using the drone in search and rescue operation will be executed in a short time.

References

1. S. Mayer, L. Lischke, and P. W. Woźniak, "Drones for search and rescue," in *1st International Workshop on Human-Drone Interaction*, 2019.
2. V. B. Hammerseth, "Autonomous unmanned aerial vehicle in search and rescue," Master's thesis, Institutt for teknisk kybernetikk, 2013.
3. H. Hildmann, "Using unmanned aerial vehicles (uavs) as mobile sensing platforms (msps) for disaster response, civil security and public safety," 2019.
4. A. Hashmi, "A novel drone-based search and rescue system using bluetooth low energy technology," *Engineering, Technology amp; Applied Science Research*, vol. 11, p. 7018–7022, Apr. 2021.
5. D. Erdos, A. Erdos, and S. E. Watkins, "An experimental uav system for search and rescue challenge," *IEEE Aerospace and Electronic Systems Magazine*, vol. 28, pp. 32–37, 2013.
6. L. Apvrille, T. Tanzi, and J.-L. Dugelay, "Autonomous drones for assisting rescue services within the context of natural disasters," in *2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*, pp. 1–4, IEEE, 2014.
7. P. Rudol and P. Doherty, "Human body detection and geolocalization for uav search and rescue missions using color and thermal imagery," in *2008 IEEE aerospace conference*, pp. 1–8, Ieee, 2008.
8. V. Lomonaco, A. Trotta, M. Ziosi, J. D. D. Y. Avila, and N. Díaz-Rodríguez, "Intelligent drone swarm for search and rescue operations at sea," *arXiv preprint arXiv:1811.05291*, 2018.
9. S. Schismenos, M. Chalaris, D. Emmanouloudis, and N. Katopodes, "Renewable energy and drones in search and rescue: Automated network for air-sea actions," in *Conference Proceedings, SafeKozani*, pp. 358–365, 2018.
10. J. V. Tomotani, "using unmanned aerial vehicles in search and rescue operation," 2015.
11. J. Brennan, A. Tang, and C. Neustaedter, "Drones for remote collaboration in wilderness search and rescue," 2019.
12. Y. Karaca, M. Cicek, O. Tatli, A. Sahin, S. Pasli, M. F. Beser, and S. Turedi, "The potential use of unmanned aircraft systems (drones) in mountain search and rescue operations," *The American journal of emergency medicine*, vol. 36, no. 4, pp. 583–588, 2018.
13. H. S. Maghded, K. Z. Ghafoor, A. S. Sadiq, K. Curran, D. B. Rawat, and K. Rabie, "A novel ai-enabled framework to diagnose coronavirus covid-19 using smartphone embedded sensors: design study," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 180–187, IEEE, 2020.
14. P. Vaishnavi, J. Agnishwar, K. Padmanathan, S. Umashankar, T. Preethika, S. Annapoorani, and M. Subash, "Artificial intelligence and drones to combat covid-19," *Preprints. org, no*, 2020.
15. M. H. Dominguez, S. Nesmachnow, and J.-I. Hernández-Vega, "Planning a drone fleet using artificial intelligence for search and rescue missions," in *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1–4, IEEE, 2017.
16. S. Karma, E. Zorba, G. Pallis, G. Statheropoulos, I. Balta, K. Mikedi, J. Vamvakari, A. Pappa, M. Chalaris, G. Xanthopoulos, *et al.*, "Use of unmanned vehicles in search and rescue operations in forest fires: Advantages and limitations observed in a field trial," *International journal of disaster risk reduction*, vol. 13, pp. 307–312, 2015.

17. C. B. et al, "Requirments and limitations of thermal drones for effective search and rescue in marine and coastal areas," 2019.
18. A. A. Zhilenkov and I. R. Epifantsev, "Problems of a trajectory planning in autonomous navigation systems based on technical vision and ai," in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pp. 1032–1035, IEEE, 2018.
19. A. W. Brown, B. E. Franklin, and J. G. Estabrook, "Sensor processing and path planning framework for a search and rescue uav network," 2012.
20. C. A. Baker, S. Ramchurn, W. L. Teacy, and N. R. Jennings, "Planning search and rescue missions for uav teams," in *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, pp. 1777–1778, 2016.
21. S. Yeong, L. King, and S. Dol, "A review on marine search and rescue operations using unmanned aerial vehicles," *International Journal of Marine and Environmental Sciences*, vol. 9, no. 2, pp. 396–399, 2015.

Authors

Tauheed Khan Mohd Tauheed Khan Mohd received his B.Tech in Computer Engineering from Jamia Millia Islamia, New Delhi, India in 2006. He received his M.S degree from The University of Toledo, Ohio in 2015, and finished his Ph.D. in Human-Computer Interaction (HCI) during Summer 2019 at The University of Toledo.

Previously, Tauheed worked for three years as Software Engineer in HCL Technologies, India followed by four years at a French Multinational company SOPRA. He worked onsite for three months at AIRBUS in Toulouse, France, and managed their onboard application called Network Server System (NSS). Tauheed worked as a Research Assistant on an NSF Funded Project called INITIATE which enables High School Students to get attracted towards STEM subjects.

His areas of research are Human-Computer Interaction, Multimodal Input, Autonomous Vehicles, Micro-controller devices including Arduino and Raspberry Pi.

Vuong Nguyen CSC Major at Augustana class of 2022, his interest are Machine Learning, AI, Unmanned aerial vehicles, etc.

Trang Hoang CSC Major at Augustana class of 2022, her interest are Web Technologies, AI, Unmanned aerial vehicles.

P. M. Zeyede CSC Major at Augustana class of 2025, his interest are Python, AI, Unmanned aerial vehicles, Computer Networks.

Beamlak Abdisa CSC Major at Augustana class of 2025, her interest are Machine Learning, Autonomous Vehicles, Unmanned aerial vehicles, Web Technologies.

EVALUATION OF MACHINES LEARNING ALGORITHMS IN DETECTION OF MALWARE-BASED PHISHING ATTACKS FOR SECURING E-MAIL COMMUNICATION

Kambey L. Kisambu[1] and Dr. Mohamedi Mjahidi[2]

[1]Msc, Cyber Security, Department of Computer Science, University of Dodoma
[2]Lecturer, Department of Computer Science, University of Dodoma, Tanzania

ABSTRACT

Malicious software, commonly known as Malware is one of the most significant threats facing Internet users today. Malware-based phishing attacks are among the major threats to Internet users that are difficult to defend against because they do not appear to be malicious in nature. There were several initiatives in combating phishing attacks but there are many difficulties and obstacles encountered. This study deals with evaluation of machine learning algorithms in detection of malware-based phishing attacks for securing email communication. It deeply evaluate the efficacy of the algorithms when integrated with major open-source security mail filters with different mitigation techniques. The main classifiers used such as SVM, KNN, Logistic Regression and Naïve Bayes were evaluated using performance metrics namely accuracy, precision, recall and f-score. Based on the findings, the study proposed improvement for securing e-mail communication against malware-based phishing using the best performing machine-learning algorithm to keep pace with malware evolution.

KEYWORDS

Malware; Malware Analysis, Malware-based, Phishing attacks, Spams, e-mails, Machine learning, algorithms, mail filters, Detection, Mitigation techniques.

1. INTRODUCTION

In today's computerized world, especially with the spread of smart phones and Internet access, malware is becoming a major concern. Malware is software created and used by cyber-attackers to disrupt computer systems, gain computer access, or gather sensitive user information. Many problems in computer security, such as the distribution of phishing scams, are embedded in the spread of malware and botnets that are widely used in launching those attacks. While Phishing is a cybercrime model where an attacker impersonate a real person or institution by advancing them as an official person or organization between emails or other means of electronic communication [1]. Malware-based phishing attacks are among the major threats to Internet users that are difficult to track down or defend against because they do not appear to be malicious in nature. The attacker usually dispatches malevolent connections or extensions through phishing e-mails that can execute numerous tasks, such as capturing account information from the victim. A typical phishing e-mail is sent to bulk users' accounts and are dispatched to prospective victims' inboxes while consistently occurs with clickable URL links. It intends to attract the recipient into trusting that the email received is from a trusted source [2]. This attracts the recipient to visit

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 17-32, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121202

the presented website hyperlink, which connects them to fake or fraudulent websites and eventually extracts personal information.

According to statistics given by the Anti Phishing Working Group (APWG) in the 3rd quarter of 2021, the amount of phishing attacks has multiplied rapidly since the beginning 2020, and APWG observed 260,642 phishing attacks in July 2021, the exorbitant monthly attacks in APWG's reporting history (Activity & Report, 2021). According to APWG, the average wire transfer request in Business E-mail Compromise (BEC) attacks has increased from $48,000 in Q3 to $75,000 in Q4 of 2020, while the software as a service and webmail service were the mass recurring exploited by phishing in the last quarter of 2021, accounting for 29.1% of attacks. As for Tanzania, it has been noted that number of internet users in Tanzania has been increased from 27.9 million to 29.1 million from September 2020 to March 2021 respectively (TCRA, 2021). With these statistics, it shows that there is high rate of internet penetration and number of internet users who are more victims of phishing attacks across the country and the world at large. More recently, some studies such as [2] and [3] showed the number of phishing attacks have increased during the Corona virus pandemic (COVID-19) and the phishers take advantage of COVID-19 to fool their target and users especially from healthcare facilities. Many Corona virus themed spam and scam messages sent by attackers exploited people's fear of contracting COVID-19 and urgency to look for information related to Corona virus.

Even though there are email filters that use machine learning (ML) techniques and a number of researches related to phishing attacks' detection and mitigation, the interesting thing is that phishing attacks are continuing to evolve every year. Moreover, phishers and malware are becoming more intelligent and evolving through obfuscation. Thus, it was stated that the encounter between security techies and malware innovators is a continuous fight with the convolution of malware alternating as quickly as transformation heightened [5]. Consequently, it is required to keep on researching and enhancing the accuracy of the detection techniques simply because there is no single solution to the phishing problem due to the heterogeneous nature of the attack vector [5].In that aspect, there is a crucial need for evaluating machine learning algorithms for detection of malware-based phishing attacks for securing email communication.

The purpose of this study is to presents an overview about various malware based phishing attacks and various techniques used to protect users in e-mail communication. The study intends to narrow the scope and specifically deal with malware-based phishing attack identification and control techniques using ML algorithms. The study is expected to deeply evaluate the efficacy of the algorithms when integrated with major open-source mail systems' filters, as e-mail communication is the leading route used by phishers. Additionally, the research will look into the efficacy of ML in exposing phishing attacks from COVID-19 related content as some studies showed that phishing incidents massively increased during the COVID-19 pandemic era.

The remaining sections of this paper are organized as follows: Section 2 provides an overview of some literature reviews including related works; Section 3 describes methodologies used; Section 4 explores machine learning (ML) algorithms, experiment made with results and performance evaluation. In section 5, the paper provide the conclusions of the study and future work.

2. LITERATURE REVIEW

2.1. Phishing Attacks Categories

Usually, phishers conduct their attacks either by using psychological brainwashing of individuals into revealing their personal information (i.e. deceptive attacks as a form of cracking) or by

misleading users into unfolding their private information through hi-tech trickery (i.e., technical methods) by downloading malevolent code into the victim's system [5]. Although phishers prefer deceptive attacks over technical methods, mitigation of technical methods attacks cannot be overlooked. Figure 1 illustrates the types of phishing and techniques used by phishers to conduct a phishing attack whereby malware-based phishing that falls under the technical subterfuge with six (6) sub-attack techniques will be the area of study in this research. The forms of malware-based phishing attacks are described hereunder:

2.1.1. Key Loggers and Screen Loggers

Key Loggers are the type of malware used by phishers to install either through Trojan horse email attachments or through direct download to the user's computer [5]. This software monitors data and record user keystrokes and sends them to the hacker or phisher. Key loggers and screen loggers are specific variation of malware that track keyboard input and send relevant information to a hacker or phisher via the Internet [6]. They can implant themselves into users' computer browsers as small convenient plan of action that run automatically when the browser is started.

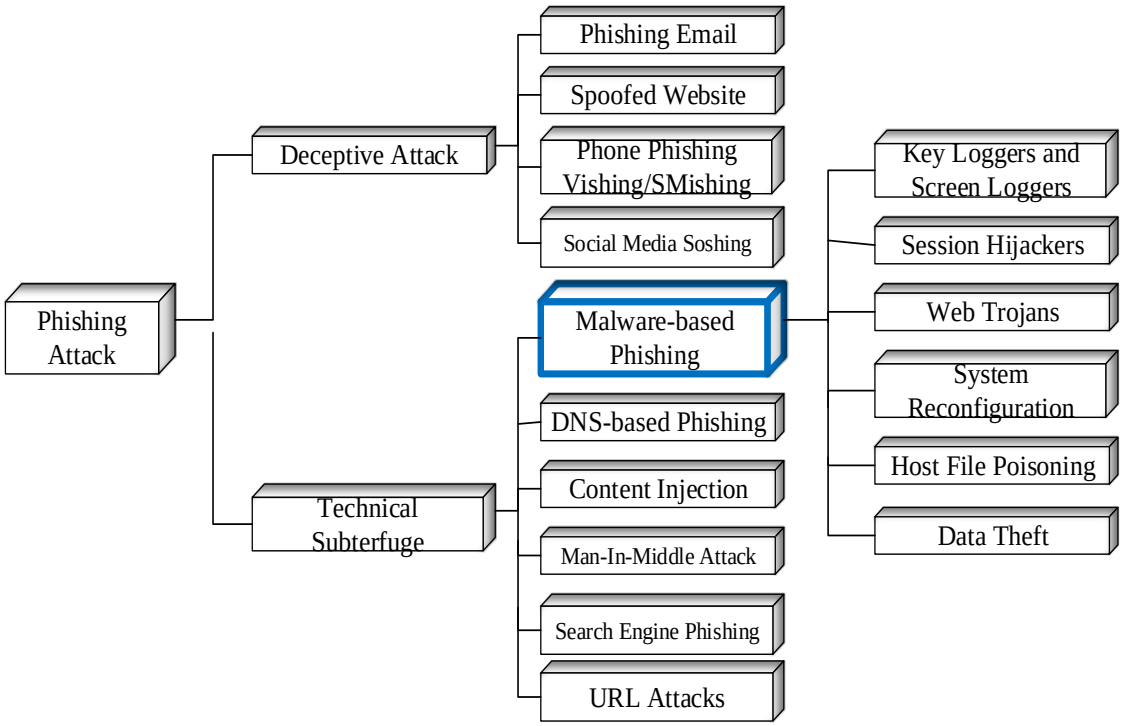


Figure 1. Types and Techniques of Phishing Attacks.
Source:[5]

2.1.2. Session Hijacking

Malware can also be used to hijack a session when a user logs into a system through a web browser to perform a transaction. The infectious software hijacks the user session and performs malicious activity once the user credentials are proved to be correct with the transacting system. In this type of phishing, the attacker observes the user's tasks by planting malevolent software inside a browser component or via network interception. Once the link is fixed, the malicious software controls and perform unwarranted actions, such as transmission of savings, without the user's knowledge [7].

2.1.3. System Reconfiguration Attacks

In this form of phishing attack, the phisher exploits the site on a user's computer for malevolent activities with the aim of compromising computer information [5]. System design can be altered using different methods, such as altering the operating system and redesigning the user's Domain Name System (DNS) server address.

2.1.4. Web Trojans

Web Trojans are malicious programs or codes that collect a user's detailed information, such as credentials, by popping up in a hidden mechanism over the login screen [5]. Phishing attacks often lead users to Web Trojans or clone websites that operate when users are trying to log on [7].These Trojans can capture important information and send them to the phisher. The sites can typically include duplicated icons and may even culminate realistic-looking SSL padlocks and third-party verification services.

2.1.5. Host File Poisoning

This kind of phishing refers to a way to trick a user into going to the phisher's site by poisoning (changing) the host's file. When the user types a particular website address in the URL bar, the web address will be translated into a numeric (IP) address before visiting the site [5]. Usually, the attacker modifies this file in order to lead the user to a fraudulent website for phishing purposes.

2.1.6. Data Theft

Data theft in phishing attacks refers to the unauthorized accessing and stealing of confidential information by a business or individual. Data theft can be done by a phishing email that leads to the download of a malicious code to the user's computer, which in turn steals sensitive information stored on that computer directly [5]. Stolen information such as system passwords, credit card information, social security numbers, and other personal data could be used directly by a phisher or indirectly by selling it for different purposes.

2.2. Malware-based Phishing Attacks Phases

A typical phishing attack includes three phases of phishers that cover several stages. To begin with, mailers send out many deceitful emails (usually through botnets), which redirect users to deceptive websites or download malicious code and install it on their machines as shown in stage 1, 2 and 3. Attackers use obfuscation techniques as the second step to conceal the malevolent texts under various layers of obscurity [8]. Various studies such as from Al-Shira'h & Al-Fawa'reh (2020) showed that constant investigation endure obfuscation and evasion attacks in most cases, while dynamic analysis itself requires a considerable amount of manual inspection for crafting detection patterns from the diversity of malware variants. Specifically, attackers try to prevent static analysis of some features by using obfuscation techniques like obfuscating the host with an IP label for malicious URLs that are statistically identical to benign ones.

Furthermore, phishers create fraudulent websites (regularly organized on compromised computers) that actively induce victims to redirect to attacker website as shown in stage 4. The victim user can also download the Remote Access Trojan (RAT) and when installed in the computer in the network, can spread in the organization network as shown in stage 5 and induce users to provide private details.

Finally, the stolen information is submitted to phisher server (stage 6) and phishers use the stolen confidential information (stage 7) to hack the user's data, such as money. The information circulation is shown in Figure 2.

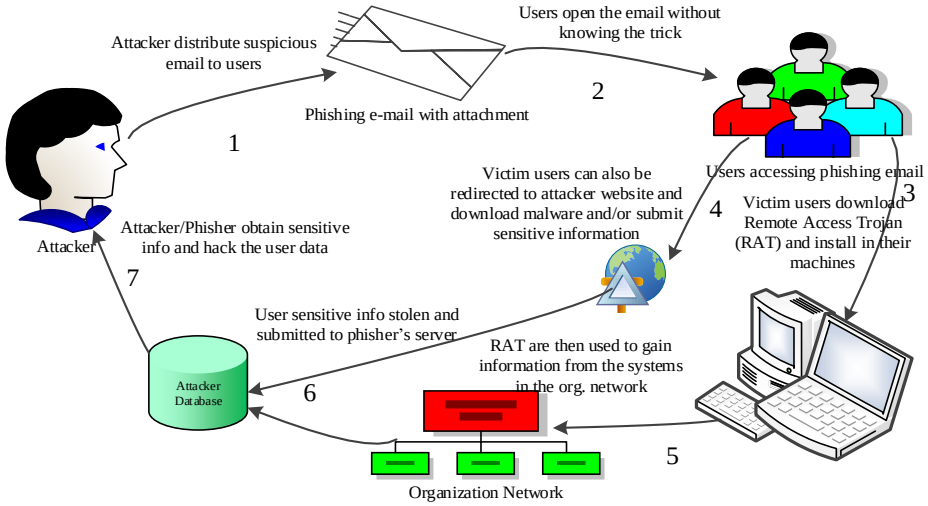


Figure 2. Information Flow in the Stages of Malware-Based Phishing Attacks

2.3. Related Works

Several studies have been conducted to address phishing attacks, detection and mitigation techniques. Each of the studies has strengths and weaknesses that could be addressed in future work by new researchers.

Rastenis et al. (2021) analyzed the existing spam and phishing email classification solutions and revealed from multiple papers that all of them are concentrated on the categorization of recognized and malicious email. As most public email datasets almost exclusively collect English emails, they investigated the suitability of automated dataset translation to adapt it to email classification written in other languages. The study focuses on solution for email classification written in only three languages, namely English, Lithuanian, and Russian [9]. The proposed solution in the study with automated translation for dataset augmentation and adaptation for the three languages prove the classification results do not decrease because of the automated translation. The result for English-only text, the accuracy was 90.07% +- 3.17% while for multi-language texts (English, Russian and Lithuanian) it was 89.2% +- 2.14%. The study was not able to demonstrate if the suggested explanation is suitable for other languages such as Swahili and how the email classification performance is affected when adapting feature optimization.

Madhavan et al. (2021) discussed the comparative analysis of disclosing fraudulent emails using various machine learning methodological analysis along with the suggested concepts with consideration of various evaluation metrics such as accuracy, efficiency, error, and evaluation time of the model. The study presented the issues based on several setbacks faced in spam filtering and classification when a particular algorithm is considered, such as evaluation time, cost, and computing resources. The study draws the variation between the strengths, weaknesses, and hindrances of some of the existing techniques that use the machine learning methodologies to identify spam emails [10]. Although the study was not able to demonstrate how efficient the developed algorithm was able to perform at best, a hybrid algorithm was suggested as the best and most feasible solution for spam detection in e-mail communication to overcome the observed challenges. Also since the study focused only on spam detection and classification, there is a need

to focus and draw contrast on the strengths, weaknesses, and limitations of malware-based phishing detection using ML algorithms and propose the best mitigation measures.

Ayman El Aassal and Shahryar Baki [11] performed a systematic study and assessment of phishing emails. The study introduced a novel taxonomy of features for phishing emails, websites, and URL detection based on their structure and how the features are processed by the web and email servers. The study proposed a novel phishing identification framework named PhishBenchused to evaluate and compare the existing features for phishing detection. The framework was also intended to act as a ready-to-use platform for security researchers. It was discovered that phishers always change their attack techniques to bypass defense mechanisms. One of the solutions suggested to minimize the attack is by retraining using a more recent dataset, as the experiment showed that slightly helps existing models to detect newer attacks. The researchers mentioned that retraining the model alone is inadequate to deal with the new attacks; unfortunately, they were not able to experiment alternative solutions.

Sanouphab Phomkeona and Okamura [12] proposed a new method to extract features from email and a deep-learning approach to detect zero-day malware spam. They extracted some features from e-mail's header and body parts that included risk words detected, machine translation detected, and other features by using several APIs. They also used four different language email datasets for more diversity and a realistic purpose to build a database of words. The experiment results showed a 78% accuracy rate for zero-day email spam detection and a 92.8% accuracy rate for normal spam. The accuracy rate for features used in the zero-day email spam detection didn't increase much because the spam email dataset used contains only normal spam and not malicious or phishing spam. Thus, there is a need to balance the dataset when conducting this study for malware-based phishing attacks, to include malicious or phishing spam dataset in order to increase accuracy and improve phishing detection and mitigation.

Gibert et al., (2020) presented a methodological review of malware identification and classification perspectives using machine learning. Different studies were reviewed, compared and examined as maintained by various factors including input features, classification algorithm, characteristics of the dataset, and the objective task. There were four main contributions, including a detailed explanation of the methods and features in a traditional machine learning workflow and literature on malware detection through deep learning. The other main contribution was a discussion on research issues and challenges faced by researchers, with emphasis placed on the problem of concept drift and the challenges of adversarial learning, among others. The study insisted on an endless battle between security analysts and malware developers due to the complications of malware development as quickly as innovation grows [4]. This study emphasizes that, there is a need to add effort in this never-ending battle of mitigating the attacks, specifically malware-based phishing attacks, for securing email communication.

Alkhalil et al. (2021) investigated problems presented by phishing and proposed a new anatomy that describes the complete life cycle of phishing attacks. The anatomy provides a wider outlook for phishing attacks with an accurate definition covering end-to-end mechanisms. The proposed new anatomy of phishing involves attacker types, attack phases, vulnerabilities, targets, threats, attack media, and attacking techniques that when combined could help in developing a holistic anti-phishing system. The study highlighted that there is no single solution for mitigating phishing attacks due to the heterogeneous nature of the attack vector but there was no any experimental setup, which prompted to conduct this research study. The study insisted on the importance of developing efficient anti-phishing techniques that prevent users from being exposed to the attack as an essential step in mitigating the attacks by detecting and/or blocking them. With regard to the stated significance, it is vital to evaluate machine learning algorithms in detection of malware-based phishing to assist in developing an efficient anti-phishing solution.

Azeez and Ajayi (2019) carried out a comparative analysis of three famous machine learning algorithms (Decision Tree, Nave Bayes and Logistics Regression Model) for verification of compromised, suspicious and fake URLs sent by spammers and phishers. The analysis determined the best of all the algorithms based on the metrics such as F-Measure, Precision, and Recall used for evaluation. The result obtained based on the confusion matrix measurement shows that the Decision Tree algorithm achieves the highest values for the three metrics and provides an efficient and credible means of maximizing detection of compromised and malicious URLs. The study cautioned on inconsistencies noticed in various researchers' findings that made corresponding results not dependable based on the values obtained and conclusions drawn from them but it was not able to provide the way forward. The authors of the study proposed that, two or more supervised machine learning algorithms can be hybridized, making one effective and more efficient algorithm for fake URL verification but were not able to implement [13]. The study aimed to design a system to detect suspicious links in e-mails and notify users instead of blocking them. The study also used only three ML algorithms to draw conclusion but some popular algorithms such as SVM could have been used for comparative analysis.

Rafatet al. (2021) showed that text pre-processing methods nullify the detection of malicious content in an obscure communication framework based on their study and experiment. They used the Spamassassin corpus as a mail filter with and without text pre-processing and examined it using machine learning (ML) and deep learning (DL) algorithms to classify it as spam e-mails. The study proposed a DL-based approach that consistently outperforms standard ML models in detecting malicious content. Although the results showed the power of DL algorithms over the standard ML in filtering spam, the effects were unsatisfactory for detecting encrypted communication for both forms of algorithms [14]. The study need to be linked with the evaluation of machine learning algorithms in detecting malware-based phishing attacks.

Sameena Naaz (2021) conducted a detection of phishing study on the Internet of Things (IoT) using a machine learning approach. The ML algorithms that include random forest classifier, support vector machine, and logistic regression have been applied to the IoT dataset for the detection of phishing attacks. The results of the study have been compared with previous studies that were carried out on the same dataset as well as on different dataset from MillerSmiles archive, PhishTank archive and Googleâ€™s. Although the study was limited to feature selection and feature extraction, as well as observation for some false alarm rates, it was found that Random Forest works better in terms of accuracy and error rate. There was a suggestion for improvement to use other methods and approaches for feature selection and feature extraction as well as the implementation of hybrid ML algorithms that improve accuracy and minimize false alarm rates [15]. However, the study was not able to mention and simulate the other methods for feature selection and feature extraction. This study will therefore focus on evaluating machine-learning algorithms in detection of malware-based phishing using different approaches for feature selection and feature extraction to improve accuracy and reduce false alarm rates.

3. METHODOLOGIES

The research methods and steps used in this study include literature review, data collection, dataset creation, practical experimentation, and integrating the ML Model with the spam filter as shown in figure 5. The steps begin with a systematic literature review that covers various studies, related works, and features for machine learning models to provide context for the topic. This is followed by data collection and then a section on dataset creation is discussed, because in order to proceed with the classifier training and testing, a dataset must be in place. Data processing, including pre-processing, classifier evaluation and results is examined. Based on the best performing machine learning algorithm, the ML model will be improved and integrated with spam filter to round up the study. In order to accomplish this study, the emulation experiment

was conducted using an environment comprising of a virtual server with Python libraries installed and mail server components such as Dovecot, Postfix, Amavis, Spam Assasin, and Webmail.

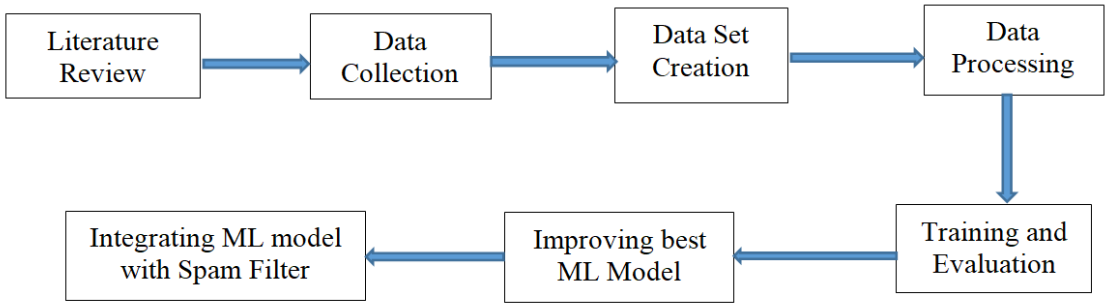


Figure 3. Research Methods Adopted

3.1 Fighting Spams and Phishing Approaches

One of the approaches most commonly used in fighting spammers is the email security filters, which use filtering techniques (spam filters). This technique is based on analyzing the message content (header and body) and other information, which can help to identify the legitimacy of the messages before they reach the user's mailbox. After identifying messages that contain scams, the action that follows depends on the settings that are applied by the mail filter itself. Some filters mostly utilize a mail server settings and usually take a separate measures of deleting the message, putting it in quarantine or labeling it as spam. However, the most appropriate method of detecting malware and spam is by using ML because they have some characteristics that are learned by the machine with the help of previously collected data in the ML algorithm [16].

Figure 4 shows the main steps taken in spam and scam mail filtering using machine learning technique. When the message is received, the initial course of action in the process is to extract the words from the message body (tokenization). This is followed by the subsequent step, which is to modify the words to their base form (lemmatization, e.g., "extracting" to "extract"). Tokenization is therefore the process of making larger words into smaller words and put into appropriate data type while lemmatization is the process of converting a word into its natural base form [17]. Also, the stop-words removal takes place by eliminating words that transpire frequently in many texts (e.g., "the," "you," "and," "to," "a," and "for") [18]. The conventional features that are usually used in spam filtering are Term Frequency with Inverse Document Frequency (TFIDF) but there were studies such as Malero, (2014) that presented alternative approach of Relative Frequency with Power Transformation (RFPT) coupled with lemmatization technique and it considerably showed improvements over TFIDF [19]. Finally, the presentation changes the messages in a format that a machine learning algorithm can use for classification.

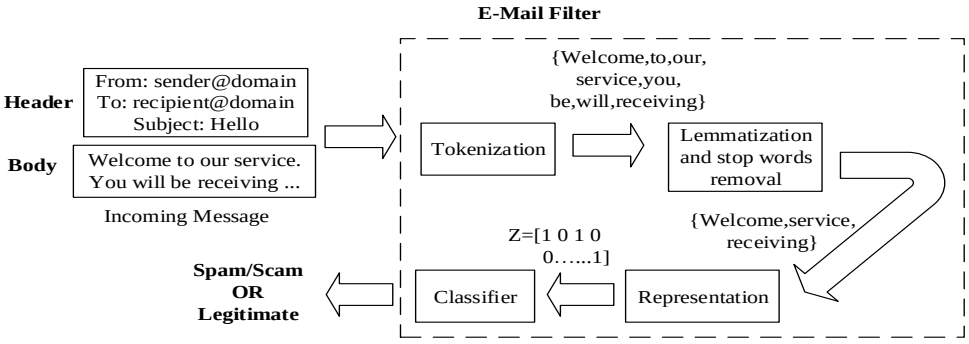


Figure 4. E-mail filtering Process

3.2. Tools Used

Scikit-Learn (SKLearn) is an environment that is incorporated with Python programming language and it is widely used in machine learning experiments. The library offers a wide range of supervised algorithms that will be suitable for this study. The library offers high-level implementation to train with the 'Fit' methods and 'predict' from an estimator (Classifier).

3.3. Machine Learning Algorithms Used

The classification techniques used in mail filtering can be grouped as content-based filtering techniques, case-based spam filtering methods, rule-based spam filtering techniques, previous likeness-based spam filtering techniques, and adaptive spam filtering techniques [20]. Various studies such as [20], [21], and [9] revealed that the most popular ML algorithms used in text classification are the Nave Bayes Classifier (NBC), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). This subsection explain each of the ML models that will be implemented to achieve the aim of this study.

3.3.1. Logistic Regression

Logistic Regression is a classification algorithm which is based on the probability concept and its cost function lies between 0 and 1. In this algorithm, the sigmoid function is used to model the data as shown in the function $g(z) = 1/(1 + e^{-z})$.

3.3.2. Naïve Bayes (NB)

Naïve Bayes model is used to resolve classification problems by using probability techniques defined by the following formula:-

$$P((\text{Phish OR Ham})|\text{WORD}) = \frac{P(\text{WORD}|(\text{Phish OR Ham})|) \text{ X } P((\text{Phish OR Ham})|)}{P(\text{WORD})}$$

There are three types of Naïve Bayes algorithms, which are Multinomial, Gaussian and Bernoulli. Multinomial Naïve Bayes (MNB), algorithm that uses Multinomial Distribution for each given feature, focusing on term frequency, has been selected to perform the spam email identification because it is text related and outperforms Gaussian and Bernoulli as per various studies.

To test this algorithm, MNB module was loaded from the Scikit-learn library. The parameters for this model are optional. If none is specified, the default values are: Alpha value set to `1.0', Fit Prior is set to `True' and Class Prior is set to `None'.

3.3.3. Support Vector Machine (SVM)

This algorithm plots each node from a dataset within a dimensional plane and through classification technique the cluster of data is separated by a hyper plane into their respective groups and is defined as:- H = VX +c

where c is a constant and V is the vector.

The Stochastic Gradient Descent (SGD) classifier, which is the linear model was loaded from scikit-learn library. SGD is the optimized version of SVM algorithm and it provide more accurate results than SVM itself [22]. Also there is a disadvantage of working with SVM algorithm since it cannot handle a large dataset, whereas SGD provides efficiency and other tuning options.

3.3.4. Decision Tree (DT) Classifier

The Decision Tree model is based on the predictive method and it creates a category which is further distributed into sub-categories or sub trees. The algorithm usually runs until the user has terminated or the program has reached its end decision. Similar to MNB and SGD, DT algorithm was loaded from the Scikit-learn library and it is executed on the default parameters which are `Gini' for Criterion and `best' for Splitter.

$$\text{Gini: } Gi = 1 - \sum_{K=1}^{n} P_{(i;k)2}$$

3.3.5. Random Forest Classifier

Random Forest (RF) algorithm can be used for both classification and regression whereby the algorithm predicts the classes by using multiple decision tree, where each tree predicts the classification class. This module was loaded from Scikit-learn library and it is based on the depth of the tree and number of DT to be produced. The termination criteria is usually considered as the more the depth and number of trees the more the computational time required for the algorithm.

3.3.6. K - Nearest Neighbor (KNN)

KNN algorithm calculates Euclidian distance and ranks the samples according to the distance between the neighbors. It makes use of the concept of similarity that helps to classify spams based upon the distance between the new mail that is to be classified and mails in the training set.

3.3.7. Multilayer Perceptron (MLP)

The MLP is a feed-forward Artificial Neural Network (ANN), which is a supervised method that includes non-linear hidden layers between the input and the output layer. The algorithm works with the linear activation function on a training dataset set by default known as Hyperbolic Tan.

$f(\bullet) : R^m R^o \rightarrow$

where ‘m’ is the input (spam words in this case) and ‘o’ is number of outputs from the function.

3.4. Datasets, Model Training and Testing Phase

As discussed through this paper, supervised learning methods were used and the model was trained with known data and tested with unknown data to predict the accuracy and other algorithms performance measures. K-Fold cross validation method was applied to acquire the reliable results although the method have disadvantages such as having a chance that the testing data could be all spam or scam emails, or the training set could include the majority of spam and scam emails. The weakness was resolved by Stratified K-fold cross validation, which separates the data while making sure to have a good range of Spam/Scam and Ham into the distributed set [22]. The parameter tuning was lastly conducted with the Scikit-Learn to improve the accuracy.

In case of datasets, the study accessed the publicly available datasets and included each email as an individual text file since the text files were string based. A list of the few spam and phishing email datasets from the public repository that were used in this study are:

(i) Ling-Spam dataset

The datasets are divided into 10 parts from the `bare' distribution that includes individual emails as a text files. This data is typical primary data since it is not pre-processed, and it includes numbers, alphabets and characters. Each part of the data was trained and tested.

(ii) Spam Assassin dataset

The dataset is more advanced with email text files and header information such as source or From address, IP address, return path, message ID and delivery information.

(iii) Enron Dataset

Enron dataset includes 6 separate datasets that contain 3000-4000 individual emails as text files. The dataset includes numbers, alphabets and characters.

(iv) Kaggle Dataset

The dataset have header and body information and the source dataset is raw as it is not pre-processed. The dataset used here contains 5568 instances with 5568 rows and 2 columns labelled as ‘Category’ and ‘Message’ respectively as shown in figure 5 and figure 6.

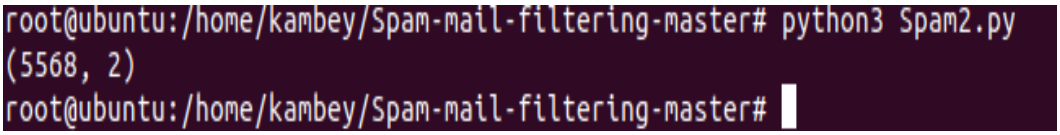


Figure 5. Kaggle dataset display

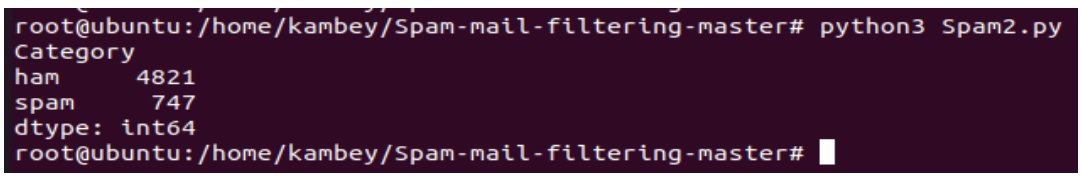


Figure 6. Kaggle dataset classification

The Python code snippet used to show dataset classification is:-

```
%matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import neighbors
data = pd.read_csv ('spamham.csv')
data1 = data.copy()
print(data1.groupby('Category').size())
```

Table 1 presents the dataset comprising of Spam/Scam and Ham with spam/Scam rate shown.

Table 1. Datasets

Dataset Name	Repository URL	Spam/Scam +Ham=Total	Rate of Spam/Ham	Published Year
Ling-Spam	http://www.aueb.gr/users/ion/data/lingspam	591 + 2304 =2895	20%	2000
SpamAssassin	https://spamassassin.apache.org/old/publiccorpus/	1918 + 4379 =6297	30%	2002
Enron dataset	http://www2.aueb.gr/users/ion/data/enron-spam/	18564 + 18261=36825	50%	2006
Kaggle dataset	www.kaggle.com	747 + 4821 =5568	13%	2012

4. RESULTS AND EVALUATION

Machine Learning algorithms play a crucial role when it comes to spam and phishing classification. Seven (7) major machine learning algorithms that are used in spam classification were discussed and experimented in this paper. The algorithms that were discussed are evaluated for their performances measure using Python Scikit-Learn tool based on the performance metrics.

4.1. Performance Metrics

A. Confusion Matrix

Though confusion matrix by itself is not a metric for performance evaluation, its components are important for the evaluation of algorithms. As the name suggests, it produces the result in the matrix form and has TP, TN, FP and FN values.

	Actual Phishing	Actual Ham
Predictive Phishing	True Positive (TP)	False Positive (FP)
Predict Ham	False Negative (FN)	True Negative (TN)

where:-

- ✓ TP indicates True Positive (correct prediction of positive case),
- ✓ TN indicates True Negative (correct prediction of negative case),
- ✓ FP indicates false positive (incorrect prediction of positive case) and
- ✓ FN indicates False Negative (incorrect prediction of negative case).

B. Classification Accuracy

The classification accuracy metric tells us that how many instances are correctly classified out of the total classified instances

Accuracy = (TP + TN)/(TP + TN + FP + FN)

***C.** Precision*

Precision indicates the number of correct prediction of positives (TP) divided by correct prediction of positives and incorrect prediction of positives. This indicates that when a model predicts positive, the precision ensures that the items are correctly labeled as positive. Hence a high precision value shows that the algorithm has returned a relevant result.

Precision =TP / (TP + FP)

D. Recall

Precision indicates the number of correct prediction of positives (TP) divided by correct prediction of positives (TP) and incorrect prediction of negative (FN). This indicates that when a model predicts positive, the precision ensures that the items are correctly labeled as positive. Hence a high precision value shows that the algorithm has returned a relevant result.

Recall = TP /(TP +FN)

Recall finds out the ratio between true positive and the sum of true positive and false negative. This will be helpful when the cost of false negative is high.

E. F1 Score

F1 score is calculated by combining precision and recall to evaluate the overall accuracy of the algorithm. Hence a low false positive and low false negative value gives a good model which has predicted the result accurately. F1score is calculated using the following formula
F1 = 2 X (precision X recall) / (precision + recall).

4.1. Performance Evaluation

The performance measures of the machine learning algorithms from the datasets presented in Table 2 of this study were simulated and analyzed. The experiment was conducted using the four (4) datasets and the average was taken. Stratified K-Fold Cross Validation (SKFCV) was applied to all the machine learning models to ensure high accuracy since the more the training data, the better accuracy the testing data provides.

The dataset were therefore split into 80:20 for training and test dataset respectively. The results obtained from the algorithms were tabulated in Table 3 below for comparison and it showed all algorithms provided 90% and above accuracy for spam/scam email detection except Random Forest classifier. Amongst the seven (7) algorithms, RF has performed poorly and SVM with optimized version (SGD) is the highest performing algorithm along with MNB that came second. The F-Score that is a measure of a model's accuracy on a dataset, for SVM is 94.81% which indicating almost perfect precision and recall.

Table 2. Performance Measures for the Machine Learning Algorithms used

	Evaluation Metrics			
ML Algorithm	**Accuracy (%)**	**Precision**	**Recall**	**F1 Score (%)**
Logistic Regression (LR)	91.76	0.54	0.95	68.86
Decision Tree (DT)	92.37	0.57	0.96	71.53
Random Forest (RF)	89.7	0.53	0.97	68.55
Multinomial Naïve Bayes (MNB)	95.6	0.66	0.99	88.35
K - Nearest Neighbor (KNN)	93.24	0.63	0.98	76.7
Support Vector Machine (SVM)	97.85	0.84	0.98	94.81
Multilayer Perceptron (MLP)	94.29	0.67	0.99	79.92

When considering the best two performing algorithms, the confusion matrix, accuracy and F-score measures is shown in Figure 7 using python SKLearn for reference.

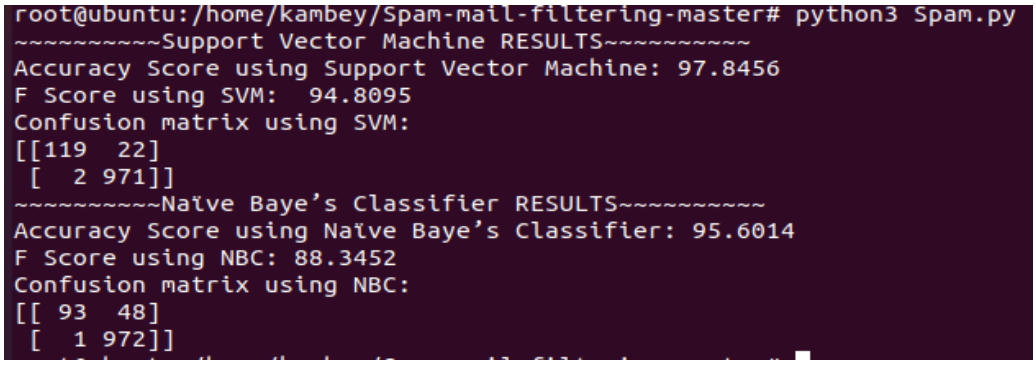


Figure 7. Performance Metrics for SVM and NB Classifier

5. CONCLUSIONS AND FUTURE WORK

This paper presents a systematic evaluation of machine learning algorithms in detecting malware-based phishing attacks. Through this study, seven machines learning algorithms were used for datasets from four different sources and the averages were calculated. This assisted in selecting the best performing algorithm based on the features considered in detecting a phishing e-mail, Also it helps develop hybrid algorithms through a combination of algorithms as their peer review is made. It is clear from the results that Support vector machines (SVM) outperforms other algorithms including closest rival Naïve Bayes (Multinomial) in detection of spam and phishing mails. Even though it is a small difference compared with MNB that also does a decent job, the better machine should always be used in solving problems such as filtering spam and malware-based phishing mails from ham mails.

As observed from all the models of classification in the field of machine learning, every method that is considered has its pros and cons. In the experiment of this study, the two best performing algorithms took a considerable computational time than the other algorithms although the time depends on the depth of a dataset and the classification. Consequently, for an efficient algorithm to be developed that performs at best even when any parameters like evaluation time, acquaintance cost and the memory of allocation, other parameters should be considered.

Therefore, hybrid algorithms seems to be the best and feasible solution for Spam and phishing detection in e-mails. In order to achieve the best detection performance in organization mail systems, it is often better to have enough training samples with balanced distributions for both malware-based phishing and benign files.

The future work that can be performed in fighting phishing attacks involves enhancing the model with more evaluation parameters for effective spam and malware-based phishing filtering.

REFERENCES

[1] M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," *6th Int. Symp. Digit. Forensic Secur. ISDFS 2018 - Proceeding*, vol. 2018-Janua, no. August, pp. 1–5, 2018, doi: 10.1109/ISDFS.2018.8355389.

[2] A. Ozcan, C. Catal, E. Donmez, and B. Senturk, "A hybrid DNN–LSTM model for detecting phishing URLs," *Neural Comput. Appl.*, vol. 0123456789, 2021, doi: 10.1007/s00521-021-06401-z.

[3] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommun. Syst.*, vol. 76, no. 1, pp. 139–154, 2021, doi: 10.1007/s11235-020-00733-2.

[4] D. Gibert, C. Mateu, and J. Planes, "The rise of machine learning for detection and classification of malware: Research developments, trends and challenges," *J. Netw. Comput. Appl.*, vol. 153, no. November 2019, p. 102526, 2020, doi: 10.1016/j.jnca.2019.102526.

[5] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy," *Front. Comput. Sci.*, vol. 3, no. March, pp. 1–23, 2021, doi: 10.3389/fcomp.2021.563060.

[6] G. RamT and R. Kumar NSR, "Analysis of Phishing in Networks," *Int. J. Sci. Eng. Res.*, vol. 4, no. 9, pp. 196–201, 2013, [Online]. Available: http://www.ijser.org

[7] S. Sagar, Shivani, and V. D. Chakravarty, "Phishing attacks and defences," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 894–897, 2019.

[8] M. Aldwairi, M. Hasan, and Z. Balbahaith, "Detection of drive-by download attacks using machine learning approach," *Int. J. Inf. Secur. Priv.*, vol. 11, no. 4, pp. 16–28, 2017, doi: 10.4018/IJISP.2017100102.

[9] J. Rastenis, S. Ramanauskaitė, I. Suzdalev, K. Tunaitytė, J. Janulevičius, and A. Čenys, "Multi-language spam/phishing classification by email body text: Toward automated security incident investigation," *Electron.*, vol. 10, no. 6, pp. 1–10, 2021, doi: 10.3390/electronics10060668.

[10] M. V. Madhavan, S. Pande, P. Umekar, T. Mahore, and D. Kalyankar, "Comparative analysis of detection of email spam with the aid of machine learning approaches," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012113.

[11] A. El Aassal, S. Baki, A. Das, and R. M. Verma, "An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs," *IEEE Access*, vol. 8, pp. 22170–22192, 2020, doi: 10.1109/ACCESS.2020.2969780.

[12] S. Phomkeona and K. Okamura, "Zero-day malicious email investigation and detection using features with deep-learning approach," *J. Inf. Process.*, vol. 28, pp. 222–229, 2020, doi: 10.2197/ipsjjip.28.222.

[13] N. A. Azeez and A. A. Ajayi, "Performance evaluation of machine learning techniques for identifying forged and phony uniform resource locators (URLs)," *Niger. J. Technol. Dev.*, vol. 16, no. 4, pp. 155–169, 2019, doi: 10.4314/NJTD.V16I4.2.

[14] K. F. Rafat, Q. Xin, A. R. Javed, Z. Jalil, and R. Zeeshan, "Evading obscure communication from spam emails," vol. 19, no. November, pp. 1926–1943, 2021.

[15] S. Naaz, "Detection of phishing in internet of things using machine learning approach," *Int. J. Digit. Crime Forensics*, vol. 13, no. 2, pp. 1–15, 2021, doi: 10.4018/IJDCF.2021030101.

[16] S. C. Et. al., "A Survey on Machine Learning Approach to Detect Malware," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 2309–2314, 2021, doi: 10.17762/turcomat.v12i2.1961.

[17] T. D. S, S. Nithya, S. P. G, and E. Pugazhendi, "Email Spam Detection and Data Optimization using NLP Techniques," vol. 10, no. 08, pp. 38–49, 2021.

[18] U. De Barcelona and V. Carvalho, "EMAIL FRAUD CLASSIFIER USING MACHINE LEARNING," 2020.

[19] A. Malero, "Applying feature transformation using Relative Frequency with Power Transformation and Lemmatization in automatic Spam Filtering," vol. 2, no. 10, pp. 21–27, 2014.
[20] E. G. Dada, J. S. Bassi, H. Chiroma, S. M. Abdulhamid, A. O. Adetunmbi, and O. E. Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems," *Heliyon*, vol. 5, no. 6, 2019, doi: 10.1016/j.heliyon.2019.e01802.
[21] S. Nandhini and D. J. Marseline, "Performance Evaluation of Machine Learning Algorithms for Email Spam Detection," Feb. 2020. doi: 10.1109/ic-ETITE47903.2020.312.
[22] S. Gibson, B. Issac, and S. Member, "Detecting Spam Email With Machine Learning Optimized With Bio-Inspired Metaheuristic Algorithms," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3030751.

AUTHORS

Kambey L. Kisambu is currently pursuing Msc in Cyber Security at the University of Dodoma, Tanzania. He earned his Bsc. Degree in Telecommunications Engineering in 2010 at the University of Dar es Salaam. His current Research interests include but are not limited to Cyber Security and Forensics, Ethical Hacking, Malware Analysis, Machine Learning and secure e-mail communication.

He has worked on several independent projects such as system development, System review, ICT security and vulnerability assessment.

He has been working in ICT industry and has full work experience for more than 10 years. He can be reached at kambeylk@gmail.com

Dr. Mohamedi Mjahidi is currently a Lecturer in the Department of Computer Engineering and Applications at the College of Informatics and Virtual Education of the University of Dodoma in Tanzania. He earned his PhD degree from the School of Natural and Applied Science at GAZİ University, Turkey in 2020. His research and teaching interest areas include but are not limited to Artificial Intelligence and Machine Learning.

He has published various research papers and he can be reached at mmjahidi@gmail.com

AN INTELLIGENT NEWS-BASED STOCK PRICING PREDICTION USING AI AND NATURAL LANGUAGE PROCESSING

Sirui Liu[1] and Yu Sun[2]

[1]Orange Lutheran High School, 2222 N Santiago Blvd, Orange, CA 92867
[2]California State Polytechnic University, Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

How do you know which stock is the right stock to invest in and have no risk of losing their money [1]? Even though there are analysis specialists out there to collect data to calculate which stock is good to be invested in, ultimately people could not afford the cost of specialists and specialists are not able to be there every minute that you want to find them. Therefore, the app Stock Recommendation is created to solve this problem where stock investment suggestions are available in touch anywhere and anytime [2]. This application helps us with what we want to invest in and gather information from recent news to show us about the public opinions towards the stock that we are looking for. Investors will no longer struggle with the problem that is the stock that they want to invest in, a good stock or a bad stock, so no money will be lost from the investor's pocket and rather, they will gain my money [4].

KEYWORDS

Stock, machine learning, AI.

1. INTRODUCTION

The Great Depression led to the first time that the stock market officially went into everyone's sight with how much the stock market changed society and people's life [9].

statistic #1: On any given day, stocks have roughly a 53 percent chance of rising and a 47 percent chance of falling. Over any given 3-month period, stocks rise 68 percent of the time, dropping the other 32 percent of the time.

statistic #2: A year after the Covid pandemic shut down the economy, stocks have gained 79% from the lows and the market is in a solid position to continue to rally. It's now being led by sectors that had been very unlikely leaders — like energy and industrials.

Some of the people had discovered and studied multiple ways to predict and proposed how the stock system will go while calculating based on the articles and statics that websites like CNN and Yahoo gives on News that allowed the stock buyer to get an understanding of which stock they should be invested in bases on the news articles that they have published about the recent stock movement that was happening [3]. However, a huge percentage of people who buy stocks do not go to a professional stock adviser but instead, they watch news to see the numbers of the stock. Their sources of information are very limited in the limited source of information that they observed from, with samples given of CNN and Yahoo News being mostly the only two sources

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 33-41, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121203

that normally people who buy stocks get information from. The limitations of only two sources are there to provide information and create a lot of limitations by the source preferences and their subjective opinions about a stock. Other techniques to calculate for should a person invest in a stock, for example, stock analysts [5]. They not only take a lot of time to analyze and give advice, but also charge a lot of money to do the stock investigation suggestions. However, plenty of time, stocks are not able to be calculated (eg. GameStop), and often results in losing huge amounts of money with investigation. A second practical problem is that giant amount of stock buyer do not relate the recent activities of one company to its stock, which forms the problem that they might keep in or sell of a stock because it shows that it is losing money or gaining money based on what the curve shows right now, and not looking forward to the future possible incomes.

Yahoo finance and CNN are two resources of stock market movement that a huge amount of people look up to for the purpose of seeking information and carefully think about their investigation towards a stock [6]. Yahoo finance and CNN are both news resources and search engines that provide information about daily stock's curve and statistics about how much a stock increases or decreases.

The app that is being built is an app that shows either positivity or negativity about whether you should invest in a stock or not. The app uses Google API to find, point out, and analyze the positive or negative words that are being found in the news resources linked in the app, and rate how positive and how negative the news is about the stock and the company that you want to invest it in.

The app gives actual suggestions about should you invest in the stock compared to Yahoo finance that just gives you a lot of statistics about the stock but not real suggestions about should you invest in the stock or not. Also the app links to more search engines compared to Yahoo finance that only has its single source of statistics.

In the application of the stock predictor, we will have two ways to demonstrate the usage. First, we show the validity of the prediction results by separating the training set and validation set. By different ways of partitioning data into training set and validation set, we can validate the accuracy of the stock prediction at each given period from the training set. By comparing the result of each training set with the validation set, a validation matrix can be computed. Through the validation matrix we can analyze the potential fitting of the machine learning function. Second, we analyze the usability of the application through a user likeability survey. Different users will try out the application and provide a subjective response based on their interaction experience. It will be analyzed with whether they agree with the trending or not. They will also rate the application based on its aesthetic value.

Introduction of the background, open problem, solution and special contribution, and paper structureThe rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Understanding the Reports

For most people yahoo finance and CNN data reports are too complicated to understand which lead to a point that they start to buy the "wrong" stock that will make them lose their money in the stock market [7]. By 2018 this dropped 8 percentage points to 34%. More alarming, less than one-third of adults understand three basic financial literacy topics by age 40, although many important financial decisions are made decades earlier. This becomes a very important problem since people start to take risks in the stock market and put plenty of their money into stocks, which they end up buying the wrong stock that makes them lose all of their money.

2.2. Choosing a Method to Enter the Stock Market

There are too many resources for stock, so it is overwhelming for people to try to enter into the stock market [8]. Thousands of websites, books, magazines, and it can be very overwhelming when people want to find out one single piece of information that they need. Most times too many news articles and news information resources have identical but conflicting sources, which about one thing, each website might have the same information but each website has a different opinion. As a result of causing confusion that makes people not understand and know which stock they should have invested in or which stock is the correct stock to invest in.

2.3. Understanding Machine Learning Websites

A lot of websites already use machine learning, but it's complicated and hard to understand. Some already use linear regression, etc [11]. We use sentiment analysis which is very good at classifying whether pages are happy or sad, good or bad news, so it's easy for people to understand how it works and what it does and it makes people feel more comfortable using our app.

3. SOLUTION

Figure 1. Stock recommendation

The application has been implemented using python and flutter, and it is carefully developed to serve as a multi-functional platform to support the visually-impaired population in navigation, during natural disasters, and in the midst of the COVID-19 pandemic [13]. The application intends to take all aspects into consideration when it provides features like QR code login, locative marker placement, vibration when detected obstacles, alert in face of disasters, GPS-frequency database, and sanitation reminder [14].

The result shows the company's movements that end up reflecting whether you should invest in or not invest into a company based on their recent news articles. By the calculations that API do, they are able to catch emotion words and rate the emotions inside of the words.

The blue links are clickable to actually browsing the website that the API gets information from since the API is still just robots, people might end up having different feelings towards the same word. So blue links that direct to the actual website are provided to let users read it themselves and think about it if they do not trust the result the app gave.

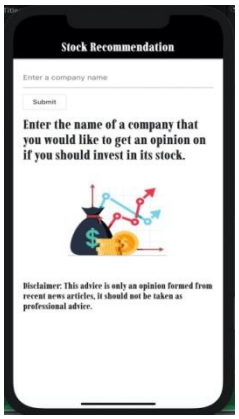


Figure 2. Screenshot of using page

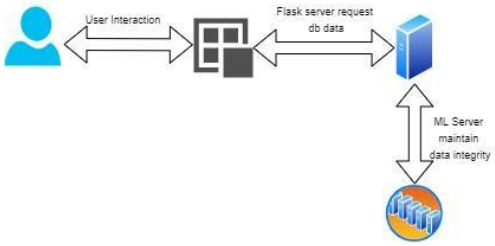


Figure 3. An overview of the project

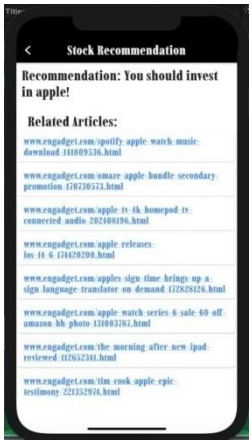


Figure 4. Screenshot of related articles

The application is designed with two main components, the front-end display and the back-end server that provides the data to be displayed. For the front end we used flutter to design the front end through the graphical interface. In the front end a user input box is provided for the user to enter the name of the company for the stock they wish to gain more information with. The user input will be saved and sent to the server through GET request. The server is written in flask with several API that provide necessary information for the front end such as articles through json.

```
35
36  class MyHomePage extends StatefulWidget {
37    MyHomePage({Key key, this.title}) : super(key: key);
38
39    // This widget is the home page of your application. It is stateful, meaning
40    // that it has a State object (defined below) that contains fields that affect
41    // how it looks.
42
43    // This class is the configuration for the state. It holds the values (in this
44    // case the title) provided by the parent (in this case the App widget) and
45    // used by the build method of the State. Fields in a Widget subclass are
46    // always marked "final".
47
48    final String title;
49
50    @override
51    _MyHomePageState createState() => _MyHomePageState();
52  }
53
54  class _MyHomePageState extends State<MyHomePage> {
55    Widget build(BuildContext context) {
56      Timer(
57        Duration(seconds: 5),
58          () => Navigator.of(context).pushReplacement(
59              MaterialPageRoute(builder: (BuildContext context) => InfoPage()))
60      );
61      return Scaffold(
62        backgroundColor: Colors.black,
63        body: Container(
64          margin: EdgeInsets.fromLTRB(20, 90, 20, 0),
65          child: Column(
66            mainAxisAlignment: MainAxisAlignment.start,
67            crossAxisAlignment: CrossAxisAlignment.center,
68            children: <Widget>[
69              FittedBox(
70                fit: BoxFit.contain,
71                child: Text('Stock\nRecommendation',
72                  style: TextStyle(color: Colors.white, fontSize: 60, fontFamily: 'Imbue', fontWeight: FontWeight.w600),
73                ),
74              ),
75              Padding(
76                padding: EdgeInsets.fromLTRB(0, 40, 0, 40),
77              ),
78              Image(
79                image: AssetImage('assets/images/stockicon.png'),
80                height: 250,
81              ),
82              Padding(
83                padding: EdgeInsets.fromLTRB(0, 40, 0, 40),
84              ),
85            ],
86          ),
87        ),
88      );
```

Figure 5. Screenshot of code 1

The design of the front end is shown above. There are three main components in the front end, which are the main page, info page, and the results page. The main page displays the question to the user to ask for the company they wish for more information about. After the user enters the information, it will parse the information, send to the information page then redirect to the request to the server (Served at https://Stock-thing.oxxxm.repl.co/results/$company). The API

will then return a json list to the result page, where the returned information will be split into a list and displayed to the front end.

```
def getArticles(company):
  # Init
  global articles, topic
  topic = company

  newsapi = NewsApiClient(api_key='5ab7a52681914e49813c2ee13f4141e4')

  d = datetime.datetime.strptime(str(date.today()), "%Y-%m-%d")
  d2 = str(d - dateutil.relativedelta.relativedelta(days=7))

  # /v2/everything
  all_articles = newsapi.get_everything(
    q = topic,                          sources = 'ars-technica, business-
insider, the-verge, bloomberg, engadget, fortune, techcrunch, techradar, the-wall-
street-journal, wired',
domains = 'marketwatch.com, fool.com, finance.yahoo.com, morningstar.com,
seekingalpha.com, investopedia.com, zacks.com,aaii.com, barrons.com,
kiplinger.com, cnbc.com, thestreet.com',
from_param = str(date.today()),              to = d2[:10],
language ='en',                         sort_by ='popularity',
page_size = 100,                        page = 1)

  articles = all_articles['articles']

```

Figure 6. Screenshot of code 2

The servers are written in Python using the flask library, where the application is created using app=Flask(app) command. The server hosts one API which is retrieveJson. In its parameter a company is entered which is passed from the front end. Based on this company name it calls the getArticles(Company) function where we use news api to search for related information. Once the results are fetched from the api, the information will be saved in a global variable called links. The function named getScored will then be used to parse each link's information semantically and generate a score for each returned article. If the article has a score between -15 to 15, the article will be used and displayed at the front end. Lastly, once this information is correctly scored and a result list has been finalized, the toJson function will jsonify the results package into json format, and the json results will be returned to the front end.

4. EXPERIMENT

4.1. Experiment 1

A good user experience is as important as a good product. So a perfect solution should have excellent user experience feedback.In order to prove that our solution has the best user feedback, we specially designed a user experience questionnaire base on the US system usability questionnaire rulesWe statistics the feedback result from 100 users, Show the user our app for 1 - 5 minutes, let them explore freely on the functionality. We divide those users into Five different groups. The first group of users ages from 10 - 20, the second group of users ages from 20 - 30, the third group of users ages from 30 - 40, the fourth group of users ages from 40 - 50, the fifth group of users ages from 50 - 60The goal of the first experiment is to verify high feedback scores

shows high performance We collect the feedback scores form these 5 different group of users and analyze it. Experiments have shown that users who ages from 30 - 40 give the highest result feedback to our app. Which may because of the age between those range are more likely to put their money in stock market [10]. The experiment graph shows below:

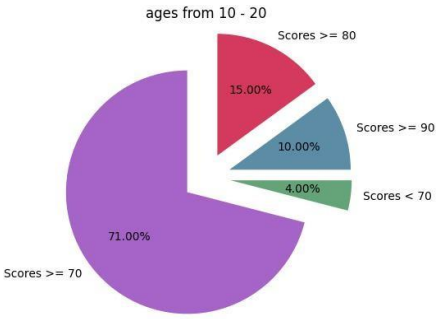


Figure 11. Results of age 10-20

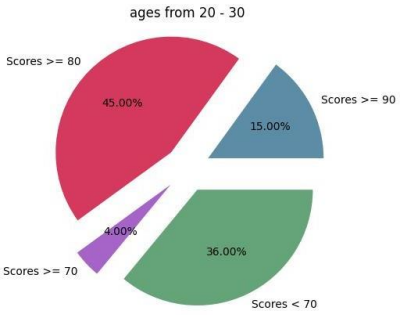


Figure 12. Results of age 20-30

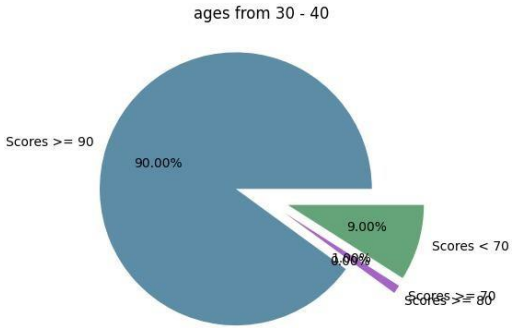


Figure 13. Results of age 30-40

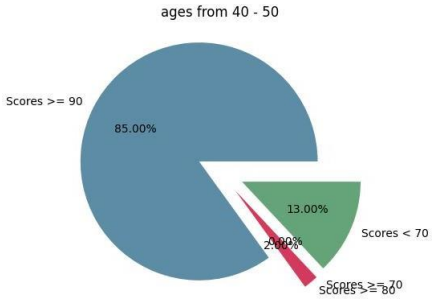


Figure 14. Results of age 40-50

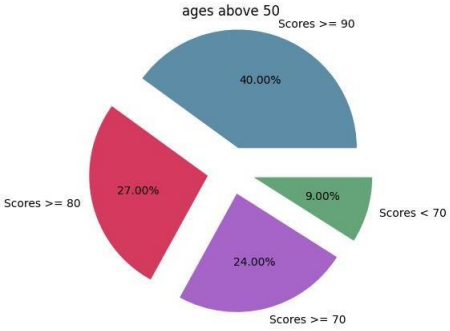


Figure 15. Results of age above 50

5. RELATED WORK

The main contribution is that the data and the api reflect the recent news of a company, and show investors that if the company was positive or negative on the recent news reports and websites.

The application was done by using repl.it and flutter which repl.it was used to write down the code to run the program of the app and the flutter was used to build the app. For example, the font, the color and different pages of the app were built in flutter. Android Studios was used to test out the app on the phone to show the errors and what needs to be improved.

6. CONCLUSIONS

In this paper, we propose a stock trading information collection system to help stock traders to acquire the information they need. We designed a user-end application using flutter,

- Propose a method/an application

  In this paper, we proposed a stock information collection system based on a flutter platform using machine learning and front-back end development [12]. Through this application users will be able to enter the name of the company they wish to find more related information and the application will search through the yahoo database then display the results [15].

- Apply the method/application to experiment

  The application was then tested through a usability test with participants number of xxx. Each of the users used the application for five minutes and rated the application through a systematic usability test survey.

- Experiment results indicate its effectiveness and solve challenges

  The result indicated that the overall usability is above average according to the usability organization, with a score of xx it indicates that the application has an above average usability score and is easy to adapt as a system.
  The application has several limitations. First it only allows you to search the keywords for the company only. Yet the application does not have ways to provide more interactive searching methods.

REFERENCES

[1] Cutler, David M., James M. Poterba, and Lawrence H. Summers. "What moves stock prices?." (1988).
[2] Walker Z, McMahon DD, Rosenblatt K, Arner T. Beyond Pokémon: Augmented Reality Is a UniversalDesign for Learning Tool. SAGE Open. October 2017. doi:10.1177/2158244017737815
[3] Gidofalvi, Gyozo, and Charles Elkan. "Using news articles to predict stock price movements." Departmentof Computer Science and Engineering, University of California, San Diego (2001): 17.
[4] Barber, Brad M., and Terrance Odean. "The behavior of individual investors." Handbook of the Economicsof Finance. Vol. 2. Elsevier, 2013. 1533-1570.
[5] Moshirian, Fariborz, David Ng, and Eliza Wu. "The value of stock analysts' recommendations: Evidence from emerging markets." International Review of Financial Analysis 18.1-2 (2009): 74-83.
[6] Xu, Selene Yue, and C. U. Berkely. "Stock price forecasting using information from Yahoo finance andGoogle trend." UC Brekley (2014).
[7] De Bondt, Werner FM, and Richard Thaler. "Does the stock market overreact?." The Journal of finance 40.3 (1985): 793-805.
[8] Barro, Robert J. "The stock market and investment." The review of financial studies 3.1 (1990): 115-131.
[9] Barsky, Robert B., and J. Bradford De Long. "Why does the stock market fluctuate?." The Quarterly Journal of Economics 108.2 (1993): 291-311.
[10] Aggarwal, Rajesh K., and Guojun Wu. "Stock market manipulations." The Journal of Business 79.4(2006): 1915-1953.
[11] Su, Xiaogang, Xin Yan, and Chih-Ling Tsai. "Linear regression." Wiley Interdisciplinary Reviews: Computational Statistics 4.3 (2012): 275-294.
[12] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects."Science 349.6245 (2015): 255-260.
[13] Ciotti, Marco, et al. "The COVID-19 pandemic." Critical reviews in clinical laboratory sciences 57.6 (2020): 365-388.
[14] Tiwari, Sumit. "An introduction to QR code technology." 2016 international conference on information technology (ICIT). IEEE, 2016..
[15] Callery, Anne, and Deb Tracy Proulx. "Yahoo! cataloging the web." Journal of internet cataloging 1.1 (1997): 57-64.

AN INTELLIGENT LOCK SYSTEM TO IMPROVE LEARNING EFFICIENCY USING ARTIFICIAL INTELLIGENCE AND INTERNET OF THINGS

Ivy Chen[1] and Ang Li[2]

[1]Troy High School, 2200 Dorothy Ln. Fullerton, CA 92831
[2]California State University, Long Beach,
1250 Bellflower Blvd, Long Beach, CA 90840

ABSTRACT

According to recent statistics, 75.4% of people with access to the internet are addicted to their phones. 78 percent of teenagers check their mobile devices at least hourly [2]. The purpose of this paper is to propose a tool that lowers users' dependence on their electronic devices. The tool named Phone Cage is created with the aim of locking electronic device for a set period of time. The application involves the user setting a specific mobile application for a specified amount of time. The phone cage provides the user a display countdown of the remaining time frame through which the locked application is inaccessible. The app provides access only when the set timer reaches the zero mark. This tool is created using Tinker cad, 3D- printer, Thunkable, Firebase console, and Raspberry Pi Zero. This will act as perfect remedy for individuals with addiction to their phones. It will also be a way for parents to control their children's use of mobile phones. Therefore, noting that a significant number of people lack self-control when it comes to cell phone usage, the cage will be of great help. The project will therefore have great impact to the community by allowing families to spend more time together and not on their phones. It will also help adults place more focus on their jobs and not on their phones.

The application has been tested by distributing the Phone Cage to ten randomly selected people across all age groups and conducted a qualitative evaluation of the approach. The result shows that the app has tremendously shrunk their work time and produced work with equal, if not higher quality.

KEYWORDS

Phone cage, Smartphone, Raspberry PI, IOS/Android.

1. INTRODUCTION

Addiction to phone usage has grown to be a global problem faced by both male and female, young and old. Estimates show that 80% of people with smart phones spend a large proportion of their time on screen based activities. Phone overuse has various adverse side effects including psychological and physical health effects [19]. A larger percentage of young people is more affected effect when compared to the older generation. This self-imposed challenge has impacted their emotional well-being, causing poor social skills and unhealthy weight gain. Also, excessive usage of smartphones, tablets, or computers can cause the young to be addicted, decreasing overall productivity and leading to needless procrastination. Smartphone addiction is caused by an Internet overuse or addiction problem, known as "nomophobia". A study done in 2012 showed

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 43-56, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121204

that in 2018, nomophobia grew from 53 percent to 66 percent [9]. However, things started getting worse in the year 2022. The Corona-virus pandemic forced people to stay indoors, increasing smartphone usage. According to a smartphone addiction poll, 99.2 percent of users experience fear and anxiety if they don't have their phone, indicating nomophobia. 37 percent of the 99.2% have mild nomophobia symptoms, 50 percent have moderate symptoms, and 13 percent have severe symptoms [13]. Nomophobia can lead to symptoms like depression, rapid heart rate, increased blood pressure, anxiety, nausea, and many more [23]. Knowing the severe cause of overusing smartphones, there is a need to start taking action to prevent phone addiction.

Nagarajan and Arthi developed a IOT smart locker. This locker is used to be a safeguard for user's personal items [7]. Both of our application used micro controller board. The project entailed components like biometric scanner and door lock system in order to make their locker more secure. Their application focuses more on security. The application can send a notification to the phone, but cannot really control the smart locker using phone. Alqahtani, Albuainaim et al. created an IOT smart locker for college [8]. They created this lock to make it more secure and convenient. The project has a keypad to open the lock if they don't have their phone. However, it uses blue tooth technique, which can be unstable. Lubans, Smith, Skinner, and Morgan created a smartphone app to help teenagers reduce screen time. The phone contained features such as 'my step', 'my workout', and 'my goal' to personalize their use of the app [24]. Since it is a digital app, the user can still have access to the phone. They can be distracted and use other entertaining app since the phone is in the user's control.

The key reason behind overuse of electronic devices is the lack of self-control [1]. For many people a phone is a luring bait. Most people do not have control of how they use their phones [20]. They feel the need to use the internet even during work or study. Following the high rate of nomophobia, there was need to find a solution hence an inspiration for creation of the Phone Cage. There is need for this app to help people curb their desire to use the phone. The application will allow individuals to operate effectively with fewer distractions.

There are various techniques used to prevent against phone addiction. They include turning on the "do not disturb" mode, silencing the phone's notifications, and deleting distracting apps. Application have been created to help set a time limit for an individual not to use their phones. However, the user can change the Screen Time settings or allow more time when the app's limit expires by typing in a pass code [17]. The proposals therefore result in lack of positive results since they assume the users already have the ability to control themselves by actively choosing to turn off their phones. In reality, people have low tendency to actually practice self-control especially when it comes to their phones. In addition, most of the available methods have easy counter-methods that undo the whole purpose: users can easily turn the notifications back on or reinstall the deleted apps. The methods are, therefore, temporary and do not result in any lasting effects. The common problem shared among all these solutions is that the phone is in the user's hand, thus, they can easily modify the setting anytime they want.

Alternative mechanism used to prevent over usage of smartphones would be more passive. Most teenagers have the experience of being restricted from their phones and or having only limited phone time. Despite them not having their phones, they are still not productive. As young adults, many teenagers desire independence [11]. They wish to gain freedom from the rules set by their parents despite lacking the skills to support themselves. Therefore, simply taking their phone away leads to reverse effects and in severe cases could worsen the parent-child relationship. Furthermore, this mechanism only works for certain age groups: specific children with phones under parental guardians. Age does not limit one from being addicted to their phone. Many adults are just as addicted to their phones and teenagers are.

Noting a need for a solution to the phone addiction, this article proposes a new idea, a physical Phone Cage. It will have a physical feature of a jail cell for electronic devices and a mobile app associated with it. The proposed method would allow users to lock their phone in a cage within a reasonably set time using the associated app. The method effectively resolves the problem on phone addiction.

The Phone Cage's unique physical feature is the ability to lock the phone making it inaccessible for the set period. Once the user places their device into the cage and starts the timer, the cage will only open when the timer goes off. This prevents users from getting distracted in the middle of their work or study as they cannot pause the locked time. The idea provides a long lasting counter for overusing phones [10]. People who use their phone simply because "it is there," now have the best solution as it prevents the user from having access to their phone whenever they wish. The 'lingering bait' is no longer able to lure the user's attention, allowing him/her to focus solely on their work.

The phone cage gives users freedom but to a certain extent. Thanks to the self-set timer feature, teenagers no longer need parents to enforce passive rules onto them. Instead, they can actively choose to lock their phones away and decide how long to lock for. This will train teenagers to decrease their dependence on electronic devices without causing issues. Phone Cage does not have an age limit [21]. Anyone from any age group can have access to this tool and have the ability to use it. It's aim is to promote productive population.

To illustrate the success of the project, two application scenarios are used. An examination of the device has been done to prove it works as expected. After rounds of trial and error, the Phone Cage worked successfully for fifty consecutive times. An illustration of the usefulness of the approach in a real-life experiment has been done. It entailed giving Phone Cage to ten of my class mass and teacher. After using the phone cage for a week, a survey was done and data collected from each user. Based on the data collected and the survey from each user, 80% of the user have increased their average worktime, and 70% of the users have increased their work quality.

The paper is organized as follows: Section 2 gives details on the challenges faced during the experiment and when designing the sample; Section 3 focuses on the details of the solutions applied in correspondence to the challenges mentioned in Section 2; Section 4 presents the relevant details of the experiment done. Finally, Section 6 gives the conclusive remarks, as well as pointing out the future work of this project.

2. CHALLENGES

When building the Phone Cage, the following challenges were encountered.

2.1. How to build the perfect phone cage

The first challenge encountered was designing the correct size and thickness of the cage and lock. The cage was designed over and over again in order to get the perfect size and thickness. The first design produced a short cage with extra spaces with poor design. We also forgot about the raspberry pi that needs to be inside the cage. After fixing the dimensions of the phone cage, we reprinted it and found out the thickness was also a problem: it breaks too easily. We decided to change the width wider and changed the 3d-printer infill density from 15% to 20% to make it stronger. Lastly, when we tried to lock the cage, we found out the lock was too tall and that it didn't fit in the slide lock. In order to fix this problem, we redesigned the lock by making the height shorter. Figure 1 is an image of all the failed attempts of the phone cage.

Figure 1. Failed attempts of the phone cage

2.2. How to get the correct angle for the micro servos

The project design was also faced with a problem in the micro servos. The micro servos were not as powerful to turn the exact angle as intended. In order to get the correct angle, we tested many times for the correct length of the pulse. After many attempts, we ended up having the pulse be 50 to 125 and the lock was able to perfectly slide out and lock the cage.

Figure 2. Length of the pause

2.3. How to let the raspberry pi run the program automatically

A problem was noted after the coding for the Raspberry Pi and the app had been completed. I had to open the raspberry pi the whole time in order to let the phone cage run the program. To solve this problem, we tried many ways to let the raspberry pi run the program automatically so my code will function at Startup. The solution lied in saving the program into raspberry pi and use of a python script program to listen to the situation.

3. SOLUTION

Phone Cage is a physical box that is created using Tinker cad, 3D- printer, Thunkable, Firebase console, and Raspberry Pi Zero. The main reason behind creating the cage is to;

1. Prevent/lower phone addiction
2. Increase productivity by isolating distraction.
3. Motivate one to be more self-controlled.

The tool was created using Tinkercad, 3D- printer, Thunkable, Firebase console, and Raspberry Pi Zero.

- The Tinker cad was used to design the overall Phone Cage and lock;
- The 3D- printer was used to print out the physical Phone Cage [5];
- Thunkable was used to create the Phone Cage app, which allows the user to set the time using a slide bar;
- Firebase console was used to store and conserve the data, Inspect the timestamp, unlock time, and whether the Phone Cage was locked or not
- Raspberry Pi Zero was used to control the micro servos arm to turn the slide lock

I started off my project with building the hardware prototype by 3D printing the physical box and lock. Then I created the phone cage app (timer) using Thunkable. This app allows the user to control how long they want to store their phone. The app is also able to connect and read data from the fire base. Lastly, I used a raspberry Pi to control the sliding lock and build connection between the phone cage and the firebase database. After finishing building the phone cage, we did system integration and publishing the apps [12]. The user can simply use this tool by putting their electronic into the Phone cage and user the app to set a time from1 min- 3 hours. When the user click on the "lock' button in the app, the micro servo will turn and slide in the lock. Our tool consist three main component

-a physical box to store the phone
-a app to control the lock/unlock time of the cage and connect and read data from firebase
-Raspberry Pi to control the micro servos arm angle of the lock and the connection between firebase and phone cage.

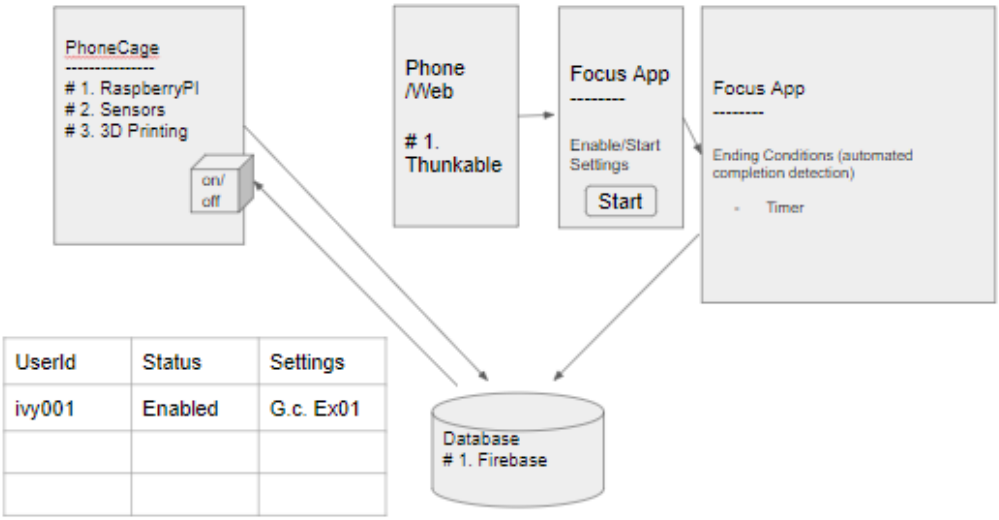


Figure 3. Overview of the system

My physical phone cage was developed by an online 3d modeling program called Tinker cad. After finishing designing the cage, I used the 3d printer and printed out the cage and the sliding lock. This step was done over and over again in order to get the perfect size and thickness of the cage and the filling for the printer.

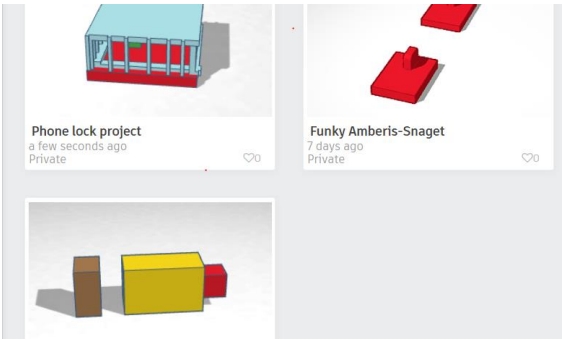


Figure 4. Online design

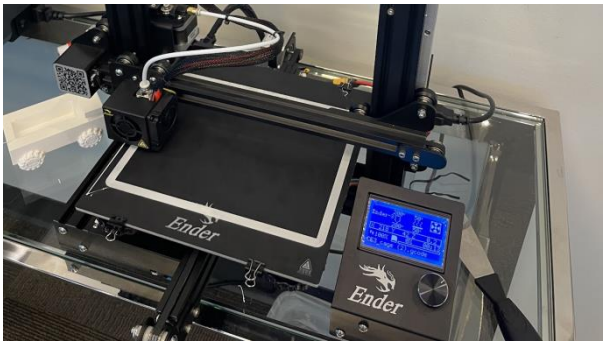


Figure 5. 3D printer

I created my application using Thunkable. In the app, the user can set the amount of time the phone is in the cage by using the slide bar. When the user clicks on the " Lock" button, the timer will start and there is no way to stop the timer. The cage will only open until the timer goes to zero. Figure 6 is an example of one of my methods, times remaining. I first set the variable to (unlock time- the time right now). If the time remaining is less than 0.1, the box will unlock so it will set the lock variable at the fire base to false. On the screen, It will return the remaining time. In the first line, I divided 3600 from the time remaining. This will show how many hours will the cage be locked. The same thing will go for minutes and seconds. I also make sure that there is a zero for the single digit. For example, instead of just showing 9 minutes, it will be 09 so the format will be correct.

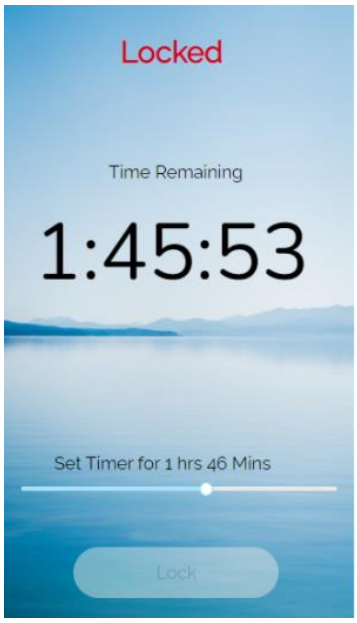


Figure 6. Screenshot of locked page

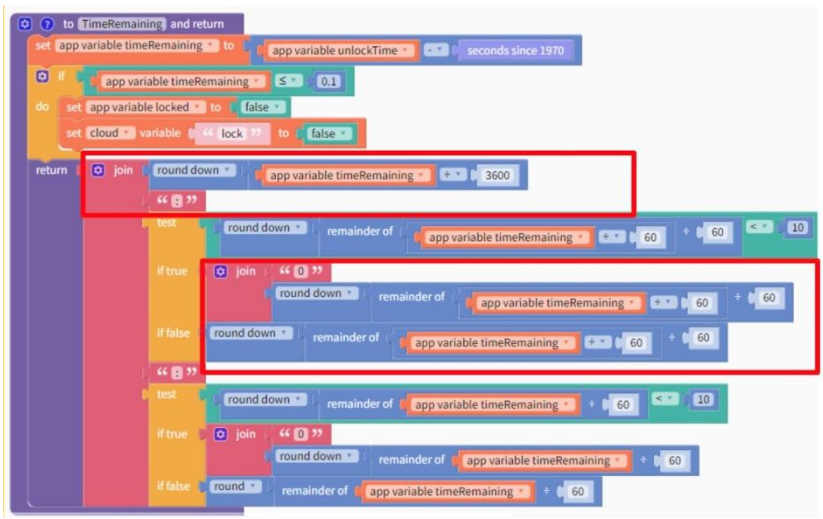


Figure 7. Screenshot of Time remaining

We connect and save data to the fire base real-time database. The fire base is linked to the Thunkable, checking the timer, timestamp, unlock time, and if the phone cage is locked or not. When the screen starts on the phone cage app, it will get the fire base data from the fire base real-time database. When the user clicks on the "start" button, the app will change and update the fire base lock in the fire base real-time database. It will display the lock variable to true, timer to (timer's value*60), timestamp to the current time, and unlock timer to (time slider's value *60 + current timestamp).

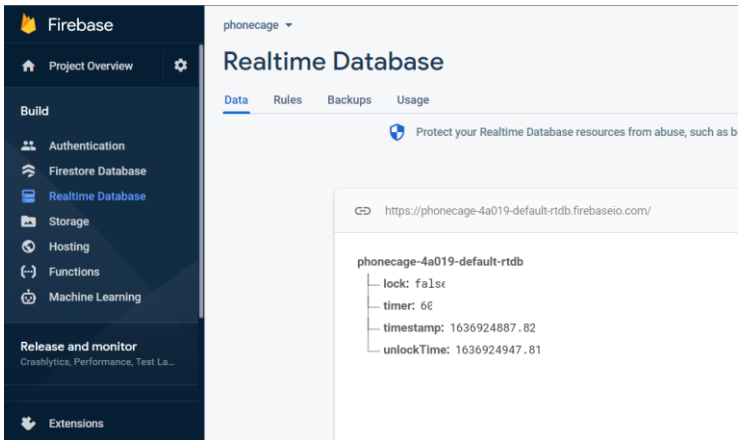


Figure 8. Screenshot of Firebase

Lastly, I used raspberry pi zero for my project. It is used to control the micro servos and the length of the pulse. The micro servos arm will turn at an angle to control the slide lock on the cage. The code below takes pin 18 to be a PWM, Pulse-width modulation, output. After setting the correct PWM mode for pin 18, It will give a clock to show how long the square wave will be. We first created a function called lockup phone and input of lockup time. We write pin 18 and give it a pulse which corresponds to the angle. In the code, the 50 and 125 is the length of the pulse. From 50 to 150, the micro servos arm will slowly increase the angle over time from opening to closing at the speed of 1, which is how smoothly the box will open/close. When the micro servo arms turn to that angle, it will result in the Phone cage being locked. When the Phone is locked, it will sleep every one second of the lock-up time. For example, if the lock time is five seconds, it will sleep five times one second. During the lock, it will also check if the micro servo is at the correct angle. After the lockup time, the length of pulse will go backward to unlock the box.

```
# set #18 to be a PWM output
wiringpi.pinMode(18, wiringpi.GPIO.PWM_OUTPUT)

# set the PWM mode to milliseconds stype
wiringpi.pwmSetMode(wiringpi.GPIO.PWM_MODE_MS)

# divide down clock
wiringpi.pwmSetClock(192)
wiringpi.pwmSetRange(2000)

delay_period = 0.01

isLocked = False

def LockUpPhone(lockuptime):
  for pulse in range(50, 125, 1):
    wiringpi.pwmWrite(18, pulse)
    time.sleep(delay_period)
  print('locked')
  for i in range(lockuptime):
    print(i)
    time.sleep(1)
    wiringpi.pwmWrite(18, 125)
  for pulse in range(125, 55, -1):
    wiringpi.pwmWrite(18, pulse)
    time.sleep(delay_period)
  print('free')
```

Figure 9. Screenshot of code

4. EXPERIMENT

To evaluate the success of my application, we varied the result with two different experiments. Experiment 1 tested the function of the application by doing repeat testing. Experiment 2 proved the success of decreasing smartphone use by using real-life experiments.

4.1. Experiment 1

The first experiment was conducted by testing the phone cage fifty times. We put my phone into the phone cage and locked it for 1-5 minutes. We then record how long the phone is locked and the result on a spreadsheet listed below (table 1). As a result, during the fifty-time testing, the phone cage was working as it is expected. We tried different lengths in locked time and each time, the phone remained locked until the timer hits zero. This experiment proves the application function adequately.

Table 1. Result of Experiment 1

Trial #	Lock Period	Success/Fail	18	1 min	Success	36	3 min	Success
1	1 min	Success	19	5 min	Success	37	4 min	Success
2	2 min	Success	20	2 min	Success	38	1 min	Success
3	5 min	Success	21	3 min	Success	39	1 min	Success
4	1 min	Success	22	4 min	Success	40	2 min	Success
5	1 min	Success	23	2 min	Success	41	1 min	Success
6	2 min	Success	24	1 min	Success	42	4 min	Success
7	1 min	Success	25	1 min	Success	43	3 min	Success
8	1 min	Success	26	2 min	Success	44	1 min	Success
9	3 min	Success	27	1 min	Success	45	1 min	Success
10	1 min	Success	28	1 min	Success	46	3 min	Success
11	2 min	Success	29	3 min	Success	47	2 min	Success
12	2 min	Success	30	2 min	Success	48	4 min	Success
13	1 min	Success	31	1 min	Success	49	1 min	Success
14	5 min	Success	32	2 min	Success	50	5 min	Success
15	1 min	Success	33	1 min	Success			
16	1 min	Success	34	1 min	Success			
17	3 min	Success	35	5 min	Success			

4.2. Experiment 2

The second experiment was conducted by giving Phone cage on ten randomly selected sample users from my high school (teachers and students). We first recorded the average time they need working on homework each day after school and the quality of their work. After a week of using Phone cage each time they are operational, we surveyed each user again through their work time and whether the quality of their work increased, decreased, or remained the same.

Statistics show that 80% of the users indicate a decrease in their average work time. Of those, 50% indicate a drastic increase in efficiency as their work time decreased by more than ⅓ of their usual work time.

Table 2. Result of Experiment 2

People	Before Phonecage Average work time (Hrs)	After Phonecage Average work time (Hrs)
1	1.5	1
2	3	3
3	2.5	2.25
4	3.25	3
5	2	1
6	4	2.5
7	4.75	3.75
8	4	3.75
9	3.5	3.5
10	1	0.5

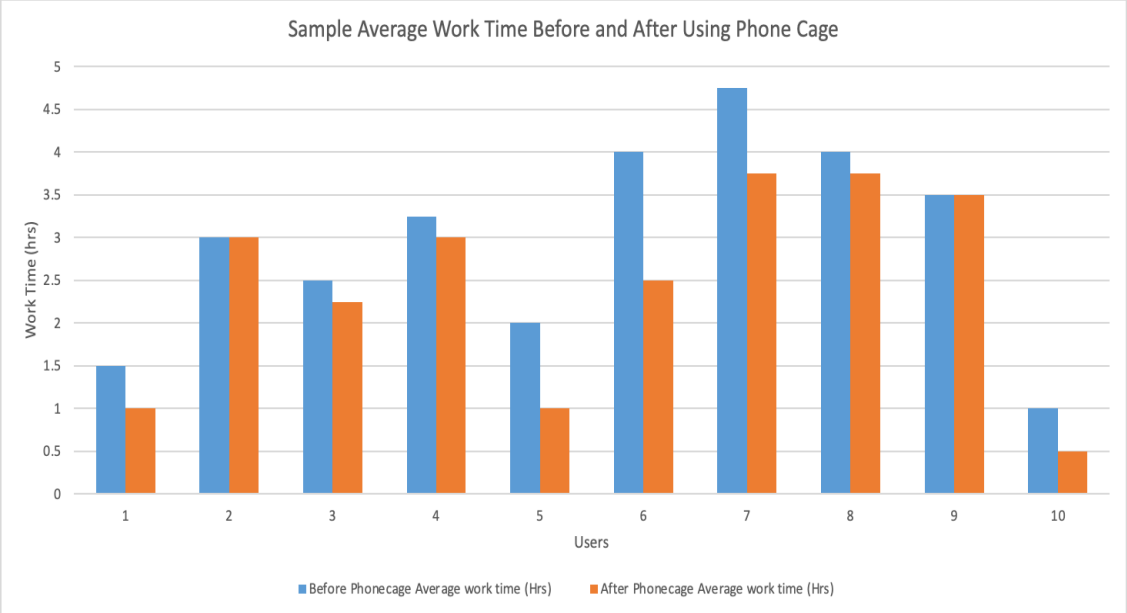


Figure 10. Chart 1 of Experiment 2

Further surveys about the quality of their work also show that 70% of the sample users admit their work quality increased after using Phone cage. The increase in quality could be demonstrated by fewer missed problems and higher performance in exams due to augmented focused study time.

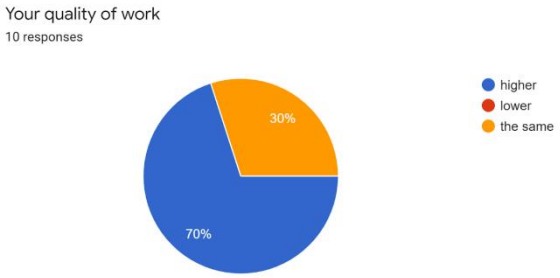


Figure 11. Chart 2 of Experiment 2

According to my own statistics, 8 out of 10 sample users we randomly selected indicate a decrease in their average work time. This significant number of users admitting positive outcomes using our Phone cage exceeded our expectations. We were confident that at least 50% of the sample would find Phone cage useful. Our data not only supports that, but also exceeds the expectations. Of the 8 that admit a decrease in their average work time, 50% indicated a drastic increase in efficiency as their work time decreased by more than ⅓ of their usual work time. (Table 2 & Figure 10) This shows the success of our phone cage to a greater extent. The sizable decrease in work time proves our point about people wasting time on meaningless electronics. Locking away phones at specific times truly boosts our sample's working efficiency. Findings from the data agrees with our central idea of decreasing the use of phones to increase people's ability to concentrate, thus producing better work at shorter times. Finally, surveys concerning the quality of work also show that 70% of the sample users admit their work quality increased after using Phone cage. The increase in quality could be demonstrated by fewer missed problems

and or higher performance on exams due to more focused study time. These line up with our expectation as we firmly believe separating useless electronics from working life will drastically increase the quantity and quality of work we get done under a specific frame of time.

5. RELATED WORK

Nagarajan and Arthi proposed the creation of an IOT smart locker dealing with security systems. It was to be applied both at home and in offices following the rising number of thefts. To provide security, the locker idea was aimed at protecting people together with their valuables [7]. Despite the proposed system being a high tech one, the team aimed at designing a low cost locker, one that could be affordable to every person in need since security would benefit all people despite their wealth. The system consists of a biometric scanner, an Arduino, piezoelectric sensor and an electromagnetic door lock system. Nagarajan and Arthi proposed a system that was pretty easy to implement and affordable hence reachable to all.

In an attempt to help people get more family time, Pamolon manufactured an easy to assemble phone prison. The project idea was aimed at improving interpersonal interactions by locking cell phones in a cell phone prison [14]. The prison was designed like a cell with a padlock and a railing. It contained a slotted bottom to allow the phone to stand neatly. The prison was not just designed for children but all individuals including adults. It had the capability of holding 6 mobile phones. It was designed in a secure way, temper proof to ensure that the phones could not be taken out unless the locker is unlocked. It contained a security locking mechanism, one that showed every times someone had accessed it. It was measured at 15 x 13 x 19cm hence spacious enough for all types of phones [22]. However, the intended phone cage to be designed will be an upgrade of this since it will incorporate a timer.

David Lubans together with his team came up with a smart phone application aimed at promoting physical activities and reducing screen time among young people. The application was created to assist in delivery of face to face obesity prevention program named Active Teen Leaders Avoiding Screen time otherwise known as ATLAS. The application was guided by social cognitive and self-determination theory [17]. It was evaluated using 361 boys from 14 different secondary schools. Following completion of the project, a group of the participants participated in an evaluation questionnaire to provide their personal perceptional of the program's performance and their experience with it. They also listed challenges encountered when using the application.

ATLAS was a multi-component and school based program. It was created with the main target being adolescent boys from low income communities and with the risk of being obese due to more screen time and lack of physical activities. The study had an ethics approval obtained from the Department of Education, the community and University of Newcastle. Majority of those who participated in using the application reported having moderate satisfaction with ATLAS and its features. However, an issue rose showing the need for more training since a significant number of people reported struggling with how to use the application. The ATLAS idea was a great one, however the phone cage one is an even better one since it is not just intended for young boys but all people in general despite their age.

6. CONCLUSIONS

To sum up my paper, my Phone cage is designed to help children, teenagers, and students and adults. It helps them work productively by limiting the use of electronics. I hope to use this project to give people just like me a motivation to change our habits of overusing phones. First in

my project, I use an online 3d modeling program called Tinker cad to design my overall phone cage and the lock. After finishing designing the cage, we used the 3d printer and printed out the physical cage and the sliding lock. This step was done over and over again in order to get the perfect size and thickness of the cage and the filling for the printer. Next, we created the app by Thunkable which lets the user can set the amount of time the phone is in the cage by using the slide bar. We connect and save data to the fire base real time database [6]. The fire base is linked to the thinkable, checking the timer, timestamp, unlock time, and if the phone cage is locked or not. Lastly, we used raspberry pi zero for my project [15]. The micro servos arm will turn at an angle to control the slide lock on the cage. We varied the effectiveness of the result by using two distinct experiments. The first experiment proves the quality of the phone cage. The second experiment proves the positive effects on the user. The two experiments prove that the phone cage is a reliable tool for the user to work more efficiently by lowering the work time and improving the efficiency. The project is therefore significant to the community since it can be used as a fun way for parents to restrict their children. It allows the children to focus on important activities such as homework and bonding with family.

There project has some limitations. First, since it is a physical phone cage, its functionality cannot be compared to that of a portable digital app. The user has to have the physical phone cage and the app in order to make this project operate normally. Also, there is only one feature for my app. Instead of just having a simply timer, I hope to create more creative conditions to open the phone cage. Lastly, I wish to add more functions for this physical phone cage. Instead of just being a simple box that is printed from the 3-d printer, I wish I can design a method that hits two birds with one stone.

I think of making tasks using my newly designed app. For example, the box will only unlock when students turn in an assignment to google classroom or finish writing 100 words on their essay. This not only motivates the student to work harder, but also reduces the boredom linked to the timer. For the physical box, I was thought of adding a portable charger or invisible shield phone sanitizer so when they use the box, they can charge and clean their phone at the same time [14].

REFERENCES

[1] Bhattacharya, Sudip Bashar, Md Abu, Srivastava, Abhay & Singh, Amarjeet (2019) “Nomophobia: no mobile phone phobia”, *Journal of Family Medicine and Primary* Care vol. 8 No. 4, pp1297-1300.

[2] Sim, Ida, (2019) "Mobile devices and health", *New England Journal of Medicine,* vol. 381, No. 10, pp956-968.

[3] De-Sola Gutiérrez, José, Rodríguez de Fonseca, Fernando & Rubio, Gabriel (2016) "Cell-phone addiction: a review", *Frontiers in Psychiatry*, Vol. 7 article 175.

[4] Kelly, James Floyd, (2014) *3D modeling and printing with Tinkercad: create and print your own 3D Models*, Que Publishing.

[5] Roberson, D. A., Espalin, D. & Wicker, R. B. (2013) "3D printer selection: A decision-making evaluation and ranking model", *Virtual and Physical Prototyping*, Vol. 8, No. 3, pp201-212.

[6] Moroney, Laurence, (2017) "The firebase realtime database." *The definitive guide to firebase*, Apress, Berkeley, CA, pp51-71.

[7] Nagarajan, L., & Arthi, A. (2017) "IOT based low cost smart locker security system", *International Journal of Research, Ideas and Innovations in Technology*, Vol. 3, No. 6, pp510-515.

[8] Alqahtani Hanan F., Albuainain, Jeehan A., Almutiri, Badriayh G., Alansari, Shahad, AL-awwad, Ghaliah B., Alqahtani, Nada N., Masaad, Samia & Tabeidi, Rania (2020) "Automated smart locker for college", *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, pp1-6.

[9] Choi, Kwisook, Son, Hyunsook, Park, Myunghee, Han, Jinkyu, Kim, Kitai, Lee, Byungkoo & Gwak, Hyesun (2009) "Internet overuse and excessive daytime sleepiness in adolescents", *Psychiatry and Clinical Neurosciences*, Vol 63, No. 4, pp455-462.

[10] Boonjing, Veera & Chanvarasuth, Pisit (2017) "Risk of overusing mobile phones: Technostress effect." *Procedia Computer Science*, Vol. 111, pp196-202.
[11] Rosenblatt, Murray, (1961) "Independence and dependence." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability and Statistics*.
[12] Hasselbring, Wilhelm, (2000) "Information system integration", *Communications of the ACM*, Vol. 43, No. 6,pp32-38.
[13] Bashir, Hasan, Seykora, John T. & Lee, Vivian (2017) "Invisible shield: review of the corneal epithelium as a barrier to UV radiation, pathogens, and other environmental stimuli", *Journal of Ophthalmic & Vision Research*, Vol. 12, No. 3, p305.
[14] Zhao, Cheah Wai, Jegatheesan, Jayanand & Loon, Son Chee (2015) "Exploring IoT application using raspberry pi." *International Journal of Computer Networks and Applications*, Vol. 2, No. 1, pp 27-34.
[15] Dasorwala, Sakina, Patil, Aniket, Vora, Raj & Kambli, Mansi (2020) "Smart locking system", *SSRN Electronic Journal*.
[16] Lubans, David R, Smith, Jordan J, Skinner, Geoff, & Morgan, Philip J. (2014) "Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys", *Frontiers in Public Health*, Vol. 2, article 41.
[17] Rudregowda, Shashidhar (2019) "Smart door lock system", *International Journal for Modern Trends in Science and Technology*, Vol. 5, No. 2, pp36-48.
[18] Azaka, Lisa (2021) "Combating smartphones addiction", *ResearchGate*, https//doi:10.13140/RG.2.2.32365.61929
[19] Harris, Bethany, Regan, Timothy, Schueler, Jordan & Fields, Sherecce A, (2020) "Problematic mobile phone and smartphone use scales: a systematic review", *Frontiers in Psychology*, Vol. 11, article 672.
[20] Ratan, Z. A., Parrish, Anne-Maree, Bin Zaman, Sojib, Saud Alotaibi, Mohammad & Hosseinzadeh, Hassan, (2019) "Smartphone addiction and associated health outcomes in adult populations: a systematic review", *International Journal of Environmental Research and Public Health*, Vol. 18, No. 22, pp1-17.
[21] Shoukat, Sehar (2019) "Cell phone addiction and psychological and physiological health in adolescents",*NCBI*.
[22] Sunday, Oluwafemi, Adesope, Olusola O, & Maarhuis, Patricia L. (2021) "The effects of smartphone addiction on learning: A meta-analysis", *Computers in Human Behavior Report*, Vol. 4, article 100114.
[23] Swendsen, Joel (2018) "Contributions of mobile technologies to addiction research", Dialogues in Clinical Neuroscience, Vol. 18, No. 2, pp213-221.
[24] Yalçın, İlimdar, Özkurt, Burhan, Özmaden, Murat & Yağmur, Rıfat (2020) "Effect of smartphone addiction on loneliness levels and academic", *International Journal of Psychology and Educational Studies*, Vol. 7, No. 1, pp208-214.

AN APPROACH CPS FOR THE SMART MONITORING OF INDUSTRIAL SYSTEMS

Nesrine Jlassi, Cédrick Béler, Omar Khlaief and Kamal Medjaher

LGP Laboratory, Toulouse INP-ENIT, Tarbes, France

ABSTRACT

Process monitoring is an important element for the long-term reliable functioning of any automated system. In fact, monitoring system is constituted of sensors installed in the physical system, in order to analyse, observe and control production systems in real time. In network, these sensors may interact with one other and with an external system via wireless communication. With recent advances in electronics, tiny sensors have appeared. Their low cost and energy consumption allow them to perform three main functions: capture data, provide information and communicate it via sensor network. In this paper, we had interested to the Cyber-Physical System (CPS) and Prognostics Health Management (PHM) domain; The CPS is one of the most important advanced technologies, it connects the physical world with the cyber using a communication layout. In other side, PHM has become a key technology for detectingfuture failures by predicting the future behaviour of the system.

KEYWORDS

Internet of Things (IoT), Cyber-Physical System (CPS), System of Systems (SoS), Cyber-Physical System of Systems (CPSoS), Wireless Sensor Network (WSN), Prognostic and Health Management (PHM), fog computing.

1. INTRODUCTION

We are fast approaching three centuries since the beginning of the original industrial revolution, which began around 1760. This is also called industry 1.0, which was based on the “Mechanization” resulting from the invention of the steam machine. It was followed by the second “mass production” using electricity and the third "digitization" using electronics and computers, marking the dawn of the fourth industrial revolution that brings us to the Internet of Things and Cyber-Physical Systems [1]. The Internet of Things (IoT) has recently become increasingly important [9]. It concerns, with more or less blurred boundaries, the massive connectivity of objects, such as sensors, telephones, or more generally objects previously disconnected. Within the factory of the future, also considered as an intelligent factory, Cyber-Physical System (CPS) will allow communication between people, machines and products [2]. As they are able to acquire and process data, they can self-control certain tasks and interact with humans via interfaces. Indeed, even a relatively simple machine can significantly increase its value if it is equipped with an appropriate on-board system for controlling and processing the device's information.

The purpose of this paper is to evaluate the contributions of the Internet of Things (IoT) to the monitoring of industrial systems, to propose a system monitoring methodology using the connected objects and finally illustrate our proposal on a case study such as the monitoring of a wind farm.

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 57-71, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121205

2. SYSTEM MONITORING

Monitoring is only one module of a complete process that allows a machine to operate with safety, productivity and quality criteria even in case of failure. As machine maintenance technology emerged, diagnostic and prognostic progressively crossed all fields. Nowadays, there are many types of professional instruments, such as sensors, counters, controllers and calculation devices, to diagnose a certain machine. These instruments can be used to acquire and analyse signals from a machine or process.

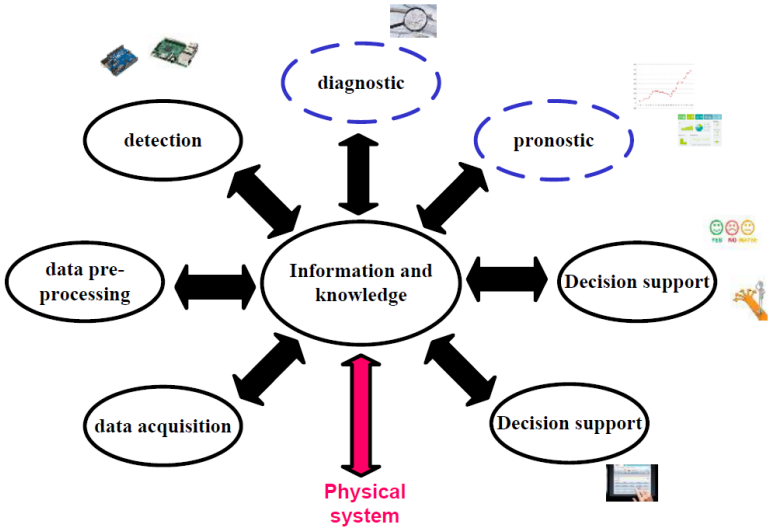


Figure 1. The main dimensions of the PHM [3]

Prognostic and Health Management (PHM) presents our application context; it aims to provide users with an integrated view of the health status of a machine or global system [16]. PHM consists of several dimensions as illustrated in Figure 1: data acquisition, data pre-processing, detection, diagnostic, prognostic and decision, except that we will limit ourselves to a few steps. Our approach is to use the different technologies of the industry 4.0 for system monitoring, and to do so, we must first define these different concepts.

3. STATE OF THE ART

After defining what system monitoring is, we will now examine what the Internet of Things consists of and then study how these two approaches can be coupled.

3.1. Internet of Things

Internet of Things (IoT) presents all physical objects equipped with information processing capacity and network connectivity to communicate with other entities: Objects, network, humans [6].

The Internet of Things aims to make it possible for things to communicate with one another, so that they can communicate with other things and users. These are integrated interconnections of various electronic devices and a fusion of two technologies: wireless connectivity and

intelligent sensors [8]. With recent advances in low-power microcontrollers, these new things are easily and inexpensively connected to the Internet [15].

3.2. Machine to Machine

Machine to Machine (M2M) is a technology that allows communication between machines without human intervention. M2M is a general term, as it does not specify specific wireless or wired networking, information and communication technologies.

This general term is particularly used by business leaders. Indeed, in M2M, four phases are involved [7]:

1. Data collection
2. Data transmission
3. Evaluation of the data collected
4. Response to the machine based on the evaluation

Machine to Machine has a wide range of applications such as industrial automation, logistics, smart grid, smart buildings, health, monitoring and security defense, automobile and transportation, etc. M2M is therefore considered an integral part of the Internet of Things and brings several benefits to industry and business [17].

3.3. Wireless Sensor Network

Wireless Sensor Network (WSN) presents the set of autonomous sensors distributed in space to cooperatively monitor and transmit their data via the network to a central location [11]. They are currently used for the real physical environment without monitoring to measure many parameters.

3.4. Cyber-Physical System

According to Bergweiler [18], the best way to represent a Cyber-Physical System (CPS) is to describe it as "systems that integrate computing and communication capabilities with monitoring and control of entities in the physical world". These systems are generally composed of a set of network agents. One of the characteristics of a CPS is that its architecture is heterogeneous, as shown in Figure 2: a system that integrates electronics and software: sensors and actuators and has communication capabilities [4] [5].

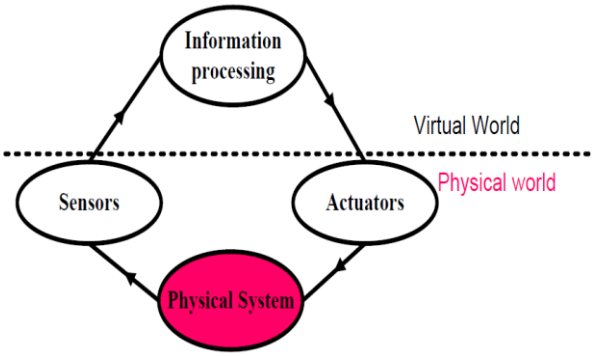


Figure 2. Principles of Cyber-Physical Systems operation

Cyber-Physical Systems perfectly integrate computation with physical processes and provide abstractions and modelling, design and analysis techniques for the integrated whole. CPSrequires computer and networking technologies to encompass not only information, but also physical dynamics. Embedded computers and networks monitor evaluate and control physical processes based on feedback control, where physical processes affect the computation process and vice versa. Within a Cyber-Physical System, the virtual world generally presents the digital twin of the real object. In fact, this twin is a way to introduce static objects into the digital world. Therefore, the result is an intelligent maintenance system that detects potential problems within the system and refines or solves the process before it becomes a problem.

3.4.1. System of Systems

Very often, we are in the systemic approach. So, a complex object is rather than a System of Systems. A System of Systems (SoS) brings together a set of cooperating systems for a task that none of the systems can accomplish on its own [10]. Each constituent system keeps its own management, goals and resources while coordinating within the SoS and adapting to meet SoS goals.

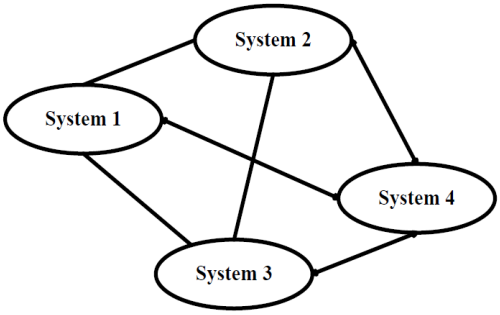


Figure 3. Architecture SoS

3.4.2. Cyber-Physical System of Systems

The combination of these SoS with Cyber-Physical Systems forms CPSoS: Cyber-Physical Systems of Systems: these are Cyber-Physical Systems that represent the characteristics of System of Systems as illustrated in Figure 4: large physical systems, often distributed in space, with complex dynamics [10].

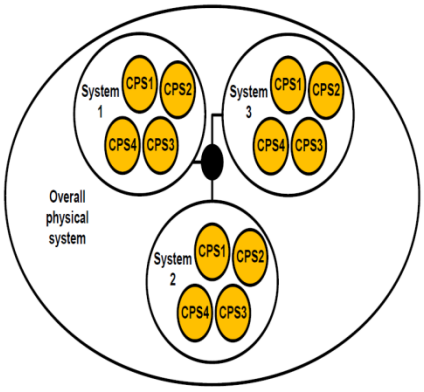


Figure 4. Architecture CPSoS

These systems allow distributed control, supervision and management and have features suchas: partial subsystem autonomy, dynamic reconfiguration of the entire system over different time scales, continuous evolution of the entire system during its operation and the possibility of emerging behaviors.

Cyber-Physical Systems of Systems can include components that are not themselves cyber-physical, for example, computer systems that manage the entire system consistently, the conceptis slightly broader than that of Cyber-Physical Systems, implying that each component of the overall system is a CPS.

3.5. Synthesis

The Figure 5 illustrates the transition of Machine-to-Machine sub-assemblies, Wireless Sensor Networks and Cyber-Physical Systems to the Internet of Things. The Internet of Things is therefore an evolutionary form of the existing ubiquitous sensor network and Machine-to-Machine (M2M).

Through these definitions, we have presented the state of the art relating to the differentconcepts of this paper in order to highlight the most appropriate methods and techniques in the field of the Internet of Things.

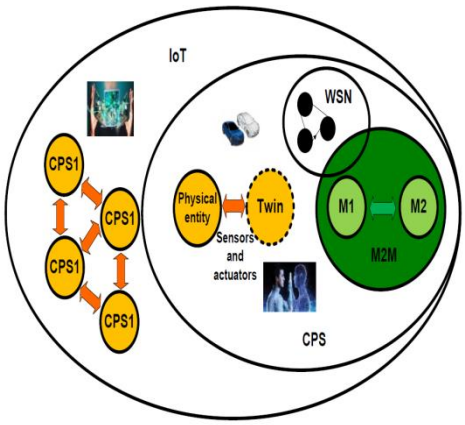


Figure 5. Synthesis diagram of IoT subsets

4. THE DIFFERENT ARCHITECTURES

Several architectures have been cited in the scientific literature: IoT, M2M [17] and 5C by Lee dedicated CPS [19].

In the IoT architecture, a construction of object meshes exists to develop increasingly intelligent systems at long distances. Thus, real-world objects transmit and receive information. The IoT architecture is made up of 3 levels: objects, fog and Cloud.

Fog level: this technology processes data from the Internet of Things locally by using clients or devices close to users to perform a substantial amount of storage, communication, control, congregation and management [14]. Finally, the fog is a new cloud paradigm designed specifically to meet the requirements of the Cloud [20].

Cloud level: cloud is a metaphor for describing the web as a space where computing has been predominant installed and exists as a service; data, operating systems, applications, storage and processing power exist on the web ready to be shared [12].

So, our contribution consists in proposing an architecture that accurately describes an intelligent monitoring method for industrial systems.

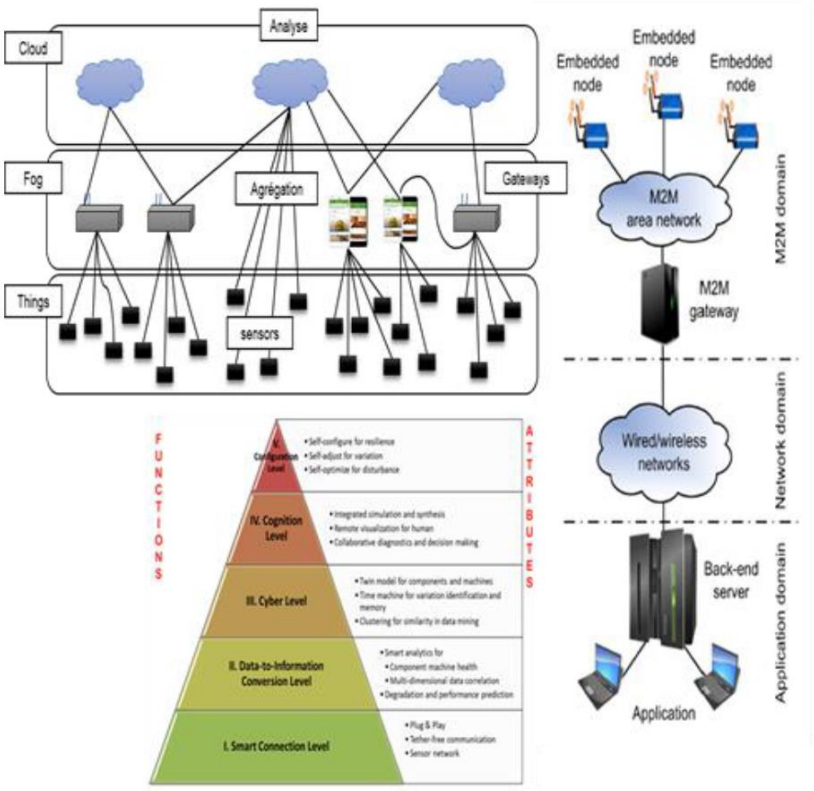


Figure 6. The different architectures

5. DISTRIBUTED INTELLIGENT SURVEILLANCE ARCHITECTURE

As illustrated in Figure 7, a Cyber-Physical System (CPS) can be modeled as a closed system loop representation. In fact, there is a similarity between the two; we also differentiate two levels when we consider a communicating object: the object itself (level 0) and the instrumented object (including communicating elements) (level 1).

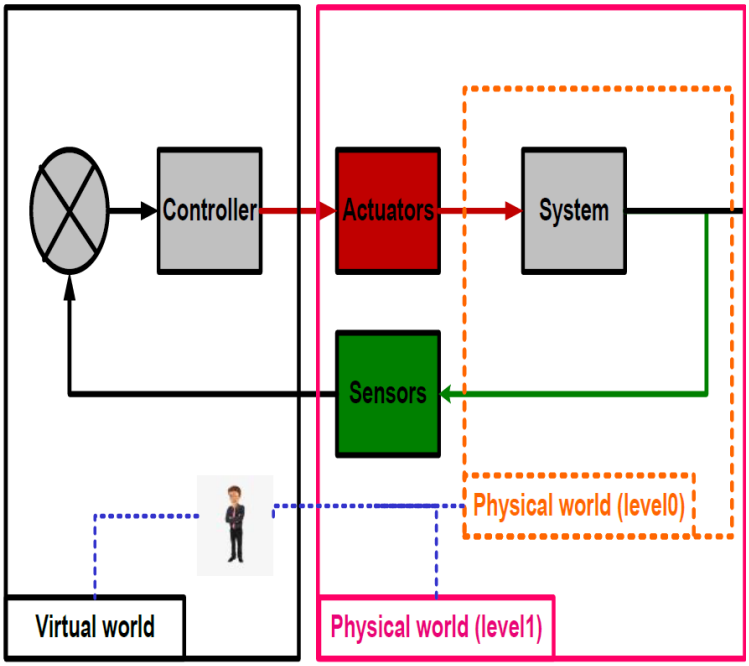


Figure 7. Closed-loop model of a CPS

Information is constantly transmitted from the physical world to the virtual space through sensors; the latter are used to collect incoming data which is then stored at the virtual world level for appropriate processing and calculation. Once this processing is completed, a generation of response actions per actuation is necessary for the good of the system. Note that this data processing is carried out at the network level through the infrastructure equivalent to fog or Cloud depending on whether we are on a local network (controlled by our equipment) or global (use of Internet, less secure networks).

5.1. Wireless Smart Node Network

In order to offer industries, the ability to monitor and control machines without any manual intervention, we have proposed a new intelligent solution based on Internet technologies for objects and Fog computing. This important solution becomes clearly necessary when dealing with a considerable number of geographically separated machines. The detailed architecture of our system is illustrated in Figure 8. As the latter illustrates, the architecture we propose is composed of three main parts which are:

The physical world: also composed of two levels; level 0, which represents the monitored system and level 1 which presents the instrumentation part, i.e. the equipment used to connect and communicate with the virtual world.

The virtual world: represents our contribution at three levels: monitoring, detection and decision support.

Fog computing: it is the infrastructure in which the virtual world is hosted.

This new surveillance architecture is characterized by intelligent wireless communication enabling a distributed, communicating and above all autonomous system.

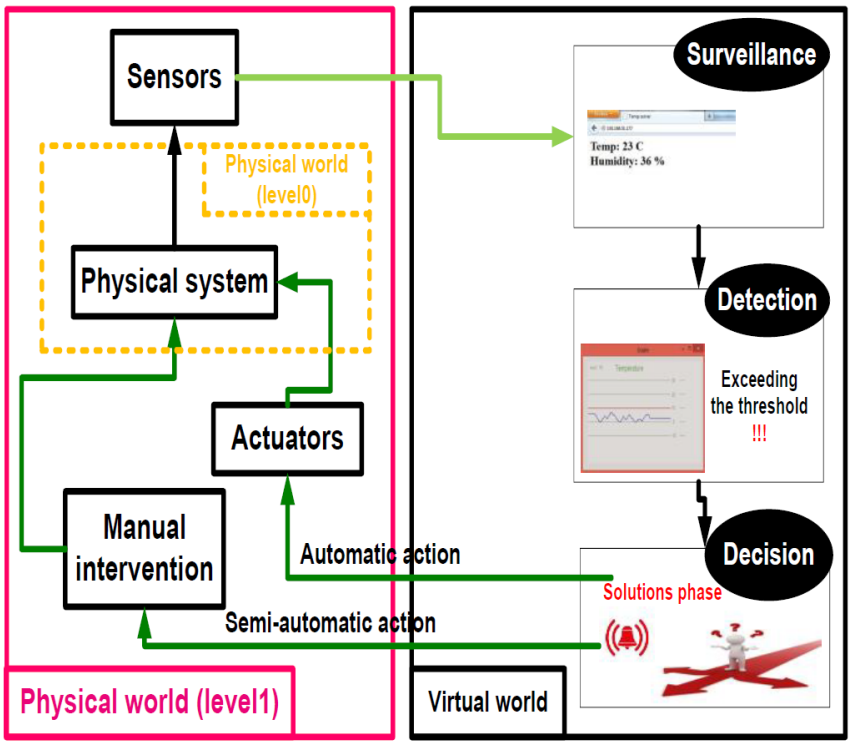


Figure 8. Proposed architecture

Wireless communication allows you to have us own processing and analysis capabilities thanks to the processor and local storage space thanks to the ROM memory. Indeed, a node is either a microprocessor or a microcontroller with wireless communication capabilities associated with sensors or actuators.

The main goal of the proposed system is to demonstrate cyber-physical systems applied to system monitoring, i.e., to build an interface for monitoring, failure detection and decision support of a system via a wireless sensor network.

In fact, the nodes of our network are distributed and wirelessly linked such that each node has processing capabilities allowing problems to be detected.

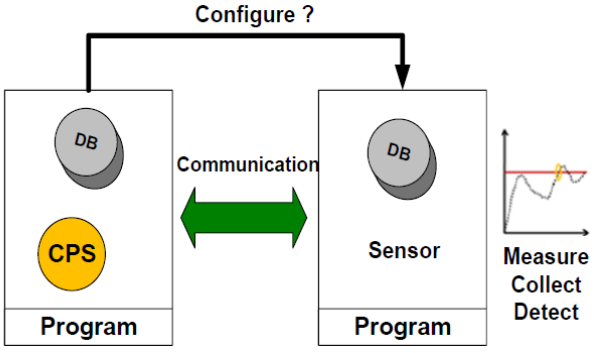


Figure 9. Example of two nodes

Let's take the example of two nodes as shown in Figure 9: the first has a CPS and the second a temperature sensor, each with a database and a program, and they are in remote communication. The sensor does three things in general: it picks up (each point = one measurement), collects and finally processes.

Data collection is not random, it is necessary to think of different scenarios that allow an intelligent distribution of decision-making responsibilities: knowing how often to collect, at what interval, at a specific request...Then, detecting any threshold overrun and therefore ensuring monitoring.

5.2. Architecture Topology

A global Cyber-Physical System (CPS) is considered to be the root of a tree that dissociates intoa set of nodes: each node has a sub-CPS that itself is formed by leaves as shown in Figure 10, so each upper level of the tree is master of the lower-level elements.

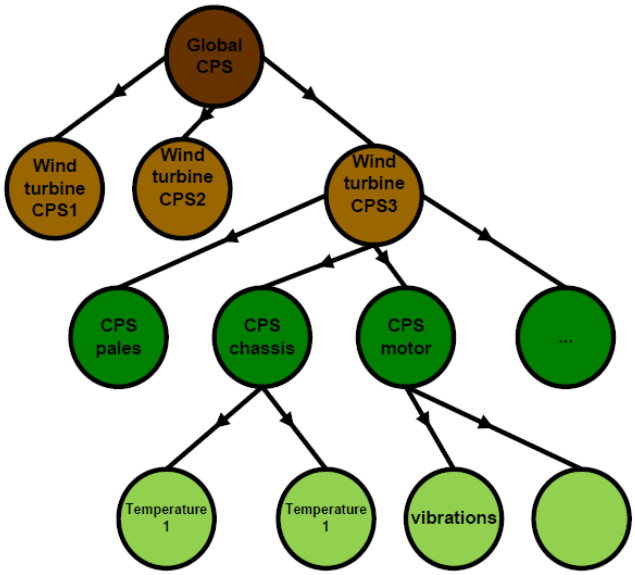


Figure 10. Tree diagram of a CPS

This tree structure allows a better localization of things and therefore the recovery of data. For example, the subdivision of a global CPS into an engine CPS and a chassis CPS allows the manufacturer to recover the operating states of the latter: temperature values as well as vibrations for the engine.

6. ARCHITECTURE PROTOTYPING

Wind energy has increased nowadays due to the proliferation of wind farms and their operation on the electricity grid by providing the electricity grid with clean and low-cost resources. As a result, there is an increasing need to establish remote monitoring of wind turbines that arehighly dependent on providing real-time safety data through a wireless connection.

The objective of our strategy is to monitor a wind farm and detect failures in order to extend the life of the turbines and thus increase productivity.

One solution is to remotely monitor a wind farm on the Internet to perform supervision, control and data acquisition tasks.

6.1. Selected Materials

The previously proposed architecture was evaluated by using a number of connected industrial devices to simulate a monitoring system. This set was composed as shown in Figure 11, a wind farm consisting of three mini-wind turbines equipped with the different modules, an anemometer which is a device for measuring wind speed, it is equipped with a mechanical sensor that rotates according to the wind power and therefore the rotational speed of the propeller is proportional to the wind speed, also a router which is a device capable of managing a small network and distributing an Internet connection to all devices connected on our "fog" network, so it is a Wi-Fi station allowing wireless devices to connect to the networks to which the router is connected and finally, a Raspberry Pi.

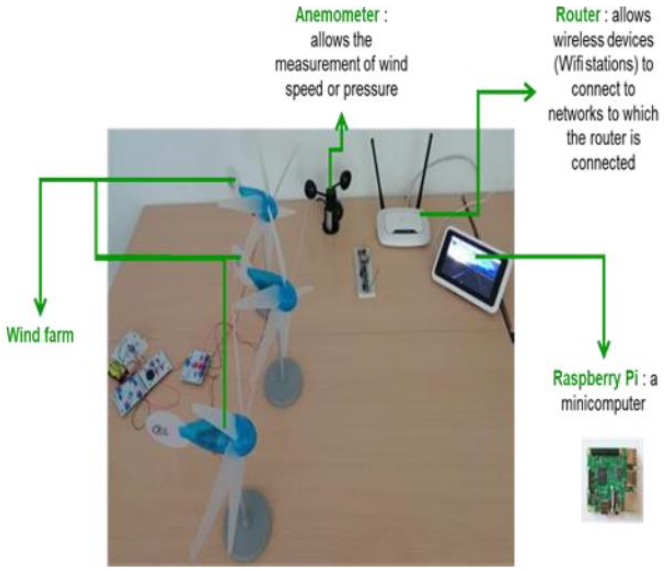


Figure 11. Selected materials

6.1.1. Arduino ESP32

Is a microcontroller designed by Espressif Systems, a Chinese company based in Shanghai. This microcontroller (figure 12) is a stand-alone Wi-Fi network solution and is also capable of running stand-alone applications. It can interface with other systems to provide Wi-Fi and Bluetooth functionality via its SPI/SDIO or I2C/UART interfaces.

Figure 12. ESP32 [21]

6.1.2. Raspberry Pi

(Figure 13) is a Linux-powered computer and is a preferred choice for Internet of Things applications since it runs on a complete kernel and has direct interfaces such as Ethernet for wired Internet as well as USB ports to connect to Wi-Fi.

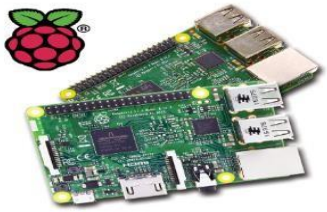


Figure 13. Raspberry PI

6.2. Data Collection

In our prototype, we used mainly two types of nodes: basic nodes and complete nodes as shown in Figure 14.

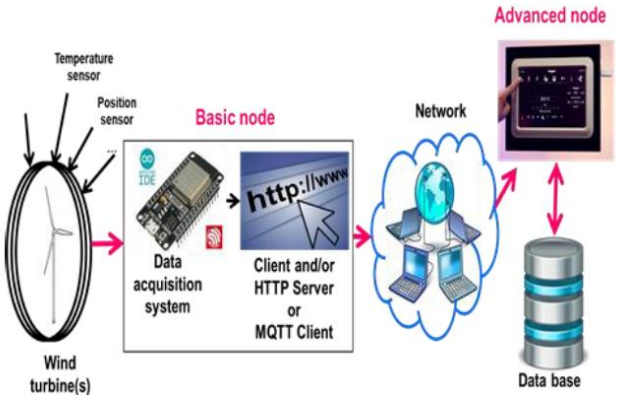


Figure 14. Data collection architecture

6.2.1. Basic node

Basic nodes contain essentially communication protocols and microcontrollers that act as a data acquisition system or are also known as data loggers; they are information systems that collect, store and distribute information. These base nodes are our server, so other systems can connect to them. Then it goes into the network. Thus, the base node can be programmed in two modes: client or server [13].

Client mode: after having correctly wired our assembly, and chosen the appropriate IDE libraries, we programmed the sensor in client Wi-Fi mode. After compilation, opening the serialconsole at a frequency of 115200 bauds gives the information captured by the sensor as shownin Figure 15.

Figure 15. Serial monitor

Server mode: to make things easier, we have installed the Arduino JSON library. Indeed, JSON (JavaScript Object Notation) is a popular data exchange format, a light text-based open standard designed for the exchange of human-readable data. It has been derived from the JavaScript scripting language to represent simple data structures and associative arrays, called objects. Despite its relationship with JavaScript, it is language independent, with analysers available for many languages.

Figure 16. Serial monitor

The serial console displays as shown in Figure 16 the IP address of our ESP32 (in our case 192.168.0.101) and there any person connected to the same local network (FOG-AP) canretrieve the information received by the sensors. So, we can see the server start and initialize its server

socket on which it listens for incoming connections. When a connection arrives, it begins to execute its request-response loop.

6.2.2. Complete Node

Aggregation: the purpose of data aggregation is to eliminate redundant data transmissions and thus improve network life. Therefore, data aggregation represents the way data is collected at sensor nodes and the routing of packets across the network.
User interface: the system developed was tested on a wind farm formed by low-power mini- wind turbines.

The user interface allows online monitoring with the goal of early defect detection preventing major component failures, facilitating a proactive response, anticipating the final shutdown of the physical system, minimizing downtime and maximizing productivity by analyzing measurements collected continuously on different types of sensors.

A user interface can be composed of several parts. For example; a part allowing the configuration of data collection. In addition, you can change the URL from where the data originates, starts data collected or even edit the period of time when the information is collected. The second part, for example, displays the values coming from the HTTP server of the monitored physical system; another part displays the evolution curves of the various monitored parameters as well as threshold values allowing the detection of any deviation from normal and thus warn the user to intervene either through an automatic action through actuators or through asemi-automatic action. Thus, the user interface allows you to display the instant control parameters of:

- The power and frequency delivered by the wind farm. In our case, we will have the power generated by the park.
- Wind speed, direction and the rotational speed of the wind turbine blades. Indeed, whenthe speed is too high, the wind turbine must be put out of service in order to avoid any damage.
- The vibrations
- The outside and inside temperature of wind turbines
- Cumulative production in 24 hours
- The state of wind turbines: either in production, starting or stopping.

Even then, we provided an image from a camera to visualize in real time the operation of the wind turbines and finally a link to current weather conditions.
Thus, the main objective of this project is achieved through the implementation of the previously proposed architecture.

The results obtained are very acceptable for remote monitoring and data acquisition. Data acquisition can be performed for all variables at an interval as small as one second, allowing accurate modeling of climatic phenomena and wind turbine operation.

7. ILLUSTRATION OF MONITORING SCENARIOS

The objective we had set ourselves was to be able to process the following 4 scenarios, which allowed us to think and design the tree architecture presented above.

7.1. Excessive Speed

Each wind turbine, regardless of the wind speed for which it has been designed, has a minimum wind speed below which it does not produce appreciable electricity and a maximum wind speed above which it must stop to avoid damage to the mechanical parts. Therefore, excessive wind can cause a wind turbine's rotation to run wild. Above a certain speed, the blades can be damaged. The anemometer is then used to detect the wind direction and speed and if the control system realizes that these speeds are high, it sends signals to the braking mechanisms to stop the wind turbine to prevent damage and therefore acts as an actuator.

7.2. Phase Failure

The rotor of each wind turbine is formed by a set of blades; it allows the conversion of the mechanical energy provided by the rotation of these blades into electrical energy. Indeed, the rotation of the rotor varies the magnetic field flux: when the north face of the rotor rotates, it causes a change in the coil pole, i.e., this pole then becomes a south pole. And conversely, when the south face of this magnet meets this same coil. This variation in magnetic field flux resultsin the production of electrical energy. The higher the rotor rotation speed, the greater the variation in magnetic field flux and therefore the greater the production of electrical energy. So, in our monitoring scenario, the phase failure of a wind turbine causes the rotor to lock, stopping power generation and the system fails.

7.3. Network failure

Communication becomes impossible, there is no more interface, no more parameters to monitor.

CONCLUSIONS

Through this project, we have succeeded in setting up a communication network in a wind farm that allows remote access to the devices connected to this network within the "fog computing" and therefore real-time monitoring of the system status.

This intelligent, efficient and robust monitoring strategy reduces maintenance costs and ensures production continuity. Indeed, it allows early detection of electrical or even mechanical defects, preventing component failures, minimizing downtime and maximizing productivity.

REFERENCES

[1] Klingenberg, C. O., & do Vale Antunes Jr, J. A. Industry 4.0: what makes it a revolution?
[2] Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration.
[3] Medjaher, K., Zerhouni, N., & Baklouti, J. (2013, July). Data-driven prognostics based on health indicator construction: Application to PRONOSTIA's data. In Control Conference (ECC), 2013 European (pp. 1451-1456). IEEE.
[4] Mois, G., Sanislav, T., & Folea, S. C. (2016). A cyber-physical system for environmental monitoring. IEEE Transactions on Instrumentation and Measurement.
[5] Lee, E. A. (2008, May). Cyber physical systems: Design challenges. In 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing (ISORC) IEEE.
[6] Vashi, S., Ram, J., Modi, J., Verma, S., & Prakash, C. (2017, February). Internet of things (IoT): a vision, architectural elements, and security issues. In I-SMAC (IoT in Social, Mobile,Analytics and Cloud)(I-SMAC), 2017 International Conference on (pp. 492-496). IEEE
[7] Kalyani, V., Gaur, P., & Vats, S. (2015). IoT:‘machine to machine’application a future vision. Journal

of Management Engineering and Information Technology (JMEIT).
[8] Aazam, M., & Huh, E. N. (2014, August). Fog computing and smart gateway based communication for cloud of things. In Future Internet of Things and Cloud (FiCloud), 2014 International Conference on (pp. 464-470). IEEE.
[9] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645- 1660.
[10] Cyber-physical Systems of Systems – Definition and core research and innovation areas. [online]. Available: http://www.cpsos.eu/wp-content/uploads/2015/07/CPSoS-Scope-paper-vOct-26-2014.pdf.
[11] Ahmed, M. R., Huang, X., Sharma, D., & Cui, H. (2012). Wireless sensor network: characteristics and architectures. World Academy of Science, Engineering and Technology, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 6(12), 1398-1401.
[12] Meraghni, S., Terrissa, L. S., Ayad, S., Zerhouni, N., & Varnier, C. (2018, March). Post-prognostics decision in cyber-physical systems. In 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET) (pp. 201-205). IEEE
[13] Adhya, S., Saha, D., Das, A., Jana, J., & Saha, H. (2016, January). An IoT based smart solar photovoltaic remote monitoring and control unit. In Control, Instrumentation, Energy & Communication (CIEC), 2016 2nd International Conference on (pp. 432-436). IEEE.
[14] Mutlag, A. A., Ghani, M. K. A., Arunkumar, N., Mohamed, M. A., & Mohd, O. (2019). Enabling technologies for fog computing in healthcare IoT systems. Future Generation Computer Systems, 90, 62-78.
[15] Sethi, P., & Sarangi, S. R. (2017). Internet of things: architectures, protocols, and applications. Journal of Electrical and Computer Engineering, 2017.
[16] Bouzidi, Z., Terrissa, L. S., Zerhouni, N., & Ayad, S. (2018). An efficient cloud prognostic approach for aircraft engines fleet trending. International Journal of Computers and Applications.
[17] Bojkovic, Z. O. R. A. N., Bakmaz, B. O. J. A. N., & Bakmaz, M. I. O. D. R. A. G. (2014, October). Machine to machine communication architecture as an enabling paradigm of embedded internet evolution. In Proc. 13th Int. Con. on Applications of Computer Engineering (pp. 40-45).
[18] Bergweiler, S. (2015). Intelligent manufacturing based on self-monitoring cyber-physical systems. UBICOMM 2015
[19] Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry4.0-based manufacturing systems. Manufacturing Letters, 3, 18-23.)
[20] Nandyala, C. S., & Kim, H. K. (2016). From cloud to fog and IoT-based real-time U-healthcare monitoring for smart homes and hospitals. International Journal of Smart Home.
[21] IOT Made Simple: Playing With the ESP32 on Arduino IDE. [online]. Available: https://www.instructables.com/id/IOT-Made-Simple-Playing-With-the-ESP32-on-Arduino-/

ARE YOUR SENSITIVE INPUTS SECURE IN ANDROID APPLICATIONS?

Trishla Shah, Raghav Sampangi and Angela Siegel

Dalhousie University, Halifax NS B3H 4R2, Canada

ABSTRACT

Android applications may request for users' sensitive information through the GUI. Developer guidelines for designing applications mandate that information must be masked/encrypted before storing or leaving the system. However not all applications adhere to the guidelines. As a prerequisite to tracking sensitive input data, it is essential to identify the widgets that request it. Previous research has focused on identifying the sensitive input widgets, but the extraction of all layouts, including images and unused layouts, is fundamental. In this paper, we propose an automated framework that finds sensitive user input widgets from Android application layouts and validates the masking of these inputs. Our design includes novel techniques for resolving the user semantics, extraction of resources, identification of potential data leaks and helping users to prioritize the sharing of sensitive information, resulting in significant improvement over prior work. We also train track the obtained sensitive input widgets and check for unencrypted transmission or storage of sensitive data. Based on a preliminary evaluation of our framework with some applications from the Google Play store, we observe notable improvement over prior work in this domain.

KEYWORDS

Android applications, sensitive, secure, GUI, layouts, framework.

1. INTRODUCTION

Applications may seek various user information through the graphical user interface (GUI), including sensitive user information such as credit card numbers, health card numbers, and social insurance numbers (SIN). Such applications must take precautions in data management and storage to protect sensitive information of users. To ensure that privacy and security requirements of users are appropriately met, it is important to vet such applications to identify abnormal behaviour.

Layouts and widgets are used to create the GUI in the Android application [1]. Layouts are containers that control widgets which are referred to as ViewGroup. Widgets are different UI components such as text boxes, labels, and buttons which are referred to as Views. These ViewGroups and Views are defined in XML files, and they may contain different attributes such as id, layout_height, and layout_width to present information to users and to collect user information. Attribute id is defined as "@+id/name", and it is unique for each widget in a layout. Developers may give any value to the IDs of such widgets [2], and, while against developer best practice guidelines, there is no current way to enforce such guidelines. This means that the IDs of widgets can take on any value and are not required to be related to the type of information being collected by the widget. Furthermore, the widgets may be placed in any order in the XML, which may be different from the order in which widgets are placed in the GUI layout, as shown in Fig.

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 73-89, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121206

1. For example, the user sees the credit card text box below the "Enter credit card number" label, but it need not be written in the same order in the XML file as shown in Fig. 1. Due to the discrepancy in the name and order in which the labels and their respective input widgets are placed in the layouts, it becomes difficult to identify which information is asked in the respective input widget. Hence, analyzing the GUI of the applications and finding the correct widget that stores sensitive information opens up research directions in this area.

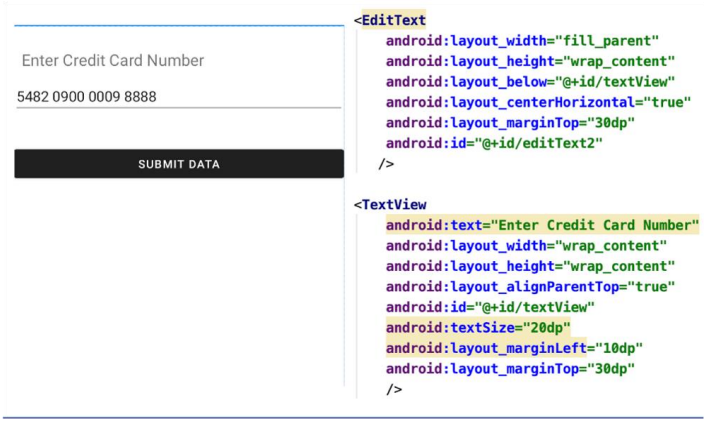


Figure 1. Sample layout and XML file

Prior research in this field has focused on resolving the widgets that seek sensitive user information [3-7]. However, such a focus on only sensitive user inputs may impact the ability to process the context and relevance of the underlying real-world information being collected, thereby affecting the semantic understanding of the context in which this information is being collected, and the overall accuracy of such observation. This motivated us to consider a deeper analysis of layout widgets, the identifiers used, the overall context in which this information is being collected, and to explore the possibility of weighting user inputs that are resolved using such a deeper analysis. This allowed us to conceptualize a framework that aims to provide users with insight about how their sensitive information is managed by an application, enabling them to make an informed decision about whether or not to share such information with the application.

Our proposed work enhances the accuracy of finding the sensitive user input widgets and advances the current state of the art by extracting more elements from the application's layouts and reducing the occurrence of false positives. We further train track these obtained input widgets to identify any sensitive data leaks. Our proposed framework will help users to make an informed decision when they use applications downloaded from the Google Play store or third-party Android app stores. While third-party Android app stores are not the recommended source to download apps [19], considering changes such as Google's announcement to increase the commission percentage for app developers [18] may consider switching venues in which their apps are made available. Such considerations pose a risk, which necessitates evaluation and mitigation of these risks prior to use of apps made available through third-party stores or the official Google Play Store.

In summary, the paper makes the following contributions:

- We develop a novel technique for analyzing the sensitive inputs of the user through the GUI improving the current state of the art (12% - 18% more accurate in detecting layouts over prior work and resolving all unused layouts in the application). We perform a direct comparison over the work done by SUPOR and UiRef by implementing their work.

- We obtain the sensitive information by analyzing the GUI of the applications and further train track this information to identify the security violations while the sensitive user data is within the application. Prior work has focused on using Data Dependency Graph (DDG) to identify security violation limited to payment data only. We design tests that apply for all the sensitive data of the users not just payment data.
- Our study highlights the loopholes in the design of current risk assessment tools (an experiment with Google's Play protect is conducted) and the need to take into account the user inputs in the security and privacy analysis of mobile applications.

The remainder of this paper describes the design and implementation of our technique to analyze the semantics of user inputs through GUI and train track of this information to identify potential security violations.

1.1. Motivation and Research Questions

Large Android application stores such as Google Play Store have implemented a "Play Protect" security feature to identify malicious applications [11]. Google Play Protect runs a safety check on applications from the Google Play Store before users download them. It also runs a safety check in the mobile phone on applications installed from outside of the Google Play Store. While running Play Protect on different sample applications which accepted and stored sensitive information such as credit card numbers, health card numbers, and SIN without masking, we found that Play Protect did not show an alert for these applications even if they were not masking user-sensitive information.

As shown in Fig. 2, we created a sample application that asked for sensitive user details such as credit card number, SIN, and passport number. The application was developed for testing purposes only and was not made available on any play stores for users to download. The sensitive user information captured by our sample application is not stored in encrypted form. It is a violation of the user's privacy and should have been detected [20]. When we passed the sample application to Google's Play Protect, it did not raise any security alerts, as shown in Fig. 2. The experiment helped us to understand the current state of art of risk assessment systems and the need to protect users' sensitive information.

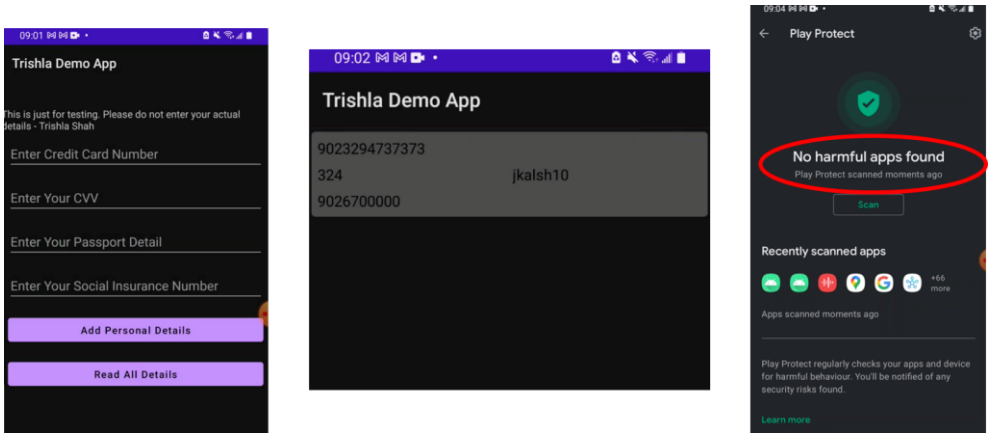


Figure 2. Google Play Protect Experiment

This prompted us to consider developing a framework that would analyze applications and identify sensitive user information, its masking, and storage of sensitive authentication data.

1.1.1. Research Questions

Based on the simple experiment conducted in the previous section, we understood the impact of poorly defined semantics in designing the application's layout. Hence, it is crucial to protect the user's sensitive data and verify that it follows appropriate privacy and security standards.

Our work focuses on automatically resolving the sensitive user inputs in Android applications through the GUI, and further train track this information to identify the potential security risks associated with it. We also calculate the weights of the identified inputs which helps to prioritize the sharing of sensitive information and inform decision making.

The following research questions guided the work:

1. Which input fields seek sensitive information?
2. Is the sensitive information masked before storing or leaving the system?
3. Is the sensitive authentication data stored after authorization?
4. For each detected input widget, what is the level of sensitivity measured by the system?

To answer the above research questions, we started our analysis by conducting a detailed literature survey of existing work in this field.

2. RELATED WORK

We previously researched the occurrence of a relay attack during credit card payments at the point of sale (POS) and designed protocol to prevent relay attack [17]. Our research then extended in the direction of securing sensitive user input data shared through the GUI of Android applications. Our work is not the first in this area of resolving user inputs and further train track the obtained information. Prior research has attempted to identify user input widgets through analysis of the GUI of Android applications [3-7] as shown in Table. 1.

Recent research in this direction [7] focuses on understanding the intentions of the icons and the sensitive GUI widgets in Android applications. However, they do not consider the problem of word ambiguity which leads to a drastic drop in the accuracy of detecting the sensitive widgets. Also, the set generated for sensitive terms and categories of user data is very limited. FlowCog [8] focuses on context-aware semantic extraction and analysis of information flow leaks in Android applications. When extracting the layouts for such semantic analysis, unused layouts must be considered to reduce the occurrence of false positives. However, no technique is incorporated in [8] to address the issues with unused layouts as well as word ambiguity.

UiRef [3] has shown better results in addressing user input semantics; however, it faces the issue of occurrences of false positives in the final result due to unused layouts in the candidate set. Also, UiRef does not consider images while extracting layout elements resulting in incomplete extraction of resources which leads to inaccurate results [3]. SUPOR[4], UIpicker [5], and AppsPlayground [6] also attempted to detect precise and scalable sensitive user input, but did not address the issue of word ambiguity and the occurrence of false positives due to unused layouts. Moreover, they do not handle custom layouts of applications which leads to insufficient analysis as most of the current applications use custom views for designing their GUI. All the existing research focuses on the extraction of resources for static layouts and do not consider dynamic layouts in their analysis. Also, no existing research in this domain focuses on measuring the weights of the detected sensitive input by the framework.

Table 1. Identification of user inputs requesting sensitive information.

Research	Primary Goal	Open Research Gaps
IconIntent (2019) [7]	Understanding the intentions of icons and identifying sensitive GUI widgets	1) Only detects sensitive icons, 2) Relies on SUPOR to identify sensitive text widgets, 3) does not consider word ambiguity problem
FlowCog (2018) [8]	Context-aware semantics extraction and analysis of information flow leaks	Does not address problem of word ambiguity and unused layouts
UiRef (2017) [3]	Resolving the sensitive user inputs	Does not considers unused layout and extraction of images
SUPOR (2015) [4]	Precise and scalable sensitive user input detection	Does not consider unused layout and extraction of images. Also, it does not address problem of word ambiguity and custom layouts
UIPicker (2015) [5]	User-input privacy identification	Same as SUPOR
AppsPlayground (2013) [6]	Automatic security analysis of smartphone applications	Same as SUPOR

To answer our research questions RQ2 & RQ3, we study the existing literature that focus on identifying and verifying the privacy and security constraints of sensitive user information in Android applications. This is achieved by tracking the flow of sensitive information in Android applications and identifying potential data leaks and security violations. These violations include 1) unencrypted or unmasked storage of sensitive user input within the device, 2) unencrypted or unmasked transmission of sensitive user input from the application, and 3) storage of authentication data within the device such as CVV.

Prior work in this direction (as shown in Table. 2) has focused only on payment data, and the test cases derived are specific to Payment Card Industry Data Security Standard (PCI DSS) [16]. Cardpliance showed promising results in detecting security violations for payment data [14]. However, it showed dependency in obtaining the user inputs on work done by UIRef. This leads to a few inaccuracies, as mentioned by UiRef, one of which is undetected user inputs. Table. 2 outlines the open research gaps by the existing work in this direction and we address it in the following subsection.

Table 2. Identification and notification of any potential data leak

Research	Primary Goal	Open Research Gaps
Cardpliance (2020) [14]	PCI DSS Compliance of Android applications	It does a keyword-based search to identify inputs related to the credit card from the list obtained from the UiRef module. It only focuses on credit card-related inputs. It relies on UiRef, which implies that it does not consider image. Also, it does not address if any extra identifiers may exist because of unused layouts which were not removed in UiRef.
FlowDroid(2014) [15]	Static taint analysis for Android applications.	Difficult to extend and configure custom test.

2.1. Research Gaps

We have identified the following research gaps based on our literature survey:

1. Prior techniques do not consider images and dynamic layouts when resolving user input widgets. These elements must be considered to improving the accuracy of detection.
2. Unused layouts must be removed to reduce the occurrence of false positives.
3. The problem of multiple-input widgets for a single label needs to be resolved to address the issue of incorrect pairing of label and text input.
4. The framework must assign some weightage to the final result which helps users to prioritize the sharing of sensitive information and to make an informed decision.
5. The above research gaps opened up directions where possible contributions can be made in this field.

3. PROPOSED WORK

Our work focuses on automatically resolving the sensitive user inputs in Android applications through GUI and further train track this information to identify the potential data leaks, achieving three main goals. First, we identify any types of sensitive information that the application requests. Sensitive information is any input that reveals users' personal information such as health records, payment information, SIN, and passport number. Second, we identify if the application follows appropriate Android developers' security and privacy guidelines [20] while storing or transmitting the requested sensitive information to avoid data leaks. Third, we assign weights to the obtained results, representing our detection strength. It will help users prioritize sharing sensitive information with a particular application. It also enables users to make informed decisions. Our framework is as shown in Fig. 3.

We divide our work into three sections: 1) Identification of user inputs requesting sensitive information 2) Identification and notification of any potential data leaks, and 3) Assigning weights to the obtained sensitive inputs.

3.1. Identification of user inputs requesting sensitive information

We focus on resolving the data semantics of user input widgets in Android applications by analyzing the GUI of these applications. As shown in Fig. 3 the module 1 has three main components: Extraction and optimization of resources, filtering of sensitive labels and mapping of sensitive labels with relevant user input. Extraction and optimization of resources will extract the layouts and exports the rendered layouts and resources for further analysis. We then filter the sensitive labels from the extracted layouts and resources, which is analyzed further for mapping of sensitive labels with the input widgets. We also apply text analytics technique to extract the sensitive labels and its context to a given layout.

For the identification of user inputs that request sensitive information (Module-1), we bifurcate our work into the following segments: 1) Extraction and optimization of resources, 2) Filtering of sensitive labels, and 3) Mapping of sensitive labels with relevant user inputs

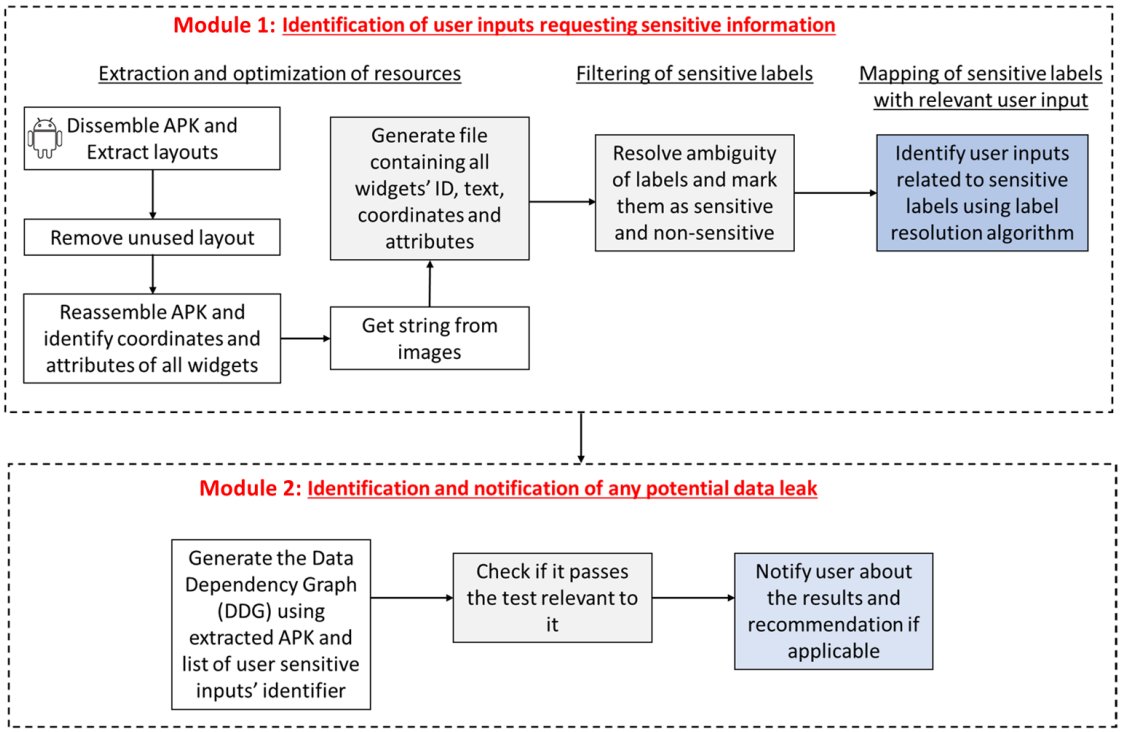


Figure 3. Proposed framework

3.1.1. Extraction and optimization of resources

In Android, there are different types of layouts: linear layout, constraint layout (relative layout), and frame layout. Each of these layout will have a different placement of the labels and input widgets (as shown in Fig. 4) The developer uses these layouts to design the GUI of the application. Whenever a layout is created, its corresponding XML file also gets generated (as shown in Fig. 4).

Statically parsing the XML file to extract the widgets is one of the methods. However, only parsing the XML file statically can lead to undetected widgets as most applications use a custom view [3]. Android does not force the developers to place widgets in a specific order and only provides guidelines. Hence, it is not easy to find the position of widgets based on the order in which the user sees them. In our work, we examine the XML files statically to retrieve all widgets and then render these widgets on the Android virtual device (AVD) to identify their coordinates which will be used further to map labels and sensitive user inputs.

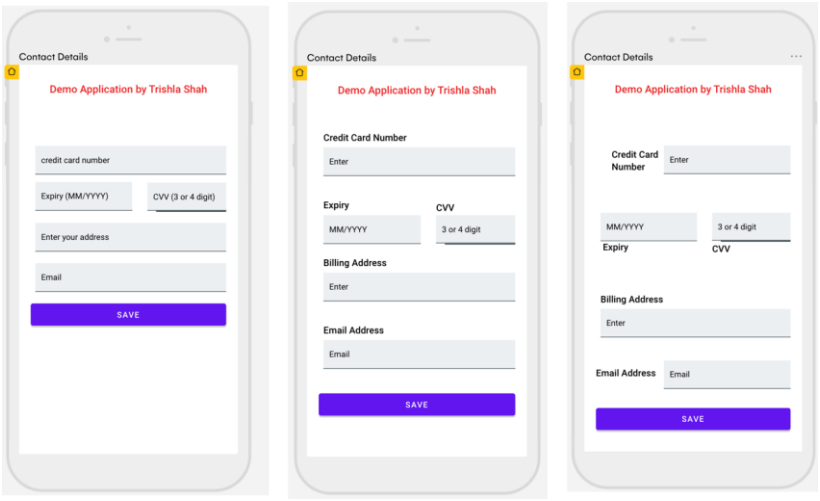


Figure 4. Different types of layouts

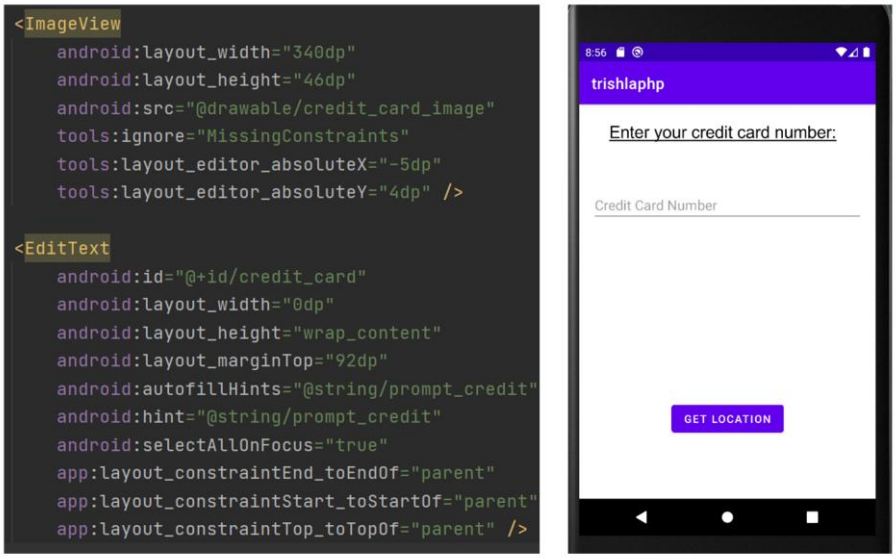


Figure 5. EditText with a label as an ImageView

The goal of this module is five-fold:

1. Extracting all the layouts from the applications
2. Removing unused layouts
3. Identifying widgets such as EditText and labels
4. Identifying images and extracting the text from them
5. Retrieving coordinates of all widgets

While developing the application, there may be a few layouts that the developers design but are not used when the application is executed. These layouts are known as unused layouts. During our initial analysis, we found that most of the current applications have unused layouts in the application package. Prior research [3] shows that the presence of unused layouts increases the occurrence of false positives, which reduces the overall accuracy of sensitive user input detection.

Our layout extraction module focuses on removing the unused layouts, thus decreasing the occurrence of false positives.

Another challenge while extracting resources is identifying images that are used as labels for prompting user to enter sensitive information. For example, the widget used in Fig. 5 looks like a text view – “Credit Card” but when the XML file is analyzed, the text displayed is actually an image. These images would be discarded from the analysis if we only focus on extracting widgets from the layouts. We convert the images to text and then extract it as a part of our analysis, which no prior research has focused on.

The extraction process is done by disassembling the APK using APKTool and extract all layout files and resources [9]. Developers may compile and publish applications without cleaning them. After disassembling, we first remove any unused layouts using Android Lint tool [12]. Android Lint is a static analysis tool that optimizes the application by removing unused layouts. If the layout file or widget is present and not referenced anywhere, we consider it as an unused layout file or widget. For example, two credit card input widgets are present with different styles, but only one credit card input widget is used in the application. We then remove the widget/layout which is not used by the application during execution. This process helps us to address the issue of unused layouts and remove them from our analysis.

The next process is to extract the images used in the layouts and convert them into strings using optical character recognition (OCR) techniques. This helps to reduce the false negatives. We use pytesseract OCR tool to read the text embedded in images [13]. We take into consideration formats of images such as jpeg, png, gif, bmp, and tiff. We then create list of widgets used in the application and their resource identifiers. Once our resource identifiers are collected, we add one custom activity file, which renders all collected resource identifiers using setContentView(). We use APKTool to recompile the APK file with payload (list of resource identifiers) and custom activity. We then replace the old signature with the new one and align the new APK using APKSigner [10]. Then we run the newly generated APK on the AVD, which will extract coordinates, visibility attributes, hint, and text strings of all widgets. This information will be stored in a separate file with their identifier and used in the next module.

3.1.2. Filtering of sensitive labels

This module aims to resolve the semantics of the input data type and the descriptive text associated with it. We only focus on the sensitive terms and the associated descriptive text. It is crucial to identify the correct terminology and concept of the associated input data type. These can be single words or phrases. While mapping different terms to the same concept, there are two issues identified by previous literature [3]. These issues are: Synonymy - different terms may represent the same concept and Polysemy - same terms may represent different concept [21]. While extracting sensitive terms from the layouts, the concept of polysemy and synonymy helps to understand the context of the associated text. It is important to extract single words, phrases and resolve the concept that each of these sensitive terms relate to. This helps to reduce the ambiguity of words. We filter each of these sensitive terms (single words and phrases) by extracting them individually and comparing it with our dictionary of sensitive terms.

The process starts by lemmatizing all extracted text of labels, identifiers, hints and removal of stopwords. We then replace the special characters with the associated words. For example, if we find #, we replace it with a "number" word. The next step is to identify the sensitive terms by doing a word match with the predefined list and extracting it. This predefined list is developed by collecting inferences from prior work [4] and the Android guidelines for sensitive information [20]. If a match exists within the list, the label will be marked as "sensitive", "non-sensitive"

otherwise. There is a possibility that some of the labels are sensitive but could not be matched because of ambiguity. For example, address can be used for "postal address" and for" IP address". Hence, when address is matched, it is important to resolve the concept to which it refers and then labelling it as "sensitive" and "non-sensitive". To protect sensitive information, we must remove ambiguity from these words. Before we disambiguate the results it is important to understand the concept i.e. the different meanings of a particular term (polysemy or synonymy).

We use the word sense disambiguation method to resolve ambiguous words. Word sense disambiguation helps us to resolve meaning of a specific term. While performing the word match of individual terms/phrases, we perform the word sense disambiguation method to resolve the concept of the identified terms. We collect all the text within the layout to find the semantic group of the matched term/phrase. We also maintain a list of possible ambiguous words by exploring different Android applications, and those collected by prior work [3]. This helps to perform a word sense induction on the terms that are ambiguous but not identified as "sensitive". The concept of the ambiguous words is resolved as shown in Fig. 6.

Once we have the set of terms that relate to the ambiguous words on hand, we calculate the score for those participating terms . For example, the words such as city, postal code, and country will cause it to replace the address with "postal address" as it will have the highest score. We use the word only once for the calculation even if it appears multiple times. If two groups have the same score, then we calculate the score based on the number of times each word appears. If "city" appears three times, then we assign a score of 3 to it. If we get the same score for the two participating groups, then we keep both the groups. For example, if the postal address and IP address have the same score, it will show "postal address or IP address" after the resolution. We will mark it sensitive if at least one of the groups is from our sensitive word list.

Since we rely on the text collected from the layout, it does not disambiguate the word if the layout has only one label in which the ambiguous word is present. It is fair to assume to make that layout with a single widget will not have any ambiguous words present, since it will make it difficult for users to enter the required information in the input widget. This limitation can be addressed in future work by considering all layout files to resolve the ambiguous words.

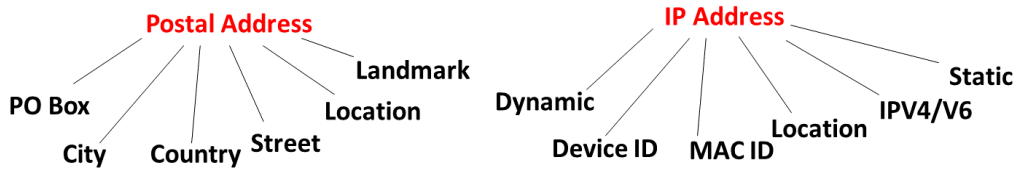


Figure 6. Sample ambiguous word groups

At the end, we remove all non-sensitive labels. However, we consider all the input widgets from the layouts without discarding any as it helps in the later stage to match the sensitive labels with all the extracted input widgets. Removing all non-sensitive labels helps to avoid incorrect pair-value matching, thus improving accuracy.

3.1.3. Mapping of sensitive labels with relevant user inputs

The main task of this module is to identify the input widgets to which the sensitive labels refer. Identifying the input widgets helps to retrieve the sensitive value entered by the user and thus train track it in the later stages of our framework. The first step is to identify the input widgets

that are related to the matched sensitive labels. To achieve this, we pass the generated sensitive labels from the semantic resolution module and find its corresponding widgets.

We devise two different algorithms to identify pairs of input widgets and sensitive labels. The two algorithms are as follows:

1. Generate candidate sets and vectors
2. Find the optimal set

Algorithm 1 - Generate candidate sets and vectors: The algorithm starts by generating candidiate sets of all the sensitive labels and input widgets. It then creates a set of vectors which represent the Euclidean distance within the GUI and direction from each sensitive label to all potential input widgets in the layout. The method calcSmallestVector() of UiRef [3] is used to generate the three smallest vectors that represent the potential distance of the sensitive term with the input widget. In this method, two vectors are generated from the two closest corners of the label to corresponding corners of the input widget. If the label is directly above, below, or on the sides of the input widget, then a third vector is created. A vector will not be considered as a candidate if the Euclidean distance is greater than a defined threshold, which is based on the screen size of the device being used to run the application. We obtain a set of potential candidates based on the Euclidean distance as the end result for this algorithm.

Algorithm 1 Algorithm for generating candidate sets and vectors

Input: All labels ($label$) and input widgets (iw)
Output: Candidate sets (cs)
$Initialisation : Candidates, v$
1: **if** ($i \neq$ "$sensitive$") **then**
2: Discard i
3: **else**
4: **for each** $i \in label$ **do**
5: **for each** $l \in iw$ **do**
6: $v \leftarrow calcSmallestVector(i, l)$
7: **if** ($v.distance < threshold$) **then**
8: $candidates[v].append(i, l)$
9: **end if**
10: **end for**
11: **end for**
12: **end if**
13: **return** $Candidates$

Figure 7. Algorithm 1 - Generate candidate sets and vectors

Algorithm 2 - Find the optimal set: The list of candidate sets obtained from Algorithm 1 will be passed as input to Algorithm 2, which will attempt to find the optimal set. It starts from the largest candidate set. It selects a label to input-widget pair if the input widget is directly above, below or on the side of the label. If there are multiple input widgets directly above, below or on the side of the labels, then it will choose the input widgets with the smallest distance. This process is repeated until it resolves all sensitive labels and finds the input widget pairs for it.

Algorithm 2 Algorithm to find the optimal set

Input: All labels ($label$), candidate sets (cs)
Output: Resolved labels ($ResolvedLabels$)
$Initialisation : ResolvedLabels, CountSensitive = 0$
1: **for** $i = 0$ to $size(labels)$ **do**
2: **if** $(i == "sensitive")$ **then**
3: $CountSensitive = CountSensitive + 1$
4: **end if**
5: **end for**
6: **while** $CountSensitive > 0$ **do**
7: $pairs \leftarrow GenCandidateSets(label, iw)$
8: $labels.remove(pairs.labels)$
9: $iw.remove(pairs.iw)$
10: $ResolvedLabels.append(pairs)$
11: $CountSensitive = CountSensitive - 1$
12: **end while**
13: **return** $Candidates$

Figure 8. Algorithm 2 – Find the optimal set

Towards the end of this module, we will have the sensitive inputs and its corresponding widgets which helps us to retrieve the value of sensitive input given by the user. Once we have received the sensitive labels and its corresponding widget pairs, we analyze this sensitive user information and train track it to identify security violations and possible data leak.

3.2. Identification and notification of any potential data leaks

This module is guided by the research questions " Is the sensitive information masked before storing or leaving the system?" and "Is the sensitive authentication data stored after authorization?". To answer the above research questions, we need to understand the technical approach to solving the problem and challenges associated with it. First, which requirements are considered for protecting the sensitive data in mobile applications? Second, how can these requirements be translated into program analysis tasks while minimizing false positives? Third, how can the violations associated with encryption/masking and authorization of sensitive data be programmatically identified?

In this module, we design our own static analysis method that captures the requirement of protecting user sensitive data in mobile applications and notify for potential security violations within the applications. Prior work in this domain [14] has focused on designing the static analysis tool but is limited to only payment applications. We consider identifying the potential data leaks within all the Android applications that seek for sensitive user information.

We identified the security requirements for sensitive user data in Android applications which are as follows:

- R1: Sharing of data securely across applications.
- R2: Storing of data safely within the application.
- R3: Limit the sensitive information storage and retention time.
- R4: Store only non-sensitive data in cache files.

- R5: Enforce secure communication.

Module-2 in (Fig 3) shows the high level overview of our proposed methodology in identifying the security violations of sensitive user data in mobile applications. We obtain the sensitive labels and widget pairs from Module-1 of the proposed framework and further train track this information. The next phase includes building a Data Dependency Graph (DDG) to identify the flow and context sensitive static program analysis on the .apk files. We use insights from Cardpliance [14] and choose Amandroid [22] as our application analysis tool. Amandroid allows us to perform static analysis and produce graphs upon which these tests can be performed. The sensitive input widgets received from Module-1 of the proposed framework will be passed to Amandroid that handles this information based on the tests designed and tracks the data flow of the GUI input.

We designed our own tests by taking reference from Cardpliance [14] for identifying violations and possible data leaks. These tests are built on top of Amandroid [22] which helps to perform the static analysis and generate the alerts for violations. The tests are as shown in Table. 3.

Table 3. Tests to identify violations and possible data leaks

Tests	**Identifies**	**Source (S)**	**Sink (K)**	**Required Methods (R)**
T1	Check if it is storing data or not	Activity.findViewById(ID)	DPM	-
T2	Check whether it is masking	Activity.findViewById(ID), URLConnection.getInputStream()	View.setText()	MM
T3	Storing non-obfuscated data	Activity.findViewById(ID)	DPM	OM
T4	Sharing Non-Obfuscated data	Activity.findViewById(ID)	Intent.putExtra(), SmsManager.sendTextMessage()	OM

Data Persistence Methods (DPM): java.io.OutputStream.write(), java.io.FileOutputSream. write(), java.io.Writer.write(), java.lang.System.out.println(), android.content.SharedPreferences Editor.putString(), android.util.Log.i(), android.util.Log.d()

Masking Methods (MM): java.lang.String.replace(), java.lang.String.substring(), java.lang. String.concat(), java.lang.StringBuilder.append()

Obfuscation Methods (OM): javax.crypto.Cipher.update(), javax.crypto.Cipher.updateAAD(), javax.crypto.Cipher.doFinal(), java.security.MessageDigest.digest(), java.security.MessageDigest .update()

These tests will be passed to our framework and tested on our dataset of Android applications. For the scope of this paper, we do not evaluate the results for this module of our framework.

3.3. Assigning weights to the obtained candidate sets

The goal of our framework is to inform users of the potential violations and data leaks of their sensitive information. However, it is also essential to inform the user of how confident we are about the result. Our framework might give false-positive or false-negative results in some scenarios, such as multiple ambiguous words in a single layout. For example, our framework identified that the application is asking for a credit card number because it found the "number"

keyword and a few other related terms. Hence, it will give a "very low" weight because the framework did not find any direct evidence that the input is for credit card information. If we find "credit card" or a similar term in the hint attribute of the input widget and the related label, then it will give a "very strong." weight as shown in Table. 4.

For the scope of this paper, we share the initial idea for assigning weights and do not discuss its methodology and evaluation results.

Table 4. Assigning weights to the obtained candidate sets

Scenario	**Weight**
If it finds sensitive user input using hint and label or ID	Very strong
If it finds sensitive user input using hint only	Strong
If it finds sensitive user input using both label and ID	Low
If it finds sensitive user input only label or only ID	Very low

4. EVALUATION

In this section, we evaluate the effectiveness of our framework with respect to the Module-1 of our framework (i.e., Identification of user inputs requesting sensitive information.)

To evaluate our extraction and optimization of resources module, we use some emulators and real time devices. We use 768 x 1280: xhdpi Nexus 4 emulator, running Android 10.0.

Our dataset consists of 60 Android applications from different categories of Google Play as shown in Table 5. We have created 10 of our own sample applications out of 60 applications and marked as "Other" to test different scenarios. To verify these applications' language, we use Python's langdectect module [3] for ensuring that applications with English descriptions are chosen.

Table 5. Applications for the evaluation

Application Category	**# of Apps**	**# of input**	**# of labels**	**# of images**	**# of hints**
Finance	10	128	203	41	88
Health & Fitness	10	103	143	34	76
Education	10	48	51	25	43
Entertainment & Travel	10	79	102	62	72
Shopping	10	112	214	84	79
Other	10	118	183	66	48
Total	**60**	**588**	**896**	**312**	**406**

We use UiRef and SUPOR [3, 4] as our baseline for comparison. We had no access to SUPOR's source or binary code and hence we re-implement their approach from the information gained by their publications. With UiRef, we were able to retrieve the loose code and modify it based on our comparison parameters.

Table 6. Extraction & optimization of resources

Parameters	**SUPOR**	**UiRef**	**Our Approach**
No. of widgets detected (Out of 2202)	1658	1790	2047
% of detection	75.29%	81.29%	92.96%
Unused widgets detected	N/A	N/A	103
Images to text conversion (Out of 312)	N/A	N/A	198 (63.46%)

Table 7. Filtering of sensitive labels

Parameters	**SUPOR**	**UiRef**	**Our Approach**
No. of ambiguous words detected (Out of 10)	N/A	9	9
% of ambiguous words detected	N/A	90%	90%
No. of labels (we also consider images as labels) detected (Out of 1208)	848 (70.19%)	876 (72.51%)	1074 (88.90%)
No. of sensitive labels detected	N/A	N/A	204

Table 8. Mapping of sensitive labels with relevant user inputs

Parameters	**SUPOR**	**UiRef**	**Our Approach**
No. of pairs detected - Only sensitive (Out of 206)	148	196	197
% of detection	78.72%	95.14%	96.11%
False positive (No. of sensitive pairs)	13	11	3
Accuracy in %	65.53%	89.80%	94.17%

The parameters based on which the comparison is done are chosen with the following aspects:

1. Number of elements detected has a direct correlation with the accuracy of detecting the correct input widgets. As greater no. of elements are detected, the initial dataset becomes more detailed. Our end goal is to detect correct pairs of the widgets wherein the user might enter the sensitive input. For this detection, all, or most of the elements of the layout must be detected. If the no. of elements detected is less than the actual elements present in the application, then the widget pairs generated will also be limited thus reducing accuracy of detection. As can be seen in Table. 6, our framework is detecting 12-18% more widgets than the prior work.
2. Detecting unused layouts is crucial as its presence has a direct impact on the ratio of false positive. Unused layout detection is one of our contributions and is not attempted by prior work. Hence the result comparison for this parameter is not applicable to the prior work. Unused widgets are around 5% of total widgets (as shown in Table. 6.
3. Converting images to text is one of our contributions and not attempted by prior work. Hence the result comparison for this parameter is not applicable to the prior work. We were able to convert 63.46% of the images to text. Some images did not have any text, so we could not be able to convert those. We found that accuracy can be improved by including icons, such as VISA or MasterCard, to detect the credit card field. We plan to include icons in our future work.
4. For the filtering of sensitive labels module, our main goal is to filter sensitive labels. For filtering sensitive labels, we first build a dictionary of possible sensitive terms and then do a text-matching. The same process is also carried out by prior work. However, the accuracy of this detection is based upon the ability of the algorithm to resolve ambiguous

words. Hence, the number of ambiguous words detected is crucial in determining the accuracy for this module. We had very few ambiguous words in the chosen applications. We were able to detect and resolve 90% of ambiguous words. UiRef also detected 90% of ambiguous words. Both (UiRef and our method) failed to resolve ambiguous words when only one layout and very few labels were available.

5. Prior work has focused on detecting all the labels – sensitive and non-sensitive. We first detect all labels and remove any non-sensitive labels from the list. We have included images as labels since some images contain text in them. As shown in Table. 7, we dected 18.5% - 21% more labels as compared to prior techniques.
6. Mapping sensitive labels with relevant user inputs modules generate the final candidate sets wherein the user inputs for sensitive labels are stored. The number of pairs detected does not have a correlation with the accuracy. However, the number of pairs detected will give an insight into the initial dataset generated and how many candidate sets the algorithm was able to generate. UiRef and our method detected almost the same number of pairs, as shown in Table. 8, but UiRef does not remove any unused layouts or widgets. So it is very important to identify the number of false positives and false negatives when determining the algorithm's overall accuracy. We found a 4.6% - 30% improvement in the accuracy from the prior techniques.

5. CONCLUSION AND FUTURE WORK

Our studies focus on resolving the semantics of sensitive user input through the GUI in Android applications and the train track of this information to identify potential violations and data leak. Prior studies have focused only on solving a part of this problem (either sensitive inputs or train track of information) and largely ignored the violations associated with sensitive information as a whole. In this paper, we have presented methodologies for identifying the input widgets that seek sensitive user information in Android applications and further train track of this information to identify potential data leaks and security violations. We also assign weights to the obtained results from our Module-1 which helps the users to prioritize the sharing of sensitive information through the application. This helps users to make an informed decision. No prior research in this field has attempted to provide weights to the obtained results. With our proposed framework, we are able to achieve an improvement in accuracy of 4.6 % - 30%. Our findings demonstrate that understanding the security and privacy concerns of user's sensitive inputs provide a unique vision to the mobile application's data security.

REFERENCES

[1] “Build a simple user interface,” Android, [Online]. Available: https://developer.android.com/training/basics/firstapp/building-ui. [Accessed 13 May 2022].

[2] “Layout resource,” Android Developers, [Online]. Available: https://developer.android.com/guide/topics/resources/layout-resource. [Accessed 13 May 2022].

[3] B. Andow, A. Acharya, D. Li, W. Enck, K. Singh and T. Xie, “Uiref: analysis of sensitive user inputs in android applications,” in Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2017.

[4] J. Huang, Z. Li, X. Xiao, Z. Wu, K. Lu, X. Zhang and G. Jiang, “SUPOR: Precise and scalable sensitive user input detection for android apps,” in 24th USENIX Security Symposium (USENIX Security 15), 2015.

[5] Y. Nan, M. Yang, Z. Yang, S. Zhou, G. Gu and X. Wang, “UIPicker: User-Input Privacy Identification in Mobile Applications,” in 24th USENIX Security Symposium (USENIX Security 15), 2015.

[6] V. Rastogi, Y. Chen and W. Enck, “Appsplayground: automatic security analysis of smartphone applications,” in Proceedings of the third ACM conference on Data and application security and privacy, 2013.

[7] X. Xiao, X. Wang, Z. Cao, H. Wang and P. Gao, "Iconintent: automatic identification of sensitive ui widgets based on icon classification for android app," in 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE), 2019.
[8] X. Pan, Y. Cao, X. Du, B. He, G. Fang, R. Shao and Y. Chen, "FlowCog: Context-aware Semantics Extraction and Analysis of Information Flow Leaks in Android Apps," 2018.
[9] Apktool, "Apktool - A tool for reverse engineering 3rd party, closed," [Online]. Available: https://ibotpeaches.github.io/Apktool/. [Accessed 14 May 2022].
[10] "Apksigner," Android Developers, [Online]. Available: https://developer.android.com/studio/command-line/apksigner. [Accessed 14 May 2022].
[11] "Play Protect," Google Developers, [Online]. Available: https://developers.google.com/android/play-protect. [Accessed 14 May 2022].
[12] "Improve your code with lint checks - Android Developers," Android Developers, [Online]. Available: https://developer.android.com/studio/write/lint. [Accessed 14 May 2022].
[13] M. Lee, "pytesseract: Python-tesseract is a python wrapper for Google's Tesseract-OCR," PyPI, [Online]. Available: https://pypi.org/project/pytesseract/. [Accessed 14 May 2022].
[14] S. Y. Mahmud, A. Acharya, B. Andow, W. Enck and B. Reaves, "Cardpliance:PCI DSS Compliance of Android Applications," in 29th USENIX Security Symposium (USENIX Security 20), 2020.
[15] S. Arzt, S. Rasthofer, C. Fritz, E. Bodden, A. Bartel, J. Klein, Y. Le Traon, D. Octeau and P. McDaniel, "Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps," Acm Sigplan Notices, vol. 49, pp. 259-269, 2014.
[16] "Official PCI Security Standards Council Site - Verify PCI Compliance, Download Data Security and Credit Card Security Standards.," [Online]. Available: https://www.pcisecuritystandards.org. [Accessed 14 May 2022].
[17] T. Shah and S. Sampalli, "Efficient LFSR Based Distance Bounding Protocol for Contactless EMV Payments," in Proceedings of the Future Technologies Conference, 2018.
[18] "Service fees - Play Console Help.," Android Developers, [Online]. Available: https://support.google.com/googleplay/android-developer/answer/112622?hl=en. [Accessed 14 May 2022].
[19] B. Uscilowski, "Mobile adware and malware analysis," in Symantec Corp, 2013.
[20] "User Data - Play Console Help.," [Online]. Available: https://support.google.com/googleplay/android-developer/answer/10144311. [Accessed 14 May 2022].
[21] D. Glynn and J. A. Robinson, "Corpus methods for semantics: Quantitative studies in polysemy and synonymy," 2014.
[22] F. Wei, S. Roy and X. Ou, "Amandroid: A precise and general inter-component data flow analysis framework for security vetting of android apps," ACM Transactions on Privacy and Security (TOPS), vol. 21, pp. 1-32, 108.

SURVEY OF SECURE NETWORK PROTOCOLS: UNITED STATES RELATED DOMAINS

DeJean Dunbar, Patrick Hill and Yu-Ju Lin

Department of Computer Science,
Charleston Southern University, North Charleston, South Carolina

ABSTRACT

Over time, the HTTP Protocol has undergone significant evolution. HTTP was the internet's foundation for data communication. When network security threats became prevalent, HTTPS became a widely accepted technology for assisting in a domain's defense. HTTPS supported two security protocols: secure socket layer (SSL) and transport layer security (TLS). Additionally, the HTTP Strict Transport Security (HSTS) protocol was included to strengthen the HTTPS protocol. Numerous cyber-attacks occurred in the United States, and many of these attacks could have been avoided simply by implementing domains with the most up-to-date HTTP security mechanisms. This study seeks to accomplish two objectives: 1. Determine the degree to which US-related domains are configured optimally for HTTP security protocol setup; 2. Create a generic scoring system for a domain's network security based on the following factors: SSL version, TLS version, and presence of HSTS to easily determine where a domain stands. We found through our analysis and scoring system incorporation that US-related domains showed a positive trend for secure network protocol setup, but there is still room for improvement. In order to safeguard unwanted cyber-attacks, current HTTP domains need to be extensively investigated to identify if they possess security-related components. Due to the infrequent occurrence of HSTS in the evaluated domains, the computer science community necessitates further HSTS education.

KEYWORDS

Network Protocols, HTTP Strict Transport Security, scoring benchmark, domain analysis, survey.

1. INTRODUCTION

HTTP was a pinnacle basis in defining how information was transmitted across a network during this technical era of information technology. Since 1990, the world wide web has employed the HTTP Protocol as a stateless application-level protocol for hypermedia information systems. HTTP/0.9, the initial implementation of HTTP, was designed for raw data delivery across the Internet. As new versions of HTTP were released, no security procedures for the transport of raw data were implemented [1]. Hackers can access health information, government information, and personal information. HTTPS protocol was introduced to address this significant security vulnerability.

HTTPS is a direct extension of the HTTP Protocol introduced by Netscape Communications. There were several security implementation versions of the HTTPS protocol which were SSL and TLS. SSL was first proposed in the middle of 1994 by Netscape Communications with its highest version implementation being version 3. TLS was proposed by the Internet Engineering Task

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 91-101, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121207

Force with its highest version being version 1.3. The overall goal of HTTPS was to provide security features for HTTP such as encipherment, digital signature mechanisms, data integrity, authentication exchange mechanisms, and notarization mechanisms [2]. With these additional security measures incorporated, another web security standard extension for HTTPS was introduced known as HTTP Strict Transport Security (HSTS).

The Internet Engineering task force (IEFT) proposed HSTS in 2012 and defined it as a security mechanism that restricts website access to only secure connections. This feature guards against bootstrap man in the middle (MITM) attacks. Additionally, HSTS provides security by converting a URI reference to a secure URI reference [3]. With these additional security benefits that can be added to HTTPS, the HSTS security mechanism helps a domain's network security strength to be even stronger.

This research will be structured with a preliminary finding section which will describe the hardware involved, techniques for gathering the US related domains, the database used, the scanner used for domain protocol information and python parsers used. The next section will discuss the research plan which includes considerations made in the research gathering, the proposed scoring system mechanics and research results. The next section will discuss the conclusions based off the research results. Finally, an acknowledgements section will highlight the key individuals that contributed to the overall success of this paper.

2. PRELIMINARY FINDING

In this section we will discuss the hardware involved in the experiment. The method of gathering the US-related domains will follow. The database used for domains will also be briefly discussed. Next, we will discuss the scanning technique used for the analysis of the domain's SSL/TLS versions. Finally, we will discuss the Python scripts involved that helped with additional domain analysis gathering and updates to the database.

2.1. Hardware Used

All the tools used for this research were all run in a virtual machine using Ubuntu. The computer model is an Inspiron 16 7610 running on a Windows 11 Pro x64 operating system. The computer has an i7 core and 32GB RAM. It was necessary to utilize a higher core to better utilize throughput for running multiple instances of the scanning program and the python scripts.

2.2. Techniques for Gathering US Based Domains

For this study, domains ending in ".us", ".gov", and ".edu" were grouped together. The number of ".us" domains gathered were 1814204. The number of "gov" domains gathered were 5854. The number of ".gov" domains gathered were 7671. The grand total of domains gathered were 1827729.

A ".us" zone file request was sent to the registry site ABOUT and was later redirected to GODADDY [4]. A zone file contains a list of all domains that have been registered. Following the approval of the request, the zone file was provided. A new zone file is created every day with the year, month, and day due to new/existing domains being updated. The zone file we chose for parsing was from March 26, 2022. Although domain names were included in the zone file, the file contained other information deemed unimportant. On a DELL laptop running Ubuntu virtually, a series of commands shown in Figure 1 was run to extract only the domain names from the zone file. The extracted data was saved as text files.

```
$ awk '{print $1}' us.zone > domains-only.txt
$ sort- u domains-only.txt --output domains-unique.txt
$ LC_ALL=C grep '^[A-Z0-9\-]*$' domains-unique.txt > domains.txt
```

Figure 1. List of commands used to parse through ".us" zone file

The same technique was attempted for ".edu" domains, however the organization EDUCAUSE's cooperative agreement rules would not allow them to give us access to their zone file. To accommodate the lack of a zone file, we decided to utilize two outside sources for the gathering of ".edu" domains. The first source was from a GitHub repository which provided ".edu" domains of universities from around the world [5]. The list was filtered to only the United States. The next source of ".edu" domains came from Common Crawl, a reputable web crawling service [6]. Common Crawl provided a server for their data to be queried via Amazon's AWS service, Athena [7]. A query was run against Athena to collect domain names ending in ".edu" in the year 2021.

For gathering ".gov" domains there is an actual government site that list all the currently registered government websites in the United States [8]. The list of .gov domains gathered were stored in a csv file.

2.3. Database Used

We used MYSQL database to maintain a consistent repository for the domain information used in this research. This database stores domain names, SSL, and TLS versions, HTTP status, HTTPS status, and HSTS status to aid in the analysis results section. MYSQL Workbench, a database application, was used to import all the domains that were gathered and stored as csv files in the previous section into the MYSQL database.

2.4. Scanner for SSL/TLS Identification

We used a well-documented tool called SSLSCAN [9] to scan the US-related domains for SSL/TLS protocol versions. This command-line tool accepts a file of domain names as input and returns in XML format the SSL/TLS protocol versions for each domain, if any. The domain names were obtained using a query against the MYSQL database and then converted to csv files to serve as the input for the SSLSCAN tool. Following that, ten instances of the SSLSCAN with the csv files were run to pipeline the scanning process. The total time to scan was thirteen days.

2.5. Python Parsers

To transfer data from SSLSCAN's XML output files to a MYSQL database, a Python application parsing the XML file output was written. An update statement within the script was executed for each domain parsed to keep the database in sync with the SSL/TLS protocol information from the XML files.

To determine whether HSTS was present in the domains collected, another Python script was written to request the domain's header information. This script took as input a csv of domain names. We queried the MYSQL database for HTTPS domain names that had successful scans with the SSLCANNER. As with the previous Python script, this one updated the MYSQL database in response to the presence of HSTS for each domain. The program execution took five days to process the domains.

3. METHODS (RESEARCH PLAN)

3.1. Considerations

Before going over the results of the domains it is necessary to outline some decisions made before and during the analysis. There were originally 1839452 domains gathered and transferred to the MYSQL database. 11723 of the domains did not have the correct extension stemming from the ".us" zone file provided. For example, there was a domain named "100plusus". It was ambiguous if the name should have been "100plusus.us" or "100plus.us". These 11723 domains were removed from the database to remove this ambiguity. Another consideration made was during the SSLSCANNER application being ran on the domains. There were errors logged in the XML file for each domain that encountered an issue. These issues ranged mainly from refused connections from the domains or timeouts. The total number of usable domains after the scanning was 658500. Due to the nature of scanning domains, we ensured that the results are only stored in a private repository in GitHub to ensure best ethical practices.

3.2. Proposed Scoring System

This section will describe the proposed scoring system which takes into account the protocol version and whether HSTS is being used. Additionally, this section will present some examples to properly illustrate the scoring system in practice with given domain configurations.

The scoring system grading will be in the numerical range from 0 – 100. We will consider a score of a 70 to be passing while anything lower is a failure. One assumption made for this scoring system is that if a domain supports multiple HTTPS protocol versions, then the highest HTTPS protocol version will only be considered for the domain's overall score.

The proposed scoring system is split into three tiers. Tier 1 includes domains that support HTTP protocol. These domains will automatically receive a score of 0 due to no security being available for the protocol. Tier 2 consists of the HTTPS protocol with the SSL version variations which includes SSL 2.0 and SSL 3.0. SSL 2.0 will be given a starting score of 5. The path for potential updates can be described as seen in Figure 2 where each transition to the next state is awarded 5 points. This pattern will continue until the highest SSL protocol version with HSTS is reached which is awarded 20 points. Tier 3 has the same principle as Tier 2 using TLS version variations, but TLS protocol 1.0 will start off with 30 points. Each transition to the next state for TLS versions will be awarded 10 points.

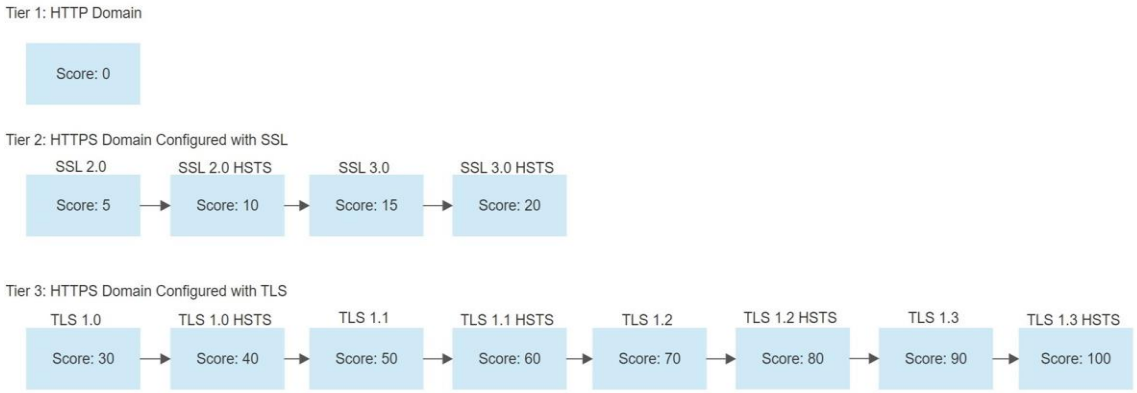


Figure 2. Three tier process for proposed scoring system

A passing score for this scoring system is when TLS 1.2 protocol is used which is awarded 70 points. The reason for this decision is due to RFC officially announcing the deprecation of protocols SSL 2.0, SSL 3.0, TLS 1.0 and TLS 1.1. The best-case scenario is that a domain contains TLS 1.3 HSTS which will be a score of 100.

We will use three examples to illustrate the scoring system. Example one is relatively simple with Figure 3 having the current configuration of just HTTP and because of this we will give a failing score of 0. Example two in Figure 4 has SSL 3.0 with HSTS. Since we are in Tier 2 category, we start off with a base score of 5. Since we transitioned three states in order to reach SSL 3.0 with HSTS we add an additional 5 points per state leading to a total score of 20 points. Example three shows Figure 5 with the current configuration of TLS 1.1 with HSTS. Notice that this figure shows the domain supporting earlier protocol versions. We ignore the earlier protocol versions and only consider the highest. Since we are in the Tier 3 category, we start off with a base score of 30. Since we have transitioned three states in order to reach TLS 1.1 with HSTS, we add an additional 10 points per state leading to a total score of 60 points.

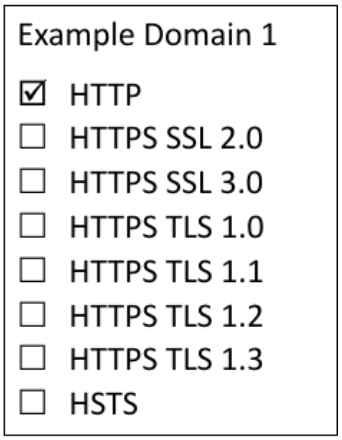


Figure 3. Scoring system example domain with HTTP configuration

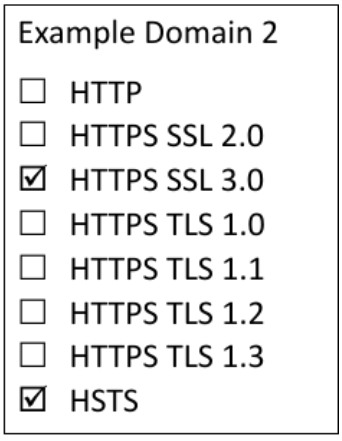


Figure 4. Scoring system example domain with HTTPS SSL 3.0 and HSTS header

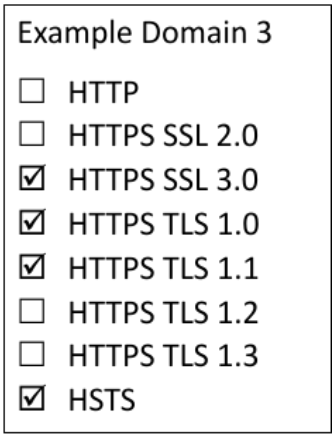


Figure 5. Scoring system example domain with HTTPS SSL 3.0, TLS 1.0, TLS 1.1, and TLS 1.2 and HSTS header.

3.3. Experiment Results/Analysis

3.3.1. HTTP vs HTTPS

Of the 658500 domains, 38% had only HTTP protocol support while 62% had only HTTPS protocol support shown in Figure 6. The noticeable percentage of US-related domains being HTTP even in modern times is alarming. A possible explanation is that the nature of the domains that contained HTTP protocol does not transmit sensitive information over the network at all which asserts that there is no need for HTTPS. This assumption is later disproven when a random sample of the analyzed HTTP related websites were chosen for investigation. We found there were several instances of .edu domains that were HTTP containing login features which is not good practice.

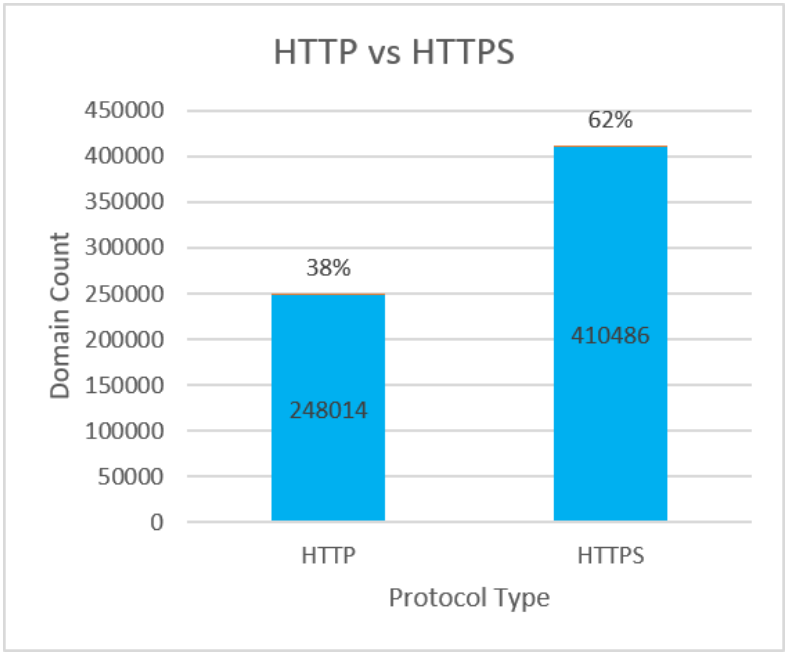


Figure 6. A bar graph visualizing the number of domains scanned that were HTTP protocol or HTTPS protocol.

3.3.2. HTTPS: SSL and TLS

We next further split the HTTPS into its individual components SSL2.0, SSL3.0, TLS 1.0, TLS 1.1, TLS 1.2, and TLS 1.3 for analysis. It is important to note that domains can have more than one version of HTTPS enabled. Of the 658500 domains analyzed, under 1% of the domains were configured with SSL 2.0 and SSL 3.0; 23% of the domains contained TLS 1.0; 24% contained TLS 1.1; 62% contained TLS 1.2; 30% contained TLS 1.3. (See Figure 7)

When analysing the HTTPS protocol versions, the SSL version 2.0 served as the minimum for the number of domains. This met expectations due to it having been deprecated since 2011 by RFC 6176. Additionally, SSL 2.0 was released over two decades ago with vulnerabilities present in them that would create a high need to transition to the TLS protocol. [10]

When analysing the HTTPS protocol versions, TLS 1.2 served as the maximum for the number of domains. This trend is furthermore supported by Qualys SSL Labs. Qualys SSL Lab's past history of domain scans from January 2021 to October 2021 revealed TLS 1.2 served as the maximum for domain usage [11].

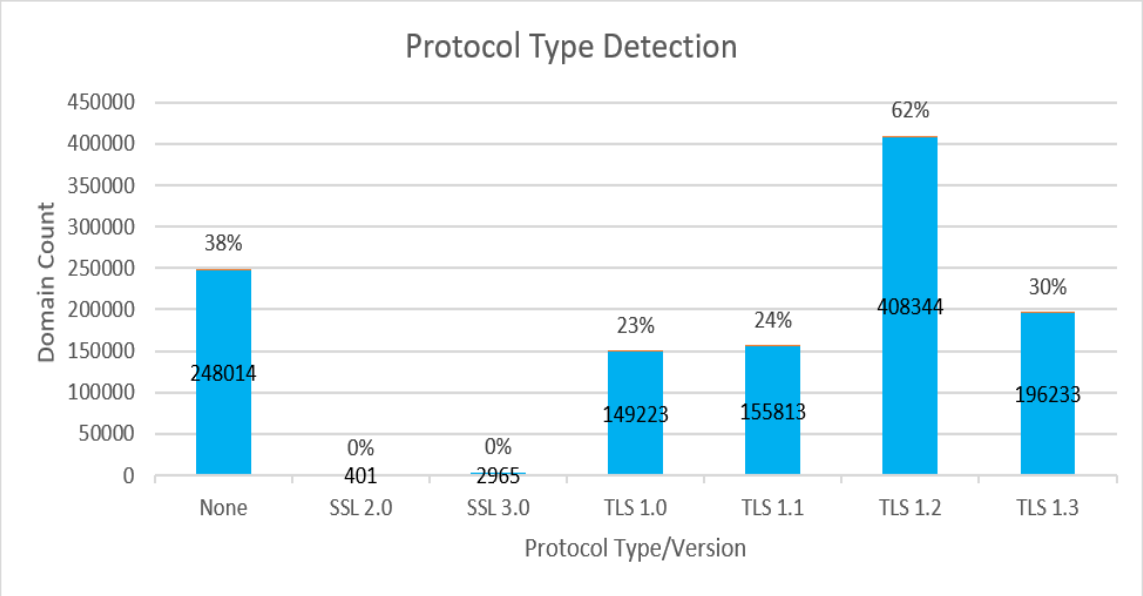


Figure 7. A bar graph visualizing HTTP domain counts along with a more broken-down analysis of HTTPS protocols with their varying versions.

3.3.3. HSTS

Finally, for HSTS detection of the 658500 domains analyzed, 5% of the domains contained HSTS headers while the other 95% contained no HSTS headers. (See Figure 8). The trend of the low amount of HSTS being detected is supported by other works conducted in the past. One research focused on government websites had a similar trend in where only 2.86% percent of those websites supported HSTS [12]. Another research paper on HSTS deployment survey conducted in 2013 revealed a similar trend. Of the 1 million websites analyzed only 277 contained HSTS headers [13]. Additional research work conducted in 2018 focused on analysing the adoption of security headers in HTTP found that from the 1 million websites scanned that only 5.41% used HSTS [14]. Another research paper conducted in July 2018 focused on analysing HSTS found that from the 1 million websites scanned that only 5.35% used HSTS [15].

We believe the main culprit for why HSTS headers are low in presence is due to users not being educated or informed about HSTS. More importantly, IT professionals or computer scientists are the community of individuals who would configure the HSTS headers for domains. To explore this theory, a survey was conducted among computer science professionals and IT professionals from a Department of Energy owned facility called Savannah River Nuclear Solutions. The results (see Figure 9) found that 80% of the surveyed individuals did not know what HSTS is.

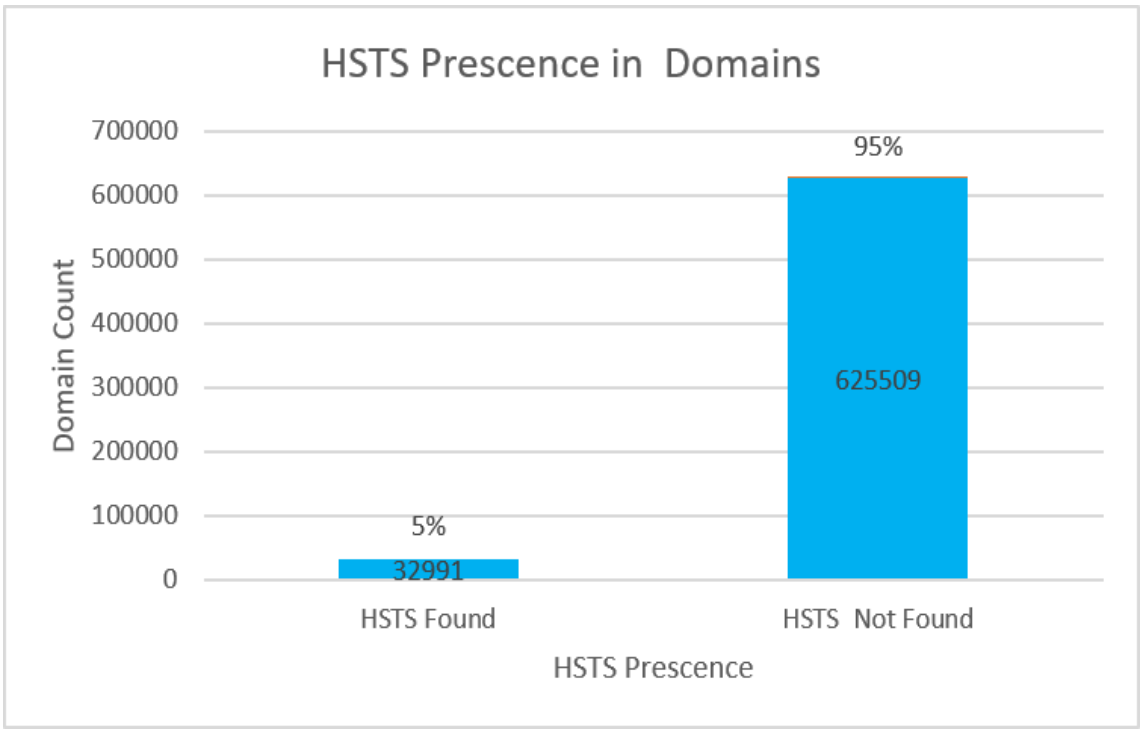


Figure 8. A bar graph visualizing the number of HSTS headers detected versus the number not detected for the scanned domains

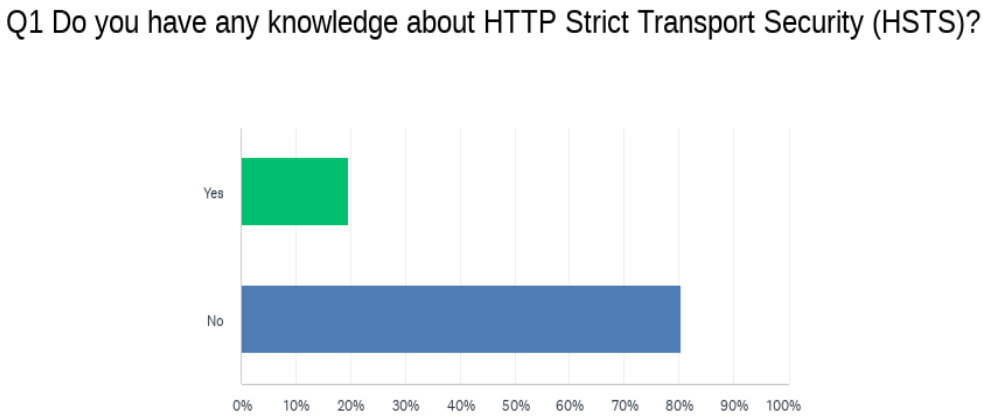


Figure 9. A bar graph from Survey Monkey visualizing the number of participant's knowledge of HSTS.

3.3.4. Scoring System Incorporation

We incorporated the proposed scoring system as part of the data analysis. The results were split into two groups. The first group included domains that scored a passing result of 70 or higher and the second group was domains that scored below a 70 which is considered a failure. Figure 10 demonstrates that based on the scoring system rules, 62% of the domains were given a passing score of 70 while 38% percent failed the scoring system. The simple benchmark scoring system can be used as a preliminary report to help establish a focus on acceptable configurations versus unacceptable configurations.

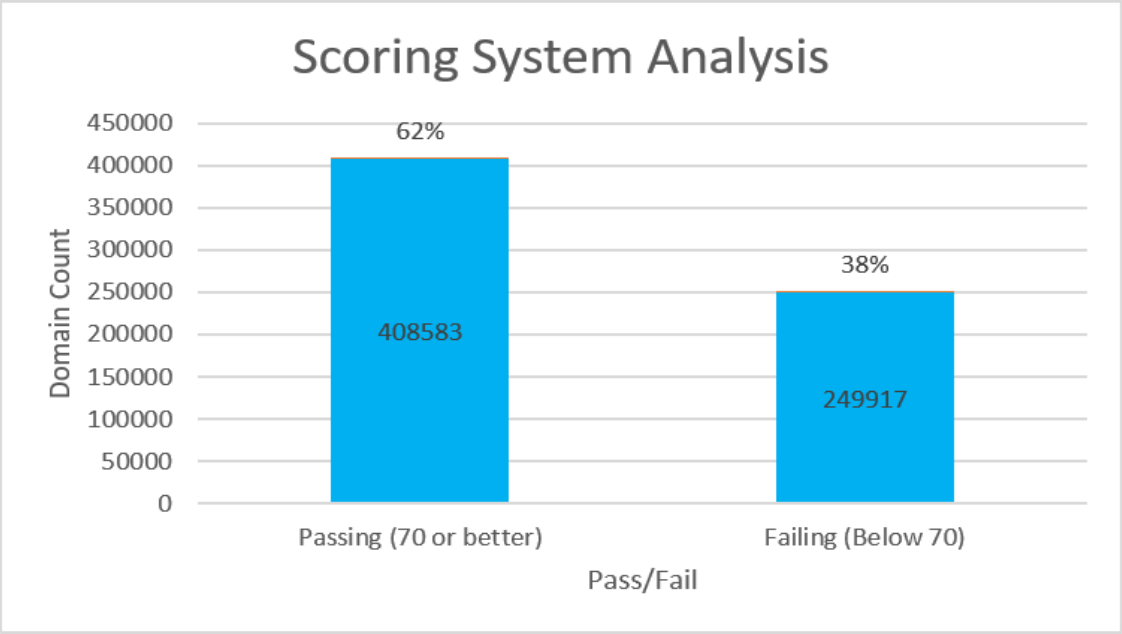


Figure 10. A bar graph visualizing the scoring system applied on the scanned domains.

4. CONCLUSIONS

This study has revealed that US-related domains are not up to date with the latest protocol security and HSTS incorporation. By creating a simple scoring system with respect to RFC's most recent deprecations, a general sense of where US based domains stand numerically was easily noticeable. The scoring system rules can be applied to any domain to get a general sense of where their network protocol and HSTS presence stands. This study has also found that HSTS usage in US based domains is low and that the lack of awareness is one of the contributing factors behind it. It is important to incorporate HSTS knowledge in the workplace for any team that works with configuring network related security which would include education for all stakeholders.

Another revelation in this study was that HTTP related sites are still prevalent. The theory of these websites having HTTP due to not needing security was disproven by analysing several HTTP websites in the study that had security related features such as login. One path forward to mitigate this issue is for registrars that give out domains to prompt the user what they would be using the website for in order to offer the user the option of having the most up to date protocol security.

It has also been revealed in the study that a high percentage of US-related domains have enabled lower versions of the HTTPS protocol which needs to be corrected to minimize downgrade related attacks. The path forward would be to notify the end user that owns the domain of this issue, however, due to ethical concerns, the domain names have been kept confidential to ensure privacy. It is therefore more effective to focus on the registrars who sell the domains pre-emptively. EDUCAUSE, GODADDY, and DOTGOV are the companies in this study that handle US-related domain extensions ".edu",".us", and ".gov" respectively.

Future work for this research includes increasing the accuracy of gathering ".edu" domains since the zone file could not be acquired. EDUCAUSE stated that they will only give zone file if they can be positively benefited. If a future researcher partners with other major universities and sends another zone file request to EDUCAUSE, then the chances of them providing the zone file will increase the coverage of ".edu" domains. An additional future work for this research is to perform another assessment of US-related domains in the next coming years and use this research as reference to show if the trend has improved or not. This study will aid in the overall understanding of US-related domains and provides a compelling argument that the organizations in charge of facilitating these domains are made aware of that there is susceptibility in many of the websites.

ACKNOWLEDGEMENTS

The assistance provided by my professor Yu-Ju Lin was greatly appreciated. He provided me a successful path of how to organize this thesis and offered valuable input during its development. Additionally, I would like to thank Patrick Hill for allowing me to build upon his research work and utilize in this paper. I would also like to thank Raymond Wilcauskas, a former employee contracted by the Department of Energy who agreed to be part of the reviewing committee for my thesis.

REFERENCES

[1] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1," www.rfc-editor.org, Jan. 1997, doi: 10.17487/RFC2068.
[2] R. Oppliger, SSL and Tls: Theory and Practice. Norwood, Ma: Artech House, 2016.
[3] J. Hodges, C. Jackson, and A. Barth, "HTTP Strict Transport Security (HSTS)," Nov. 2012, doi: 10.17487/rfc6797.
[4] "Domain Name Registry Services from GoDaddy Registry," registry.godaddy. https://registry.godaddy/ (accessed Apr. 21, 2022).
[5] "Hipo/university-domains-list," GitHub, Sep. 02, 2020. https://github.com/Hipo/university-domains-list
[6] "Common Crawl," Common Crawl. https://commoncrawl.org/
[7] "Amazon Athena - Serverless Interactive Query Service - Amazon Web Services," Amazon Web Services, Inc. https://aws.amazon.com/athena/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc
[8] "Data | .gov," home.dotgov.gov. https://home.dotgov.gov/data/ (accessed Apr. 21, 2022).
[9] rbsec, "sslscan2," GitHub, Apr. 21, 2022. https://github.com/rbsec/sslscan/ (accessed Apr. 21, 2022).
[10] N. Sullivan, "Why TLS 1.3 isn't in browsers yet," The Cloudflare Blog, Dec. 26, 2017. https://blog.cloudflare.com/why-tls-1-3-isnt-in-browsers-yet/#:~:text=The%20reductive%20answer%20to%20why%20TLS%201.3%20hasn%E2%80%99t
[11] "Qualys SSL Labs - SSL Pulse," www.ssllabs.com. https://www.ssllabs.com/ssl-pulse/
[12] P. Hill and Y.-J. Lin, "Evaluation of Trust Worthiness of State and County Government Websites." Accessed: Apr. 21, 2022. [Online]. Available: http://gator3168.temp.domains/~patriill/wp-content/uploads/2021/05/SAM21-1.pdf
[13] L. Garron, A. Dropbox, and D. Boneh, "The State of HSTS Deployment: A Survey and Common Pitfalls." Accessed: May 13, 2022. [Online]. Available: https://garron.net/crypto/hsts/hsts-2013.pdf

[14] W. J. Buchanan, S. Helme, and A. Woodward, "Analysis of the adoption of security headers in HTTP," IET Information Security, vol. 12, no. 2, pp. 118–126, Mar. 2018, doi: 10.1049/iet-ifs.2016.0621.
[15] S. De los Santos and J. Torres, "Analysing HSTS and HPKP implementation in both browsers and servers," IET Information Security, vol. 12, no. 4, pp. 275–284, Jul. 2018, doi: 10.1049/iet-ifs.2017.0030.

AUTHORS

DeJean Dunbar earned his B.S in Computer Science from Charleston Southern University in 2017. He is currently pursing his masters degree at Charleston Southern University. DeJean's interests in computer science includes automation-based systems, database management systems and networking security.

IMPLEMENTING RISK SCORE TO PROTECT FROM ANDROID PATTERN LOCK ATTACKS

Yasir Al-Qaraghuli and Caroline Hillier

Department of Computer Science, University of Guelph, Ontario, Canada

ABSTRACT

Cyberattacks on Android devices have increased in frequency and commonly occur in physical settings with shoulder surfing and brute-force attacks. These attacks are most common with devices secured by the pattern lock mechanism. This work aims to investigate the various methods that increase the security of Android lock patterns. Research showed that these pattern lock screens are especially vulnerable due to users employing a set of common lock patterns. We propose a pattern-matching algorithm that recognizes these common lock patterns and increases the Risk Score if these passcodes are attempted. The blocking of common passcodes, and identification during the unlocking, reduces the risk of the aforementioned threats to device security. The algorithm we implemented succeeds in deterring users from configuring their devices with commonly used patterns. Overall, our algorithm achieves advanced security compared to current systems by detecting unusual inputs and locking the device when suspicious activity is detected. Our test results show 80% satisfaction from human test subjects when settings the passcode. The algorithm eliminates the use of commonly used patterns and 79% acceptance using our proposed algorithm and blocks access to the device depending on the accuracy score. The proposed algorithm shows remarkable success with limiting brute-force attacks as it proves effective in denying common passcodes.

KEYWORDS

Android device, Lock pattern, Brute-force, Shoulder-surfing, Pattern Recognition.

1. INTRODUCTION

Over 2.5 billion users worldwide use Android mobile devices [1]. These devices hold a great deal of valuable personal information such as addresses, passwords, and personal documents [2]. Because of the value of this data, there can be severe repercussions if a device's security is compromised.

Research has shown that users created passcodes to fall into a small set of reoccurring patterns [2]. Complex lock patterns can be hard to memorize, and certain nodes are easier to reach while holding a device [3][4]. There is a necessary balance for users to create a unique pattern that they can remember and repeat frequently, but there will always be some human error.

Many ways that forgetful users or attackers can bypass the lock screen, though these processes are time consuming or result in data loss [5]. Because of these long processes, it is ideal for attackers to directly attack the lock screen.

Brute-force attacks occur when an attacker attempts passwords on a device with the intent to gain unauthorized access [6]. Shoulder surfing attacks occur when a criminal physically positions

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 103-112, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121208

them self in a way where they can observe device passcode entry of their target [7], as shown in (Figure 1). We found that these threats are an exceptionally severe problem with Android devices as pattern locks are amazingly easy to a brute-force attack (Figure 2). Prior research showed that six dot-length Android pattern attacks with a single shoulder surfing observation had a 65% success attack rate, which increased to 79.9% with multiple observations by the attacker [8]. Our work will investigate several ways of identifying attacks and blocking them. We propose a *Risk Score* factor that evaluates the user based on their input and determines if the entry is a brute force attack or if the user made an input error. We have implemented this risk percentage system to allow the user to choose a risk allowance on their device. Once the user reaches the predetermined risk allowance with unsuccessful pattern attempts, they would be locked out and require additional verification.

Figure 1. Diagram of shoulder surfing

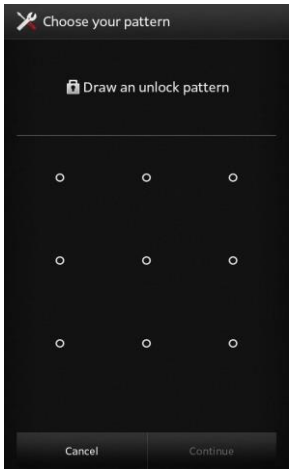


Figure 2. The famous android lock pattern

2. RELATED WORK

There has been some research regarding shoulder surfing, but the work typically focuses on the prevention, not the actual attack performance [8].

Previous studies have shown that having a six-digit pin over a four-digit (similar complexity) decreases shoulder surfing attack success from 64% to 11% [8]. The way that the user holds their phone and angle of observation also alter the attacker's success rate [8].

Other methods have utilized biometric trends in tools for the security of mobile devices. Instead of a passcode, researchers recorded users' physical traits to secure the device [3]. The biometric characteristics included keystrokes and touch dynamics [3].

Utilizing the sensors that track these characteristics can also assist in preventing attacker success, though it does lead to limitations such as having a friend use the device.

Many security professionals have aimed to investigate the vulnerabilities surrounding lock pattern systems. A computer vision algorithm showed that the analysis of fingertip motions often leads to successful shoulder surfing to brute force attacks [9]. Research showed that an attacker using fingertip analysis successfully guessed passwords in five attempts, with a 95% success rate [9].

The current methods for device locking are not sufficient and remain vulnerable [10].

Researchers have tested using an efficacy meter to enforce stronger user-created passwords [9]. The study compared the security of pattern selections of users seeing the meter and those that did not. The passwords of the participants that did not see the meter could be guessed in 16 guesses, while those that did see the meter needed 48 attempts to guess correctly [11].

Researchers have studied the implementation of blocklists to deter users from choosing the most common passcodes [1]. They found that the block list caused frustration and additional time from the user but overall reduced attack success rate [1]. 28 Android unlock patterns were identified as the most frequently used [1]. The authors describe the frequently used patterns as unsafe because they will likely be used in a brute force attack [1].

To better Android device security, we propose a system that identifies common patterns and lists them as a security risk. Our system both denies the user from using these common patterns and monitors for an attacker trying to access the system by entering these patterns. Our work aims to provide further protection for Android devices that implement a lock pattern system. Because we are also developing solutions for the security of Android lock patterns, we will use the database established in this work [1].

3. PROPOSED SOLUTION

Our system utilizes a risk scoring system to determine if the device should lock after each pattern lock attempt. The owner will set a lock pattern during setup, which is acceptable if it is not a common pattern [1]. The user will also establish their preferred risk acceptance level after an explanation of what the acceptance level means.

Our system uses a pattern matching algorithm to identify whether an inputted pattern matches a common pattern.

Since the user cannot set a common pattern as their passcode, the attempt of these patterns is a high-risk action.

Pattern matching also compares the current unlock attempt to the previous entry. Sequential unlock attempts with high variation signify that the user attempting to unlock the device does not know the passcode and is trying various patterns to unlock the phone. This behaviour indicates that the user unlocking the device is likely an attacker, thus increasing the *Risk Score*.

Actions deemed high risk will significantly raise the *Risk Score*, while actions that are low risk will marginally raise the *Risk Score*.

This ranking means that a series of minimal risk unlock attempts will gradually lock the user out, but a series of high-risk unlock attempts will lock the device more quickly.

When the device becomes locked, it may be subject to a time penalty or connection to a home device to unlock. The time penalty security measure has low efficacy if the attacker has the device and can resume their brute force after waiting a small amount of time. Increasing the lockout time may be effective against attackers but inconvenient for users. Similarly, requiring connection to a home device would be most effective as the attacker would not have access to the device but would be inconvenient for a user not near their home.

A solution for the locked-out device (Figure 3) is to have a recovery email tied to the account, shown in Figure 4. Device owners would receive the recovery email immediately after device lockout. This email system gives legitimate users an unlock link that lets them continue to unlock their devices. Attackers would not have access to the email link, meaning the device would remain locked. If a user is not currently using the device, this email could indicate that their device is being attacked. The recovery feature was implemented following user testing. In testing, we received feedback regarding "How long would [a user] have to wait until [they] can try again?" This email system complements our pattern detection system as an extra layer of security to keep brute-force attackers out.

Our algorithm provides improved security compared to previous research done on Android lock patterns. Creating a cohesive interface that blocks the use of commonly used patterns [1] and uses a *Risk Score* where users can enforce their security preference, provides strengthen security for the device locking system. Uniting the proposed security concepts allows for insight to real world integration.

Figure 3. Locked out user after multiple attempts

Figure 4. Email authentication to gain access back to device

4. METHODOLOGIES

4.1. Development

We chose to use Python to create both our front-end interface and our back-end code library. To create the Graphical User Interface, the *tkinter* interface package was used. The *smtplib* library was used to send out verification codes to user recovery email addresses.

4.1.1. Back-End Development

The 28 common patterns defined in [1] serve as the database of common passwords for this project. To allow for machine identification of the android patterns, we converted them into arrays by assigning each node with a number and then representing patterns as arrays. Pattern nodes are represented with the digits one through nine, with one being the node in the top left and nine in the bottom right. Zeroes padded the length of the array. Using these arrays, we could then perform pattern-matching by analysing the positions of node values and calculating how similar the arrangements were. We then created a library of back-end functions that could be called from our front-end code.

4.2. Pattern-Matching Algorithm

To perform the pattern-matching, we compared patterns based on their similarity to the identified common passcodes identified in [1]. Pattern location will be represented as numbers in a three-by-three grid, which can be seen in figure 6 [12]. Each pattern in the input would contribute to the *Risk Score*. If the node value appears in the same position as the correct pattern, one point is added to the *Risk Score*. If the node value is in the correct pattern but in different locations, half a point is added to the *Risk Score*. If an input is not in the correct passcode array it was being compared to, zero points will be added to the *Risk Score*. The score is then divided by the length of the pattern.

For example, if the password pattern is set to [2,4,6,1], the brute force attempts would have varying *Risk Score*s, as seen in Figure 5.

[2,8,9,1] = 2 Points (1,0,0,1) = 2/4 = 50%
[1,6,4,2] = 2 Points (0.5,0.5,0.5,0.5) = 2/4 = 50%
[7,8,9,3] = 0 Points (0,0,0,0) = 0/4 = 0%
[1,4,6,2] = 3 Points (0.5,1,1,0.5) = 3/4 = 75%
[2,4,6,1] = 4 Points (1,1,1,1) = 4/4 = 50%

Figure 5. Base pattern *Risk Scores* for varying brute force attempts

4.3. Graphical User Interface

We aimed to replicate current consumer experiences when designing the user interface. We allowed users to set their preferred passcode if it is different from an identified common passcode. Following current standards, a minimum of four nodes must be used to create a valid passcode. The *"Set"* option only becomes available once four nodes are selected during pattern creation. However, during the unlock phase, the *"Unlock"* button is available from the beginning of user interaction. This configuration avoids providing the attacker with information on password credentials or settings. Giving passcode information to an attacker, such as the minimum length of nodes, provides no benefit for device security. For this same reason,

information regarding the *Risk Score* is also not displayed towards the user. After the passcode is set, the user is given a demo on the *Risk Score* meaning. The user may then choose what level of risk acceptance they would like to set their device to for later entry.

4.4. Human Feedback

For testing, we contacted individuals that were familiar with the Android pattern unlock system and recruited them to test out our program through the user interface.

We had 12 participants join our study. The participants completed tasks such as setting a passcode with our precautions in place and participating in mock shoulder surfing to brute force attacks.

4.4.1. Setting the Passcode

The participants were instructed to navigate through the user interface without any influence from the researchers. We reduced interaction so that we were able to collect unbiased information about their experience. The participants set an unlock pattern on a device configured with our interface (Figure 6). If the user attempted to set their passcode to one of the common patterns, they would not be permitted to do so and must try again. We took feedback on their experience after their passcode was successfully set.

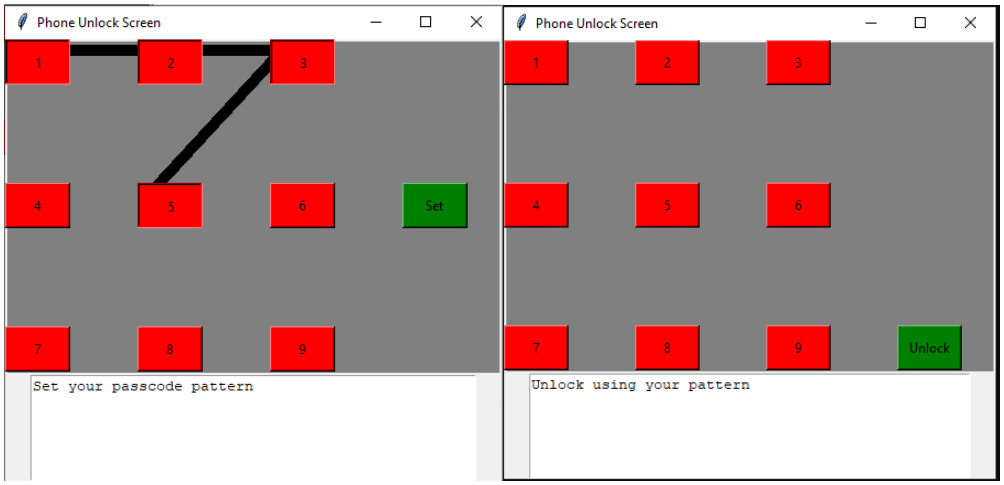


Figure 6. Setting the pattern lock

4.4.2. Simulated Shoulder Surfing Attack

In this stage, we conducted six attack scenarios that each involved two participants. One participant was instructed to sit or stand (based on personal preference) and input their passcode on the device. The other participant walked past the device user to shoulder surf. After three shoulder surfing observations, the attacker was given the device to attempt a brute-force attack. From the attack simulations and participant feedback, we acquired the following results.

5. RESULTS

5.1. Setting the Passcode

Twelve users participated in pattern setting with our interface. On average, it took participants ~ two (1.917) attempts to set an accepted pattern. User satisfaction was measured on a Likert scale of 1 to 5, with 1 being *Low User Satisfaction* and 5 being *High User Satisfaction*. The average user satisfaction was 80% (4/5), as shown in Table 1 below. The high satisfaction shows that the pattern restrictions have minimal impact on usability. The data from this testing can be found in Table 2.

Table 1. User feedback on setting new pattern

User #	Attempt Until Valid Password (/5)	Feedback (/5)
User 1	2	2
User 2	4	3
User 3	1	5
User 4	3	3
User 5	1	5
User 6	2	3
User 7	1	5
User 8	2	5
User 9	2	5
User 10	2	4
User 11	1	3
User 12	2	5
Average	1.916	4

Table 2. Attack success rate with the new imposed algorithm

User	passcode length	Attacker # of tries before phone gets locked	Attacker breach	Feedback
User 1	6	1	No	5
User 2	6	1	No	5
User 3	6	2	No	4
User 4	6	3	No	4
User 5	9	4	No	3.5
User 6	7	2	No	4
User 7	6	0	Yes	2
User 8	6	3	No	3
User 9	8	2	No	4
User 10	7	1	No	5
User 11	5	0	Yes	0
User 12	4	0	Yes	1
Average	6.333	1.583	25%	3.375

5.2. Feedback on the Algorithm

It is crucial to the success of our proposed security method that there is still a suitable level of usability satisfaction from users [13]. We tested the pattern matching algorithm with participants who are all current Android users. After the simulated attack, the shoulder surfer was then given the device and attempted to replicate the observed pattern lock. The majority, 66% (8/12), of the testers agreed that the proposed algorithm would increase the security of the pattern lock and

provide a better recovery model than the current methods on their devices. The participant feedback emphasized that email recovery is the preferred device lockout solution.

5.3. Simulated Attack Success Rates

Our pattern-matching algorithm performs best with pattern locks of length six or higher. We found that shoulder surfing attacks had lower success rates with the implementation of our algorithm. The algorithm has proved its efficiency as it reduces the 65% (13/20) success attack rate from one shoulder surfing observation to 25% (1/4) [5]. In testing, we found a user satisfaction rate of 79% (7.9/10) using patterns lock of length six or higher (Table 2).

User satisfaction for passcodes over the length of six showed a success rate of 79% and 67.5%, which includes passcodes that started from length four.

Eliminating the use of common passcodes received an agreement rate of 80%. Shoulder surfing attacks were successful on passcodes with lengths less than six as they had a 100% success rate. Interestingly, passcodes of six nodes or more only had a 10% success rate.

6. CONCLUSION

Through extensive planning, development, and testing, our algorithm improves the security of any device using a pattern unlock system.

Implementing a system that blocks the use of common patterns, users will be encouraged to create more complex lock patterns that are more difficult for an attacker to brute force. The detection of high-risk pattern unlock attempts helps to keep brute-force attackers from compromising a device.

Finally, the recovery email system helps secure the device by indefinitely blocking interaction from an attacker. There may be limitations to the email method, such as users that do not have another device to access their account on.

Testing showed high user satisfaction overall and supports that our system would increase security without having any noticeable negative impact on usability.

7. LIMITATIONS AND FUTURE WORK

Though we were successful in creating a secure system for increasing pattern unlock security, there are still areas that require improvement. A key component is the *Risk Score* which increases based on the input patterns by the user.

Our interface could improve with more features that allow the user to alter their preferred *Risk Score*. One such addition could be to implement geolocation technology. The *Risk Score* could change based on the location of the device. For example, the score would have more allowance if the phone was in a trusted area, such as the owner’s home.

Human biometrics could be incorporated into a model that recognizes the owner's behaviour, such as touch characteristics.

When setting the pattern, the user could receive better feedback about the strength of their passcode by using tools such as the efficacy meter previously discussed. These improvements could consider features such as similarity to common patterns and pattern length.

We focused on user experience for testing since we could receive direct creative feedback while evaluating our program usability. Expanding the human feedback sample size would provide higher quality insights for future work. Additionally, implementing a control group for comparison of efficacy would give meaningful comparison. Our evaluations could improve with an automated suite that could test our system's effectiveness against automated attacks.

ACKNOWLEDGEMENTS

The authors would like to thank the participants of the study that gave their time and feedback to help verify the efficacy of the proposed model. They also send appreciation to all the academic contacts that contributed suggestions and support to help this paper come to fruition.

REFERENCES

[1] C. W. Munyendo, M. Grant, P. Markert, T. J. Forman, and A. J. Aviv, "Using a Blocklist to Improve the Security of User Selection of Android Patterns," 2021, pp. 37–56.

[2] S. Higashikawa, T. Kosugi, S. Kitajima, and M. Mambo, "Shoulder-Surfing Resistant Authentication Using Pass Pattern of Pattern Lock," *IEICE TRANSACTIONS on Information and Systems*, vol. E101-D, no. 1, pp. 45–52, Jan. 2018.

[3] Y. Ku, L. Hyun Park, S. Shin, and T. Kwon, "Draw It As Shown: Behavioral Pattern Lock for Mobile User Authentication," *IEEE Access*, vol. 7, pp. 69363–69378, 2019, doi: 10.1109/ACCESS.2019.2918647.

[4] L. de Wilde, L. Spreeuwers, and R. Veldhuis, "Exploring How User Routine Affects the Recognition Performance of a Lock Pattern," in *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sep. 2015, pp. 1–8. doi: 10.1109/BIOSIG.2015.7314603

[5] V. V. Rao and A. S. N. Chakravarthy, "Analysis and bypassing of pattern lock in android smartphone," in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Dec. 2016, pp. 1–3. doi: 10.1109/ICCIC.2016.7919555.

[6] L. Bošnjak, J. Sreš, and B. Brumen, "Brute-force and dictionary attack on hashed real-world passwords," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, pp. 1161–1166. doi: 10.23919/MIPRO.2018.8400211.

[7] E. Fatima, M. Ashfaq, A. Nazir, M. H. Khan, and M. S. Umar, "A Shoulder Surfing Resistant Technique for Login on Mobile Devices," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Dec. 2018, pp. 1–4. doi: 10.1109/CCAA.2018.8777590.

[8] A. J. Aviv, J. T. Davin, F. Wolf, and R. Kuber, "Towards Baselines for Shoulder Surfing on Mobile Authentication," *Proceedings of the 33rd Annual Computer Security Applications Conference*, pp. 486–498, Dec. 2017, doi: 10.1145/3134600.3134609.

[9] G. Ye *et al.*, "A Video-based Attack for Android Pattern Lock," *ACM Trans. Priv. Secur.*, vol. 21, no. 4, p. 19:1-19:31, Jul. 2018, doi: 10.1145/3230740.

[10] Y. Wang, W. Qiu, Y. Xie, and Y. Zha, "PatternMonitor: a whole pipeline with a much higher level of automation for guessing Android lock pattern based on videos," arXiv:2102.01509 [cs], Feb. 2021, Accessed: Mar. 12, 2022. [Online]. Available: http://arxiv.org/abs/2102.01509

[11] Y. Song, G. Cho, S. Oh, H. Kim, and J. H. Huh, "On the Effectiveness of Pattern Lock Strength Meters: Measuring the Strength of Real World Pattern Locks," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2015, pp. 2343–2352. doi: 10.1145/2702123.2702365.

[12] L. Zhang, Y. Guo, X. Guo, and X. Shao, "Does the layout of the Android unlock pattern affect the security and usability of the password?," Journal of Information Security and Applications, vol. 62, p. 103011, Nov. 2021, doi: 10.1016/j.jisa.2021.103011.

[13] W. Aiken, H. Kim, J. Ryoo, and M. B. Rosson, "An Implementation and Evaluation of Progressive Authentication Using Multiple Level Pattern Locks," in *2018 16th Annual Conference on Privacy, Security and Trust (PST)*, Aug. 2018, pp. 1–6. doi: 10.1109/PST.2018.8514215.

AUTHORS

Yasir Al-Qaraghuli is currently a student at the University of Guelph in the Masters of Cybersecurity and Threat Intelligence program. He has completed a B.Sc. in Computer Science from the University of Toronto, and Applied Science University in Jordan. He has professional experience working as a Mobile Application Developer which inspires his research focus.

Caroline Hillier is a current student at the University of Guelph enrolled in the Masters of Cybersecurity and Threat Intelligence program. In 2021 she graduated from Trent University with a B.Sc. in Forensic Science and Biology. She has worked on digital security projects which inspires her research objectives of creating secure environments for end users.

MISTAKES OF A POPULAR PROTOCOL CALCULATING PRIVATE SET INTERSECTION AND UNION CARDINALITY AND ITS CORRECTIONS

Yang Tan[1] and Bo Lv[2]

[1]Shenzhen Qianhai Xinxin Digital Technology Co., Ltd, Shenzhen, China
[2]Huizhou University, China

ABSTRACT

In 2012, De Cristofaro et al. proposed a protocol to calculate the Private Set Intersection and Union cardinality (PSI-CA and PSU-CA). This protocol's security is based on the famous DDH assumption. Since its publication, it has gained lots of popularity because of its efficiency (linear complexity in computation and communication) and concision. So far, it's still considered one of the most efficient PSI-CA protocols and the most cited (more than 170 citations) PSI-CA paper based on the Google Scholar search.

However, when we tried to implement this protocol, we couldn't get the correct result of the test data. Since the original paper lacks of experimental results to verify the protocol's correctness, we looked deeper into the protocol and found out it made a fundamental mistake. Needless to say, its correctness analysis and security proof are also wrong.

In this paper, we will point out this PSI-CA protocol's mistakes, and provide the correct version of this protocol as well as the PSI protocol developed from this protocol. We also present a new security proof and some experimental results of the corrected protocol.

KEYWORDS

Private Set Intersection, PSI-CA, PSU-CA.

1. INTRODUCTION

Private Set Intersection Cardinality (PSI-CA) is an important primitive of secure two-party computation. It enables two parties, each holding a private set, to jointly compute the cardinality of their intersection without revealing any private information about their respective sets. Also, it can be extended to multi-party scenarios.

PSI-CA can be very useful in lots of scenarios. For example, in social networks(such as Facebook, Whatsapp), when two people try to determine whether they should become friends based on the number of common connections, the intuitive way is to directly exchange the information of their contacts to see if their number of common connections exceeds some threshold value. However, this method will leak private contact information. With PSI-CA, these two people can privately compute this number of common connections without revealing any private contact information.

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 113-125, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121209

Other useful scenarios include DNA sequence similarities comparison, anonymous authentication, etc.

In the existing PSI-CA protocols, the most efficient ones require linear computation and communication complexity. As far as we know, [1] is the first and the most famous protocol which achieves linear communication complexity. Following works tried to introduce more mechanisms such as bloom filters, homomorphic encryption [2], [3], [4], oblivious transfer [5] to improve the efficiency, but they basically stay at the same complexity level which is linear. A recent work [6] proposed a more efficient PSI-CA protocol, but this protocol has to sacrifice some accuracy for efficiency which makes it not suitable for some applications. Consequently, scholars [7], [8] are still inclined to reckon that [1] is one of the most efficient PSI-CA protocols.

However, in our implementation of this protocol, we found that the protocol can't return the correct result for the test data. Since the original paper is lack experiment results to back it up, the possibility of this protocol being wrong can't be ruled out. Thereby, we went through the details of this protocol and found a fundamental mistake it made. Needless to say, its correctness analysis and security proof are also wrong.

Since concerning the PSI-CA protocol, this paper is very influential and has the most cites(based on the Google Scholar search), we feel obliged to point out its mistakes and make corrections.

The structure of this paper is as follows. Firstly, we give a background introduction. Secondly, we describe some basic definitions. Thirdly, we describe the original PSI-CA and point out the mistakes it made and show the correct version of this protocol. We also show the correct version of the PSI protocol developed from the PSI-CA protocol. Fourthly, we give a correctness and security analysis of our new corrected protocol. Fifthly, we run a protocol simulation and implement the PSI-CA/PSU-CA, PSI protocols in a popular open-source federated learning framework: FATE[9]. Finally, we make a conclusion.

2. DESCRIPTION OF THE ORIGINAL PSI-CA PROTOCOL

In [1], the authors proposed a protocol to compute the Private Set Intersection Cardinality between two parties without revealing the actual private content of the set. Also, the Private Set Union Cardinality (PSU-CA) can be easily derived from the result of PSI-CA.

The authors claimed their PSI-CA protocol was secure under DDH assumption [10] in the random oracle model against semi-honest adversaries.

2.1. Definitions

Firstly, we introduced some basic definitions of the original PSI-CA.

Server A party in this paper is referred to as **S** with a private input set $s_1,...,s_w$.

Client A party in this paper is referred to as **C** with a private input set $c_1,...,c_v$. This is the party that sends a PSI-CA/PSU-CA request to the Server and it will get the final result.

Private Set Union Cardinality(PSU-CA) A protocol involving server, on input a set of *w* items $S = \{s_1,...,s_w\}$, and client, on input a set of *v* items $C = \{c_1,...,c_v\}$. It outputs $|U|$, where: $U = S \cup C$.

Private Set Intersection Cardinality(PSI-CA) A protocol involving server, on input a set of *w* items $S = \{s_1, ..., s_w\}$, and client, on input a set of *v* items $C = \{c_1, ..., c_v\}$. It outputs |*I*|, where: $I = S \cap C$.

Private Set Intersection (PSI) A protocol involving server, on input a set of *w* items $S = \{s_1, ..., s_w\}$, and client, on input a set of *v* items $C = \{c_1, ..., c_v\}$. It outputs *I*, where: $I = S \cap C$.

For both PSI-CA and PSU-CA protocols, the following privacy requirements should be met:

- **Server Privacy** Client learns no information beyond: (1) cardinality of set intersection/union and (2) upper bound on the size of S.
- **Client Privacy** No information is leaked about client set C, except an upper bound on its size.
- **Unlinkability** Neither party can determine if any two instances of the protocol are related, i.e., executed on the same input by client or server, unless this can be inferred from the actual protocol output.

One thing to note, for any set *C* and *S*, the size of the union $C \bigcup S$ can be computed as $|C| + |S| - |C \bigcap S|$. Thereby, with the result of PSI-CA, one can easily get PSU-CA.

Other definitions involved in this protocol:

- Two hash functions act as random oracles, $H1: \{0,1\}^* \rightarrow Z_p^*$ and $H2: \{0,1\}^* \rightarrow \{0,1\}^k$ given the security parameter *k*. *H*1 and *H*2 are both deterministic which means if *x* = *y*, we have $H1(x) = H1(y)$ and $H2(x) = H2(y)$.
- Two random data permutations $\prod$, $\prod$' which randomly shuffles the item order of a data set. All the calculations will happen in *G* which is a cyclic group of order *q* and with *g* as its generator.

2.2. Original Protocol

Here in Figure 1, shows the original protocol described in [1].

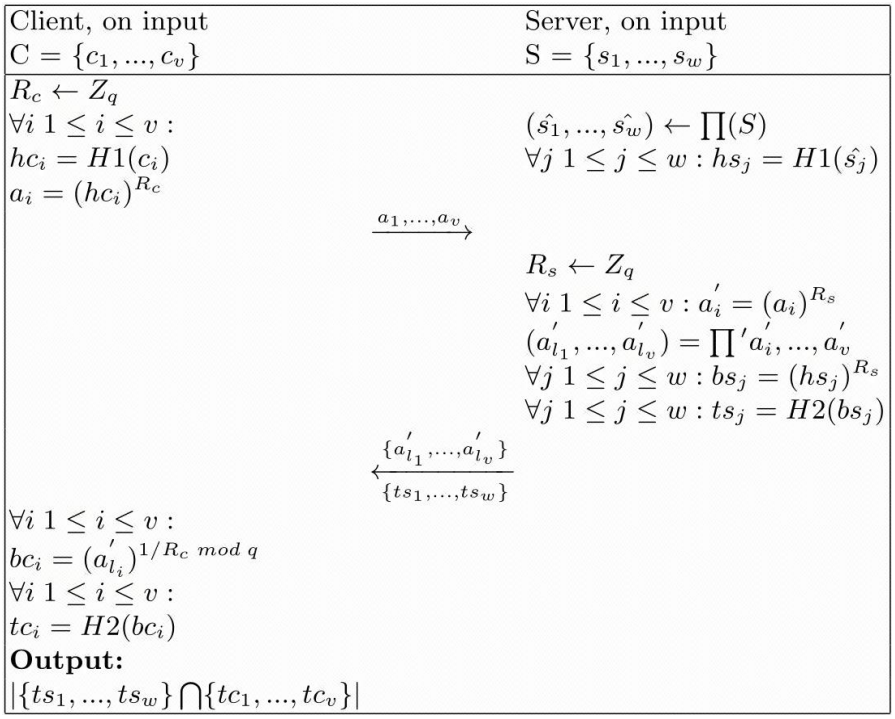


Figure 1. Original PSI-CA Protocol from [1].

The original protocol's idea comes from the following intuition.

Intuition. First, the client masks its set items c_i with hash function $H1$ and a random exponent (R_c) generated on his side and sends the resulting values (a_i) to the server which further masks them by exponentiating them with its own random value R_s. The server randomly shuffles(i.e. $\prod$') these values to prevent the client from recovering the exact intersecting item by item's order. The resulting values a'_{l_i} after shuffle will be sent back to the client.

Then, the server masks its own set items s_i by the following order: A new random shuffle $\prod$; hash function $H1$; exponentiating them with its own random value R_s; hash function $H2$. The resulting values ts_i will be sent to the client for client to do the final PSI-CA calculation.

When client receives further randomly shuffled and exponentiated items a'_{l_i}, he will strip of the initial exponent (R_c) by exponentiating R_c's reverse modular q. He then further masks the set items with the hash function $H2$ and outputs tc_i.

From the process description, we can see that, from the client's side, in the end, the data he gets tc_i, ts_i are the client and server's private set items which went through the same calculations: hash function $H1$ and exponentiating with R_s and hash function $H2$ and along with some random order shuffles in the process which won't affect the actual value but the order. Thereby, if $c_i = s_j$, then we have $tc_i = ts_i$ and the client can do the intersect cardinality computation with simple equality checks between these two data sets $\{tc_i\}$, $\{ts_j\}$. However, the client cannot recover the original private item under the DDH assumption since he doesn't know the random value R_s and how the items are shuffled ($\prod$, $\prod$').

To draw a conclusion, the client finally gets the correct result of PSI-CA without knowing the actual intersection.

2.3. Why it's wrong and its corrections

From the protocol description and the intuition behind it, the original protocol seems alright. However it made a fundamental mistake: all the computations, including hash function and exponentiating with random values, didn't happen in the chosen cyclic group. Without mapping values to the chosen group, its security dependency: DDH assumption won't stand under this circumstance. It will be just normal computations in modular *p*.

To be more specific, computations like stripping off the initial exponent R_c by exponentiating with R_c's reverse modular *q* won't work.

After figuring out what's wrong with the original protocol, we correct it by mapping the intermediate results (values after being processed by hash function *H*1) to the chosen cyclic group for both the client and the server sides. Specifically, we do more exponentiations with generator *g* as the base, and intermediate results as the exponents. With this correction, the client can successfully strip the initial exponent (R_c) off now.

In Figure 2, we show the description of the full corrected protocol.

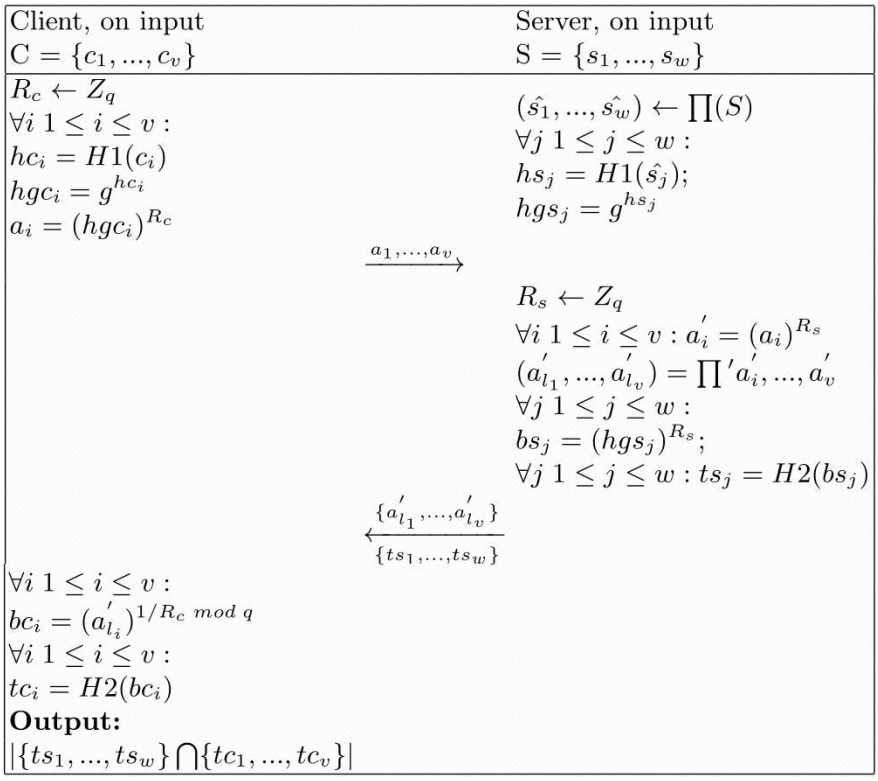


Figure 2. Correct version of the PSI-CA Protocol from [1].

Complexity. The correct version of protocol's complexity remains the same level (linear complexity) as the original paper [1] claimed which is O(*w*+*v*) computation and communication. Specifically, according to the protocol description in Figure 2, the client performs 3*v* exponentiations with |*q*|−bit exponent and |*p*|−bit modular, 2*v* hash functions and send *v* data items to server. On the other side, the server performs 2*w* + *v* exponentiations with |*q*|−bit exponent and |*p*|−bit modular, 2*w* hash functions and sends *v* + *w* data items to the client.

2.4. Corrections on the extended PSI protocols

In the original paper, based on the PSI-CA protocol, the authors extended two additional protocols.

One is called authorized PSI-CA (APSI-CA) in which client's input must be pre-authorized by an off-line mutually-trusted authority. In this protocol, all client's input $c_1,c_2,...,c_v$ will be attached with an RSA signature $\sigma_1,\sigma_2,...,\sigma_v$ released by the mutually trusted authority on its hashed value (with hash function $H1$). This time, the authors didn't make a mistake since this protocol's security no longer relies on the DDH assumption but on RSA [11]. Also, we checked the protocol's mathematical deduction. No correction is needed.

The other one is PSI protocol which was directly derived from the original PSI-CA protocol with only one additional round of communication. Ergo, the PSI protocol made the same mistakes as PSI-CA protocol. Thereby, it also needs correction. We skipped the description of the wrong PSI protocol. In Figure 3, we directly give the correct version of the extended PSI protocol. Compared to the PSI-CA protocol, we can see that PSI has an extra last step in which the client sends the intersection of $\{tc_i\}$ and $\{ts_j\}$ to the server for the server to do the final matching and output the PSI since server can deduce the original item from ts_j. Besides that, all the steps are the same with PSI-CA. Ergo, in Figure 3, we make the same correction as PSI-CA.

Client, on input $C = \{c_1, ..., c_v\}$		Server, on input $S = \{s_1, ..., s_w\}$
$R_c \leftarrow Z_q$ $\forall i\ 1 \le i \le v:$ $hc_i = H1(c_i)$ $hgc_i = g^{hc_i}$ $a_i = (hgc_i)^{R_c}$		$(\hat{s_1}, ..., \hat{s_w}) \leftarrow \prod(S)$ $\forall j\ 1 \le j \le w:$ $hs_j = H1(\hat{s_j});$ $hgs_j = g^{hs_j}$
	$\xrightarrow{a_1,...,a_v}$	
		$R_s \leftarrow Z_q$ $\forall i\ 1 \le i \le v: a_i^{'} = (a_i)^{R_s}$ $(a_{l_1}^{'}, ..., a_{l_v}^{'}) = \prod{}' a_i^{'}, ..., a_v^{'}$ $\forall j\ 1 \le j \le w:$ $bs_j = (hgs_j)^{R_s};$ $\forall j\ 1 \le j \le w: ts_j = H2(bs_j)$
	$\xleftarrow[\{ts_1,...,ts_w\}]{\{a_{l_1}^{'},...,a_{l_v}^{'}\}}$	
$\forall i\ 1 \le i \le v:$ $bc_i = (a_{l_i}^{'})^{1/R_c \bmod q}$ $\forall i\ 1 \le i \le v:$ $tc_i = H2(bc_i)$ $T = \{ts_1, ..., ts_w\} \bigcap \{tc_1, ..., tc_v\}$		
If $PolicyIsSatisfied(w, v, \lvert T\rvert):$	$\xrightarrow{T}$	$\forall tc_j \in T:$ **Output** $c_j \in S \cap C$

Figure 3. Correct version of the PSI Protocol from [1].

Remark 1: The authors seemed to forget to include the calculations of ts_j in the PSI protocol description in [1]'s Figure 3, we also correct this part.

Remark 2: *PolicyIsSatisfied* are some policies for the client to decide whether to proceed with PSI protocol based on the PSI-CA result and w, v. For example, PSI-CA result must exceed a

threshold value(e.g. 0.8*v*) to proceed PSI. Also, the roles of server and client are not fixed. They can be reversed depending on the specific application scenario.

3. CORRECTNESS AND SECURITY ANALYSIS

In this section, we will give the correctness and security proof of our corrected protocol.

3.1. Correctness

For any c_i held by client and s_j held by server, if $c_i = s_j$, hence, $hc_i = hs_j$, we obtain:

$$\begin{aligned}
&|\{ts_1,\ldots,ts_w\}\bigcap\{tc_1,\ldots,tc_v\}| \\
=&|\{bs_1,\ldots,bs_w\}\bigcap\{bc_1,\ldots,bc_v\}| \\
=&|\{g^{hs_1\cdot R_s},\ldots,g^{hs_w\cdot R_s}\}\bigcap\{g^{hc_1\cdot R_s},\ldots,g^{hc_v\cdot R_s}\}| \\
=&|\{g^{hs_1},\ldots,g^{hs_w}\}\bigcap\{g^{hc_1},\ldots,g^{hc_v}\}| \\
=&|S\bigcap C|
\end{aligned} \tag{1}$$

hence, client learns set intersection cardinality by counting the number of matching pairs (tc_i, ts_j).

3.2. Security Analysis

In this section, we will define the threat model and some security assumptions related to our protocol. Then we give the security proof of the corrected PSI-CA protocol. The PSU-CA and PSI protocol's security proofs are skipped since they are directly derived from the PSI-CA protocol and their security proof are basically the same as PSI-CA's.

3.2.1. Threat Model

Firstly, we define the threat model and classes of adversaries of our PSI-CA, PSU-CA, PSI protocols.

Semi-honest Security In the Semi-honest Security threat model, the protocol participants will always follow the protocol's procedure. However, they are curious and will try to gain extra information out of the protocol.

After giving the definition of the threat model, we can now define two classes of adversaries concerning the Client privacy and Server privacy security requirements defined in the previous section.

Client adversary The client participant of the protocol who tries to violate the Server privacy. With the data he gets on his side, he will try to recover server's private items or intersection items.

Server adversary The server participant of the protocol who tries to violate the Client privacy. With the data he gets on his side, he will try to recover client's private items or intersection cardinality.

One thing to note, adversaries outside of the protocol are not considered since they can't gain more information than the protocol participants. If the more powerful participant adversary can't break the protocol, outsider adversaries can't break the protocol either.

3.2.2. Security Assumptions

Discrete Logrithm Assumption Let G be a cyclic group and g be its generator. The discrete logarithm problem(DLP) is called (t, ε) hard relative to G if for all algorithms A runs in time t there exists a negligible function ε of security parameter k such that

$$Pr[\mathrm{A}(g, g^a) = a] \leq \varepsilon \qquad (2)$$

DDH assumption Let G be a cyclic group of order q and g is its generator and l is the bit-length of the group size. The following two distributions are computationally indistinguishable

(1)g^x, g^y, g^{xy}
(2)g^x, g^y, g^z

given x, y, z are randomly and independently chosen from Z_q. To put it in a more formal way, DDH problem is (t, ε) hard if for all algorithms A runs in time t there exists a negligible function ε.

$$\begin{aligned} &|\Pr[x, y \leftarrow \{0,1\}^l : A(g, g^x, g^y, g^{xy}) = 1] - \\ &\Pr[x, y \leftarrow \{0,1\}^l : A(g, g^x, g^y, g^z) = 1]| \leq \varepsilon \end{aligned} \qquad (3)$$

If we consider more powerful adversaries, we have following security assumptions:

One-More-DH assumption [12] DH problem is hard even if the adversary is given access to a $DH_x(.)$ oracle (i.e. given random h in group G, the oracle return h^x). Formally, let $(G, q, g) \leftarrow KeyGen(k)$, and $x \leftarrow Z_q$, we say One-More-DH is (t, ε) hard for all algorithm A runs in time t there exists a negligible function ε:

$$\Pr[\{(g_i, (g_i)^x)\}_{i=1,\ldots,v+1} \leftarrow A^{DH_x(.)}(g_1, \ldots, g_{ch})] \leq \varepsilon \qquad (4)$$

where $ch \geqslant v$ and A can make at most v queries to $DH_x(.)$.

3.2.3. Security Proof

a) *Client Privacy:*

According to the threat model, the server adversary will try to violate the Client Privacy. According to the protocol description, the data he will receive from the client is $\{a_i\}$ where $a_i = (hgc_i)^{R_c}$ in which $hgc_i = g^{hc_i}$, $hc_i = H1(c_i)$.

The private items the server adversary tries to recover from the client can be divided into the following classes:

- **Non-intersection Data** If no server item satisfies $s_j = c_i$, because of the hardness of the DLP and $a_i = g^{hc_i \times R_c}$, server adversary can't find an algorithm A that runs in (t, ε) to recover $hc_i \times R_c$. Even if we discard the DLP assumption, and let the server adversary gets $hc_i \times R_c$, without the knowledge of random R_c, $hc_i \times R_c$ is indistinguishable from a random $r \leftarrow Z_q$. Thereby, under this circumstance, Client Privacy is guaranteed.

- **Intersection Data** If server adversary has item satisfies $s_j = c_i$, then the server adversary has $a_i = g^{hc_i \times R_c}$ and some $g^{hs_j} = g^{hc_i}$. We consider the following casese:

(1) If the intersection cardinality equals 0, i.e. the private intersection set is empty, the security proof is the same as Non-intersection Data case.

(2) If the intersection cardinality equals 1 and let's assume the intersecting item is c_i, even if we discard the DLP assumption let server adversary gets $hc_i \times R_c$ from a_i, without the knowledge of R_c, $hc_i \times R_c = hs_j \times R_c$ is indistinguishable from a random $r \leftarrow Z_q$, he can't recover the hs_j to get the corresponding intersection item.

(3) If the intersection cardinality is greater than 1, let's assume it's n, and the intersection item is $s_{j1},...,s_{jn}$, then the server adversary has $g_1 = g^{hs_{j1}},..., g_n = g^{hs_{jn}}$ along with $a_{j1} = g_1^{R_c}, a_{j2} = g_2^{R_c},..., a_{jn} = g_n^{R_c}$. This case matches the One-More-DH assumption. According to this assumption, even if we give the server adversary access to $DH_x(.)$ oracle and let it recover [0, $n-1$] pairs of $(g_1, a_{j1}),(g_2, a_{j2}),...,(g_n, a_{jn})$, it can't find an algorithm A runs in (t, ε) to recover another pair of (g_t, a_{jt}) where $1 \leqslant t \leqslant n$ and use g_t to get the corresponding intersection item. Thereby, the client privacy is also guaranteed for this case under One-More-DH assumption.

Thereby, for both Intersection Data and Non-intersection Data, Client privacy is guaranteed.

b) *Server Privacy:*

According to the threat model, the client adversary will try to violate the Server Privacy. According to the protocol description, the data he will receive from the server is $\{ts_j\}$ and $\{a'_{l_i}\}$ where $ts_j = H2((hgs_j)^{R_s}) = H2(g^{hs_j \times Rs})$ and $a'_{l_i} = g^{hc_i \times R_c \times R_s}$.

For the simplicity of the proof calculations, the random permutations that prevent the client from recovering intersecting items based on the item's order are not concluded in the calculations. As long as these permutations are random, recovering items order has no advantage over blindly picking items from client's data set. Thereby, it's safe from order- recovering attack.

Similar to the previous case, the private items the client adversary tries to recover from the server can be divided into the following classes:

- **Non-intersection Data** If no client item satisfies $c_i = s_j$, since $H2$ is modeled as random oracle, ts_j is indistinguishable from random $r \leftarrow Z_q$. Even if the adversary somehow reverses the $H2$ and gets $(hgs_j)^{R_s} = g^{hs_j \times Rs}$, because of the hardness of the DLP, client adversary can't find an algorithm A that runs in (t, ε) to recover $hs_j \times R_s$. Even if we discard the DLP assumption and let server adversary gets $hs_j \times R_s$, without the knowledge of R_s, $hs_j \times R_s$ is indistinguishable from a random $r \leftarrow Z_q$. Thereby, under this circumstance, Server Privacy is guaranteed.

- **Intersection Data** If the client adversary has item that satisfies $c_i = s_j$, for this case, $H2$ and R_c can both be stripped of by client adversary to perform the attack. The client adversary will have $bc_i = g^{hc_i \times R_s} = g^{hs_j \times R_s}$ and some $g^{hc_i} = g^{hs_j}$

We consider the following cases:

(1) If the intersection cardinality equals 0, i.e. the private intersection set is empty, the security proof is the same as Non-intersection Data case above.

(2) If the intersection cardinality equals 1 and assume the intersecting item is $c_i = s_j$. Because of the DLP assumption, client adversary can't get $hc_i \times R_s$ or the specific $g^{hc_i} = g^{hs_j}$ without the knowledge of R_s. Even if we discard the DLP assumption let client adversary gets $hs_j \times R_s$, without the knowledge of R_s, $hs_j \times R_s = hc_i \times R_s$ is indistinguishable from a random $r \leftarrow Z_q$. Thereby, he can't recover hc_i to get the corresponding intersection item.

(3) If the intersection cardinality is greater than 1, let's assume it's n, and the intersection item is $c_{i1} = s_{j1}$,..., $c_{in} = s_{jn}$, then the client adversary has $g_1 = g^{hc_{i1}}, ..., g_n = g^{hc_{in}}$ along with $bc_{i1} = g_1^{R_s}, bc_{i2} = g_2^{R_s}, ..., bc_{in} = g_n^{R_s}$. This case also matches the One-More-DH assumption. According to this assumption, even if we give the client adversary access to $DH_x(.)$ oracle and let it recover [0, $n - 1$] pairs (g_1, bc_{i1}),(g_2, bc_{i2}),...,(g_n, bc_{in}), it can't find an algorithm A runs in (t, ε) to recover one more pair of (g_t, bc_{jt}) where $1 \leqslant t \leqslant n$ and use g_t to get the corresponding intersection item. Thereby, the server privacy is also guaranteed for this case under One-More-DH assumption.

Thereby, for both Intersection Data and Non-intersection Data, Server privacy is guaranteed.

4. PROTOCOL SIMULATION AND IMPLEMENTATION

After we give the correct descriptions of the PSI and PSI-CA protocol, in this section, we will make protocol simulations and implementations.

Firstly, we run a simple python simulation for the PSI-CA protocol. The python environment we use for simulation is Python 3.6.8 and we run this simulation on a Linux Red Hat 4.8.5-44 server with Intel Cure2 Duo T7700(2.4GHz) as CPU.

The parameter for the cyclic group comes from the [13]: **1024-bit MODP Group with 160-bit Prime Order Sub-group** which means a 1024-bit p and 160-bit q.

SHA256 with different salts are used to construct $H1$ and $H2$ separately.

The initial parameters for client and server are illustrated in Figure 4 (Only showed a part of p because of its great length).

```
1024bit-p:124325339146889384540494091085456630009856882741872806181731279018491820800119460022367403769795008250021191767583423221479l
160bit-q:1399252811935680595399801714158014275474696840019
client's random value rc:902159083074063151500266207821672078422883169112
server's random value rs:45405261896976746793102016638521922217200232867 7
client c_i: 3
client c_i: 4
client c_i: 5
client c_i: 2
client c_i: 6
server s_j: 3
server s_j: 4
server s_j: 5
server s_j: 7
```

Figure 4. Initial Parameters For PSI-CA.

The intermediate results for tc_i and ts_j and final outcome are illustrated in Figure 5.

```
ts_j 00f55c196284695cf1646580471bd3aff8cb4e570127db08878640150812b677
ts_j 18e95e7c065a7f51749174d26fb9b458edd1246b602d6c419059fbed9eab17c4
ts_j 7ab697f79c32457c8a337e2b4c66770ff1fa7dd85e614414400a4c906081a7cd
ts_j ac183e2d6c737cda6429e8b92e9f511f6418fb981c71252756afe71b67930602
tc_i: 00f55c196284695cf1646580471bd3aff8cb4e570127db08878640150812b677
tc_i: 4da9cb488691c688baa3076522ba42b62e70c3dfd12169baeef9d83c3d0fc086
tc_i: 6a406977a818d6e0382d080f306441e45946b7421d97087cc96710dbaf3c0b9a
tc_i: 7ab697f79c32457c8a337e2b4c66770ff1fa7dd85e614414400a4c906081a7cd
tc_i: ac183e2d6c737cda6429e8b92e9f511f6418fb981c71252756afe71b67930602
intersect of tc_i and ts_j 00f55c196284695cf1646580471bd3aff8cb4e570127db08878640150812b677
intersect of tc_i and ts_j 7ab697f79c32457c8a337e2b4c66770ff1fa7dd85e614414400a4c906081a7cd
intersect of tc_i and ts_j ac183e2d6c737cda6429e8b92e9f511f6418fb981c71252756afe71b67930602
PSI_CA: 3
```

Figure 5. Intermediate Results tc_i, ts_j and Final Output for PSI-CA.

The simulation proves our new version of PSI-CA protocol is correct.

We didn't do the python simulation for PSI. Instead, we directly implement PSI-CA, PSU-CA, and PSI protocols in a very popular (3.1k stars and 900 forks) open-source multi- party federated learning framework FATE[9] in which PSI is very common operations for different parties to perform Sample-Aligned Federated Learning. Since this framework doesn't support PSI-CA, our work can be a pretty useful supplementary for FATE.

In FATE, there are three different parties: Guest, Host and Arbiter. In our implementation, the Guest party will play the role of client which will get the result of PSI-CA, PSU-CA. The Host party will play the role of server which can get the PSI result if the Guest agrees to proceed with the PSI computation. The Arbiter is not needed in our protocols. Both Guest and Host can choose to share their side's result with another party by setting parameters in a configuration file.

For the illustration, we use the data of "epsilon_5k_hetero_guest.csv", "epsilon_5k_hetero _host.csv" to run a test on the same server as we run PSI-CA simulation. We deploy FATE in the Stand-alone Host mode and the tested data files can be found in the "examples/data/" directory on [9]'s repository.

In Figure 6, we show the result that the Guest party gets which includes PSI-CA and PSU-CA. The result indicates Guest and Host has 5000 items in PSI and 5000 items in PSU.

Figure 6. Final Output for PSI-CA and PSU-CA.

In Figure 7, we show the result that the Host party gets which is PSI: the table in this figure showed part of the intersection items.

Figure 7. Results of PSI.

The whole PSI-CA/PSU-CA and PSI protocol finished in 7.81s.

5. CONCLUSION

In this paper, we pointed out some mistakes of a popular PSI-CA protocol and its extended PSI protocol.

Since the original protocol's computations didn't happen in the chosen cyclic group, the protocol turned out to be wrong as well as its security proof.

After figuring out what's wrong with these protocols, we made some corrections by mapping the intermediate results to the chosen cyclic group. After the corrections, we also presented a new correctness analysis and security proof of the corrected PSI-CA protocol.

To further back up our corrected protocols, we ran a python simulation of the PSI-CA protocol and implemented PSI-CA/PSU-CA, PSI protocols in a popular open-source federated learning framework FATE. Turned out, our new corrected protocols worked perfectly well.

REFERENCES

[1] Cristofaro, Emiliano De, Paolo Gasti, and Gene Tsudik. "Fast and private computation of cardinality of set intersection and union", *International Conference on Cryptology and Network Security*. Springer, Berlin, Heidelberg, 2012.

[2] Egert, Rolf, et al. "Privately computing set-union and set-intersection cardinality via bloom filters", *Australasian Conference on Information Security and Privacy*. Springer, Cham, 2015.

[3] Debnath, Sumit Kumar, et al. "Secure and efficient multiparty private set intersection cardinality", *Advances in Mathematics of Communications* 15.2 (2021): 365.

[4] Debnath, Sumit Kumar, and Ratna Dutta. "Secure and efficient private set intersection cardinality using bloom filter", International Conference on Information Security. Springer, Cham, 2015.

[5] Ion, Mihaela, et al. "On deploying secure computing: Private intersection-sum-with cardinality", *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2020.

[6] Dong, Changyu, and Grigorios Loukides. "Approximating private set union/intersection cardinality with logarithmic complexity", *IEEE Transactions on Information Forensics and Security* 12.11 (2017): 2792-2806.

[7] Chen, Hao, Kim Laine, and Peter Rindal. "Fast private set intersection from homomorphic encryption", *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 2017.

[8] Shi, Run-Hua. “Efficient quantum protocol for private set intersection cardinality”, *IEEE Access* 6 (2018): 73102-73109.
[9] Webank, “Federated ai technology enabler”, https://github.com/FederatedAI/FATE, last Accessed November 18th, 2021.
[10] Dan Boneh, “The decision diffie-hellman problem,” in *International Algorithmic Number Theory Symposium*. Springer, 1998, pp. 48–63.
[11] Rivest, R. L., A. Shamir, and L. Adleman. “A method for obtaining digital Signatures and public key Cryptosystems”, communications of the ACM Vol. 27”, 21(1978) (1978).
[12] Bellare, Mihir, et al. “The One-More-RSA-Inversion Problems and the Security of Chaum's Blind Signature Scheme”, *Journal of Cryptology* 16.3 (2003).
[13] Lepinski, Matt, and Stephen Kent. Additional Diffie-Hellman groups for use with ietf standards. RFC 5114, January, 2008.

AUTHORS

Yang Tan Received his B.Eng. and Ph.D. from South China University of Technology. He now works as a cryptography researcher at Shenzhen Qianhai Xinxin Digital Technology Co., Ltd and his research interests include applied cryptography, secure computation, federated learning, blockchain, etc.

Bo Lv Received her B.Eng. and Ph.D. from South China University of Technology. She now works as a lecturer at Huizhou University and her research interests include public key cryptography, secure multi-party computation, etc.

CRYPTO YOUR BELONGINGS BY TWO PIN AUTHENTICATION USING ANT ALGORITHM BASED TECHNIQUE

Janaki Raman Palaniappan

Brunswick Corporation, USA

ABSTRACT

Everyone realize data is one of the important strategic for any company to run and win the business. Let it be a mobile apps, websites and so on, there are more chances that our personal data like images, videos, texts get expose while we share across for different purposes. Even though the company says app, website forms are encrypted, the said company itself uses the data internally for their business development. This research presents how one can secure own's data themselves before sending. There are many cryptography methods that has evolved from time to time. Upon researching and analyzing, I present a unique method to encrypt and decrypt the data, using combination of techniques such as Cryptographic technique, ANT Algorithm based formula and logic gates that would provide stronger protection to the data. Secure your images, videos with a 2-pin authentication and protection to encrypt and decrypt the data. A user must provide 2 different symmetric pins to encrypt and decrypt, where first pin shall be up to 4-digit secret pin and a second pin is a single digit pin. Single digit pin acts on how many stages the encryption takes place. The proposed method had been experimented on several images and videos. This study reveals, A combination of secret keys, ANT algorithm and Logic gates makes difficult for anyone to hack the data. This unique methodology helps us to protect our data more safely at source device itself.

KEYWORDS

Visual Cryptography, Ant Algorithm, Logic Gates Technique.

1. INTRODUCTION

In the history we know every time one or other emperors were ruling around the world. Now that we live in a computer era where only emperor named 'Internet' is ruling all over the world and entered our personal space. Data in one part of the world is accessible in other part of the world in a millisecond. In the present communication the major issue with data transfer is the security and authenticity.

On a day-to-day basis, we use mobile apps, public email services, websites for different purpose to store/share Images such as Driving License, Passport, Marksheets, etc., and videos that contains sensitive information are very crucial and must be secured. These data are used by companies directly or indirectly to build their business. Unauthorized user hacking this sensitive information could lead to lot of social problems. So, protecting self-data is our own responsibility rather than to be a victim.

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 127-133, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121210

My research objective is how one can protect owns data in source itself before we share across. Even though there are many cryptography methods available and evolved, I researched to develop a unique method to encrypt and decrypt the data. This method is a combination of visual cryptographic technique with 2 pin authentication, ant algorithm formula and logic gates technique. This makes a cryptography method unique to encrypt and decrypt the data, making it difficult for anyone to hack the data. This also ensures the data is safe at user end itself.

2. VISUAL CRYPTOGRAPHY

Cryptography is the way to keep the information secret and safe. It helps in hiding the data and allows only intended users to view the content. Cryptography is widely used due to great security. There are 2 methods that are used widely,

Symmetric Key – Both the sender and receiver should have the same key to view the information. Other name for this method is known as Secret Key cryptography.

Asymmetric Key – This cryptography uses pair key based technique i.e., public key and private key to Encrypt and Decrypt the data. It is also named as public key cryptography.

Similarly Visual Cryptography is a technique that allows the images, videos, etc., to be encrypted thatconverts in a non-readable format. Only authorized user is allowed to decrypt the data. Once decrypt takes place, the visual appears the same.

3. ANT ALGORITHM

Ant algorithm is based on the behavior of how ant's searches the food. Ant starts searches, by wander randomly. Once the ant finds the food, it picks and goes back to source place leaving the markers to show the path has food. When other ants come across the markers, they certainly follow the markers and leave their markers thus making the path stronger and shorter.

4. ANALYSIS AND RESEARCH

4.1. Problem

Daily, we use many apps in mobile, public emails, browsers, etc., We share our personal data like photos, Driving License, passport, ids, etc., for different purposes like jobs, verifications, security, etc., Upon sending, our personal data are used by companies for their business development strategy because data analytics/science is one of main key to success and growth. Securing one's own personal data lies in own hand.

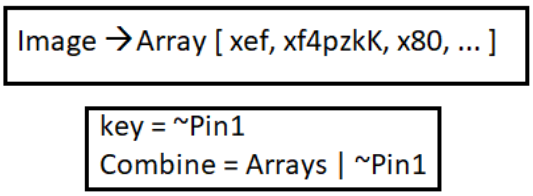
Even though there are multiple cryptographic techniques available. I have come up with a unique method of visual cryptography to safeguard the data in our source device itself like mobile, laptop, etc., before we send across internet.

4.2. Method

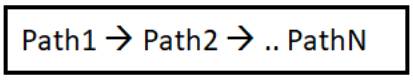
There are many methods how visual cryptography technique can be used to secure the image. In my experience I decided to take a different path to handle this technique and would like to share my research work.

Encryption using combination of multiple techniques such as 2 PIN authentication, ANT algorithm technique and logic gates technique on how the data to be encrypted. This technique provides a very high security to the data. If anyone wants to view the data, person must decode 2 secret PIN/Code. Combination of 2 secret PINs makes it more difficult for anyone to hack it.

The user who wants to encrypt the image must provide 2 secret codes (Pin1 and Pin2) and remember it. First code (Pin1) is up to 4-digit secret code and the second code (Pin2) is a single digit secret code up to 5. Initially the image will be decoded into series array of bytes, next the first secret code will be inverted to hide original code, then logic gate technique is used to combine array of bytes values and the inverted first secret code value and write them as series of bytes back into the image. At this stage, the image is encrypted partially.



Based on the second secret code value, the Ant algorithm technique is used to decide on how long the encryption technique must travel. Also, how long the travel is, the encryption will be repeated at each stage. The length of the travel is based on the second secret code value.



At each encryption hop (stage), first secret code will be changed to a different value using logic gate technique. Also generates the new series of array of bytes by combining previous stage series of array of bytes and new inverted first secret code and write them into the image. This process is repeated at each stage until it reaches final stage. This design makes the encryption key even stronger.

Once the ant reaches the final hop which is decided based on second secret code value, final time the encryption technique will be done, and the message is issued to the user. Here the image is completely encrypted. This acts as a double protection and multiple encryptions technique applied to the image.

At this stage, the user is safe to transfer the image to someone or save in an email, etc. It is encrypted.

Once the destination user receives the image, user can decrypt the image. As the symmetric key method is followed here. A user must remember both the secret codes to be able to decrypt the message.

When the user enters both the secret codes, the technique of decrypting the image starts. Remember, here the image contains series of array of bytes as it was encrypted at last step.

Ant algorithm plays a major role in decryption technique as it helps the decrypt hops to follow the travel path (markers) of encryption technique. At each decryption stage, first secret code will be changed to a different value using logic gate technique. Ant algorithm shows the travel path until reaches the final stage.It helps in optimizing the path.

At each decryption hop (stage), again the image will be decoded into series of array of bytes values, next the first secret code to be inverted once again to be able to obtain the hidden original secret code, then logic gate technique is used to combine array of bytes values and the inverted first secret code value and write them as series of bytes back into the image. At this stage, the image is decrypted partially.

Path1 ← Path2 ← .. PathN

As you know the length of the travel is based on the second secret code value. Ant algorithm helps the decrypt technique to travel the right path until it reaches the source. At each decrypting stage, the Ant algorithm technique is used to decide on how long the decryption technique must travel. Also based on the distance of travel, the decryption will be repeated at each stage.

Once the path is back to starting point. At this stage, User is authorized to view the content. Refer Figure4.2.1 for chart representation.

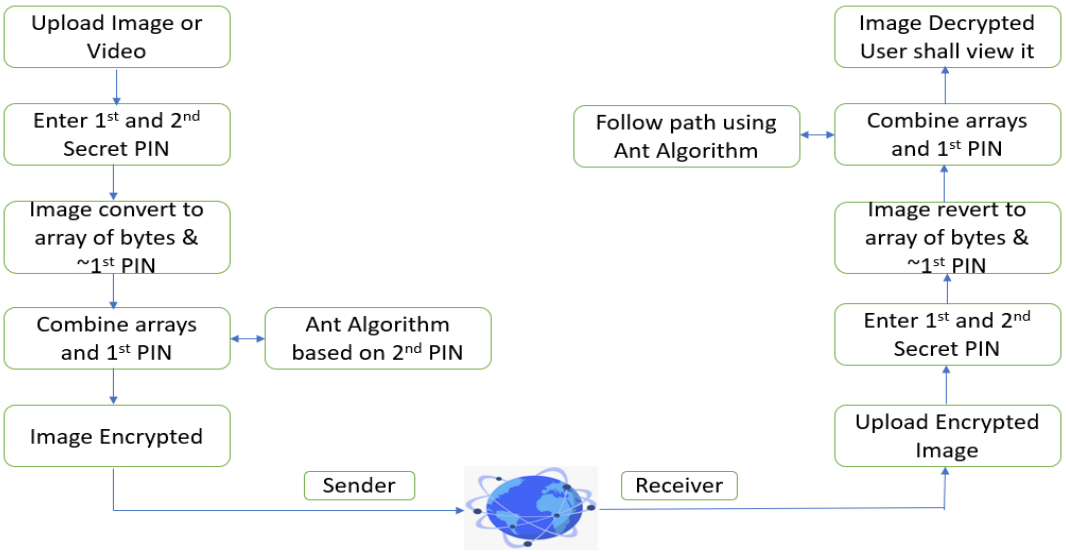


Figure 4.2.1 – Flow Chart of visual cryptography

The Limitations are,

- When the secret codes PIN are entered wrong the image or video gets corrupted. If user do not remember secret PIN, it is suggested to have a copy of the image or video to retry if PIN are remembered.
- Encryption/Decryption size of the images/videos that were tested are up to 10 MB.
- The 2nd secret PIN was tried with maximum number 5.

4.3. Sample Results

ORIGINAL .JPG IMAGE → ENCRYPTION → DECRYPTION → ORIGINAL .JPG IMAGE

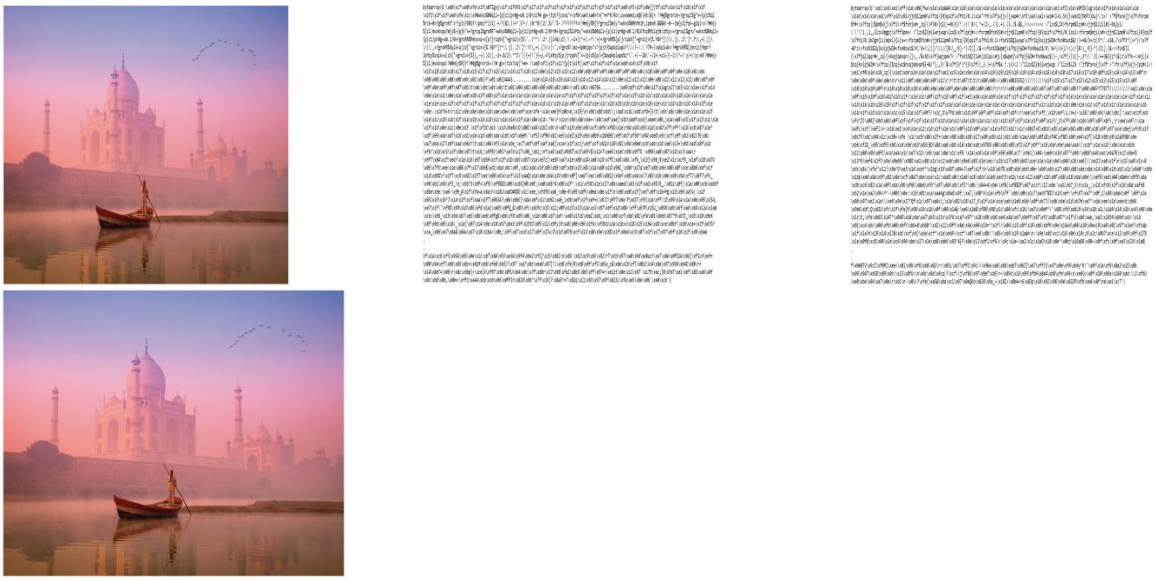


Figure 4.3.1 –JPG encrypt/decrypt output

ORIGINAL .PNG IMAGE → ENCRYPTION →DECRYPTION → ORIGINAL .PNG IMAGE

Figure 4.3.2 – PNG encrypt/decrypt output

4.4. Result Comparison

Images/Videos	Size	Degree of 2nd PIN	Time taken to Encrypt	Time taken to Decrypt
mahal.jpg	171 KB	Medium	532 ms	441 ms
mahal.jpg	171 KB	Low	308 ms	307 ms
PP.png	74 KB	Low	149 ms	123 ms
PP.png	74 KB	Medium	216 ms	202 ms
PP.png	74 KB	High	310 ms	340 ms
Egg.jpg	4.07 MB	Low	5.1 secs	5.0 secs
Egg.jpg	4.07 MB	Medium	16.8 secs	15.91 secs
Egg.jpg	4.07 MB	High	26.6 secs	26.2 secs
Sand.jpg	6.90 MB	Medium	56.12 secs	43.48 secs
Sand.jpg	6.90 MB	High	57.26 secs	1 min 1 sec
vid.mp4	193 KB	Low	342 ms	254 ms
vid.mp4	193 KB	Medium	455 ms	434 ms
vid.mp4	193 KB	High	683 ms	553 ms

Figure 4.4.1 – Table Comparison output

4.5. Tabulation Discussion for Ant Algorithm

50+ variety of images/videos with the different sizes have been analyzed and the part of results are given in the Fig: 4.4.1. The analysis is done based on the size of the image/videos and the 2nd secret PIN level of degree.

Based on the 2nd Secret PIN entered, logic gate technique combination of array of bytes values and the inverted first secret code value happens and written them as series of bytes back into the image and it is repeated at each stage.

From the analysis it has been found 90% of the scenario, time taken for the decryption is faster than the time taken for the encryption. This shows while decrypt, ANT algorithm is efficiently used to find the path. Based on the 2nd pin value, the encryption and decryption time taken varies for the same image. Also understand, there are several other factors that consume time like CPU, Memory, system load, etc.

5. CONCLUSION

The entire research shows the unique way of securing the data at source itself before we share across. 2 PIN secret code authentication makes it hard for any hacker to decrypt the image. Wrong PIN corrupts data, so better to have a copy to retry. The ANT algorithm technique helps the decryption faster. Result comparison table shows how efficiently ANT algorithm technique is used.

Combination of these multiple techniques makes sure the data is safe and that reduces user stress and shall concentrate on other responsibilities.

My future scope is to increase the 2nd PIN digits values which would make this unique visual cryptography method even stronger. Even the images and videos with higher sizes would be considered for encryption and decryption. My research and contribution would continue towards future scope.

REFERENCES

[1] Ant algorithm for grid scheduling problem - IPP – BAS, Acad. G. Bonchev, bl.25A, by Stefka Fidanova & Mariya Durchova
[2] Wayner, P. : Disappearing Cryptography, Morgan Kaufmann Publisher, 2002
[3] Copyright protection scheme for color images using extended visual cryptography by Sonal Kukreja, Geeta Kasana, Singara Singh Kasana

A PEDESTRIAN COUNTING SCHEME FOR VIDEO IMAGES

Chi-Cheng Cheng and Yi-Fan Wu

Department of Mechanical and Electro-Mechanical Engineering,
National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C

ABSTRACT

Pedestrian counting aims to compute the numbers of pedestrians entering and leaving an area of interest based on object detection and tracking techniques. This paper proposes a simple and effective approach of pedestrian counting that can effectively solve the problem of pedestrian occlusion.Firstly, the moving objects are detected by the median filtering and foreground extraction with the improved mixed Gaussian model. And then the HOG (Histogram of oriented gradient) features detection and the SVM (Support vector machine) classification are applied to identify the pedestrians. A pedestrian dataset containing 1500 positive samples, 12000 negative samples, and 420 hard examples, which gave the false discriminant results with the initial classifier, also considered as negative samples to enhance classification capability is employed. In addition, the Kalman filtering with BLOB analysis for dynamic target tracking is chosen to predict pedestrian trajectory.This method greatly reduces the target misjudgment caused by overlapping and completes the two-way counting. Experiments on pedestrian tracking and counting in video images demonstrate promising performance with satisfactory recognition rate and processing time.

KEYWORDS

Machine Vision, Kalman Filtering, Pedestrian Identification, Target Tracking, Pedestrian Counting.

1. INTRODUCTION

Pedestrian identification and people counting recently gather attentions of researchers and engineers in the fields of machine vision and intelligent security monitoring. These technologies have been useful for marketing analysis and security purposes and can be found in theatres, markets, malls, department stores, exhibition halls, transportation stations, government buildings, personnel controlled laboratories and many other places. Although pedestrian counting can be accomplished by using inferred or laser technologies, machine vision possesses advantages of broad spatial coverage, accurate results, cost effectivity, and ability to provide much more information about pedestrians by extracting features of people. Therefore, this paper aims to develop an effective and efficient framework for pedestrian identification and bidirectional counting for potential future applications.

Basically, there are two difference approaches to deal with the challenge of people counting. The first approach performs people counting based on moving regions where pedestrian is treated as a single moving object [1]. If the size of a moving region is similar to that of a pedestrian, the region will be counted as a person. However, if the size of a moving region is greater than that of a single person, the moving region needs to be decomposed into a number of areas for single

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 135-149, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121211

person according to prior knowledge regarding the pedestrian size. These methods highly rely on prior knowledge and bring about characteristics of low accuracy and fast computation. Another approach is based on image features and machine learning techniques. People counting is therefore achieved by template matching through learning process with image samples of pedestrian [2]. Although these methods provide much better performance in terms of accuracy, expensively computational cost and requirement of a large amount of image samples are their drawbacks.

Methods for pedestrian identification can be classified into four categories. The human body model approach applies geometric features of human body to justify if the region of interest is pedestrian or not. Most common features for identification of human body include head [3], trunk [4], hair [5], shoulders [6], etc. Furthermore, a straight ellipse was suggested to model the outline of the human's body [7].

The template matching approach searches for similar objects in images to identify pedestrian according to given templates. A grid-based template matching algorithm was introduced for people counting [8]. Besides, the fast Hough transform was applied to quickly search for potential position of human's head in the foreground [9]. Nevertheless, human's features cannot be always described by simple mathematical rules and extensive learning through lots of image samples is strongly required. The outline classification approach conducts object identification based on trained templates through a learning process with lots of image samples. A multilevel HOG-LBP (Histogram of oriented gradient-Local binary pattern) scheme based on the PCA (principal component analysis) was proposed to identify the profile of human's head and shoulders [10]. A boosting learning algorithm identifying human's head, shoulders, trunk, and legs was presented for detection and tracking of humans [11]. In addition, a people counting system based on detection and tracking of human face with a neural network training process was proposed [12]. The LDA (linear discriminant analysis) was chosen to enhance learning capability to human's features [13].

The movement feature approach is to perform pedestrian identification and tracking based on periodic movement characteristics of humans. Human movements can be assumed to be independent events. As a result, similar moving patterns should belong to the same human. Pedestrian identification was achieved by tracking corners with similar moving patterns followed by classification with a Bayesian framework [14].

2. EXTRACTION OF MOVING TARGETS

The purpose of pedestrian counting is to calculate the numbers of people entering and leaving a given region. In order to conduct pedestrian counting, pedestrian identification and moving objects tracking are usually involved. This paper proposes a systematic approach including foreground extraction, objects detection, objects tracking, and pedestrian counting.

Because image quality is strongly affected by weather, illumination, electromagnetic interferences, and other possible noises. It is required to include image pre-processing to remove unnecessary signals and enhance image quality at the beginning stage. The median filter is selected for image preprocessing because of its outstanding performance of noise removal and edge reservation.

2.1. Foreground extraction

Foreground extraction is to isolate regions with pedestrians in the images for the purpose of improving computational efficiency by narrowing searching area. There exist a number of popular methods for foreground extraction such as temporal difference, background subtraction, the optical flow. The temporal difference locates the moving objects relying on the difference between two or three successive images. This method is highly based on the assumption of same background contents and can only be applied to images taken under invariant illumination environment [9, 15]. The background subtraction employs the difference between the image and the known background to achieve moving objects detection. Nevertheless, a reliable background image, which can adapt to variant background, plays a crucial role to successfully extract moving objects from the background [10,16]. The optical flow method extracts moving objects according to the derivatives of brightness stemming from the constraint of brightness consistency Although this approach demonstrates satisfactory detection performance for moving objects, static objects cannot be successfully identified [11,14].

In order to extract moving and static pedestrian in video images, a foreground extraction scheme based on template matching is therefore demanded. A hierarchical template matching algorithm using contour features was developed due to its significant detection performance and computational efficiency. Lots of samples as templates are required to reflect many possible postures of pedestrian. Nevertheless, reduction of computation speed will be resulted. As a result, a combined coarse-to-fine approach in shape and parameter space was proposed to enhance computational efficiency [12,17].

The single Gaussian model, which consists of background initialization and background update, is quite applicable to interior environment and uncomplicated outdoor space. However, the single Gaussian model may fail due to non-constant illumination, varying environment, and incursion and vanishing of unknown objects. Consequently, the mixture-of-Gaussians classification model was introduced to deal with both robustness to environment and real-time capability by Stauffer and Grimson [20,18]. Three stages, establishment of background model, identification of background model, and update of background model, are involved in this technique for foreground extraction.

At the establishment of background model stage, three important parameters regarding the Gaussian distribution including mean, standard deviation, and its weighting factor, are required to be determined. As for the identification of background model, assume there are K Gaussian distributions with corresponding weighting factors w_k. a priority list can be arranged according the ratio of the weighting factor to the standard deviation, w_k/σ_k. The weighting factor and the standard deviation indicate the duration time and the stability of the Gaussian distribution, respectively. If the sum of the weighting factors for the top B Gaussian distributions in the list is larger than a given threshold value T, these B distributions can be applied as the background model and the rest distributions will represent the foreground.

After the background model is established, if pixels of a new coming video sequence agree with the model, they will be classified as the background, otherwise foreground. A standard process to justify if a pixel belongs to the background can be formulated by

$$|x_t - \mu_{t-1}| < 2.5\sigma_{t-1}$$

where x_t is the pixel information at time t, and μ_{t-1} and σ_{t-1} stand for the mean value and the standard deviation of the Gaussian distribution at time t-1, respectively. The background model can therefore be updated by modifying corresponding parameters for each Gaussian distribution.

Unfortunately, this approach may suffer from update failure due to slow learning rate. Consequently, a novel learning rate formula and an online expectation-maximization (EM) algorithm were proposed to improve convergent speed and adaptation capability to variant environment [21,19].

It was found that the mixture-of-Gaussians classification scheme was able to provide complete contours of moving objects in the foreground. However, the result also contains partial contour on the background with noises especially for situation with complicated background. As a result, the following modified online EM algorithm is proposed to maintain complete contour information of moving objects in the background and eliminate fake foreground pixels as well as background noises.

1. Initially, apply the original online EM algorithm to obtain the background image I_b and the binarized foreground image I_f.
2. Find the difference image I_d by subtracting I_b from the current image I.
3. Scan every pixel in the image I_f. If the pixel belongs to the background, leave it unchanged; otherwise, fill it with the pixel information at the corresponding location of I_d.
4. Implementation of a thresholding process to the modified foreground image I_f.
5. Calculate boundary lengths of contours and the area size enclosed by contours. If either the boundary length or the area size is too small, remove the correspond contour.
6. The foreground image is updated by filling the contour with the foreground intensity.

Remarkable performance of foreground extraction on sequences of images taken from the Intelligent Room video in SBMnet (http://scenebackgroundmodeling.net) dataset and the PETS (Performance evaluation of tracking and surveillance) 2009 dataset is illustrated in Figure 1.

Figure 1. Foreground extraction using the proposed modified EM algorithm.

2.2. Objects detection

Objects detection is to identify pedestrian from the foreground as the targets for objects tracking afterwards. After the process of foreground extraction, moving objects are successfully identified. Nevertheless, moving objects may consist of pedestrian and other matters with movement.

Basically, there are four different approaches to deal with objects detection. They are model-based, template matching, posture classification, and movement-based methods. There have been a number of feature detection methods for pedestrian identification including the Harr-like technique, local binary patterns (LBP), histogram of oriented gradient (HOG), and pyramid histogram of oriented gradients (PHOG). In order to accommodate computational efficiency and detection performance, the approach of HOG is chosen for this research.

2.2.1. HOG

The approach of the histograms of oriented gradients locates features of local regions in an image.The followings are simplified HOG procedures:

1. Set the size for the window, which stands for the region of features extraction.
2. Compute the amplitude and the angle for the gradients in the window.
3. Partition the window into a number of overlapped blocks and decompose a block into un-overlapped cells. Establish the histogram for the angles of gradients in each cell.
4. Combine all histograms in a block to form a histogram vector.
5. A complete HOG features vector is constructed by collecting histogram vectors for all blocks

After the features of an image are extracted by the histograms of gradients, the technique of the support vector machine (SVM) will be applied to determine a best hyperspace as the decision function for classification.

2.2.2. SVM

The SVM aims to mapping feature vectors to a hyperspace so that a best hyperplane can be found for classification. Assume $\{-1,+1\}$ represents two different classes. Training samples are denoted by $\mathbf{x}_i$ and its corresponding output is $y_i \in \{-1,+1\}$. Since there are only two possible classes, the decision function for classification can be formulated by $\mathbf{w}^T\mathbf{x} + b = 0$. In order to allow the data can be classified by the maximum-margin hyperplane, i.e., the distance between the hyperplane and the nearest point $\mathbf{x}_i$ from either group is maximized, the constraint equation can therefore be defined as

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$$

Consequently, the desired function becomes

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T\mathbf{x} + b)$$

where $\mathbf{x}$ is the vector of sampled data and $\mathbf{w}$ is the normal vector to the hyperplane.

If the training data is separable, two parallel hyperplanes can be chosen to divide the data into two classes. Geometrically, the distance between these two hyperplanes can be found to be $2/\|\mathbf{w}\|$.Enlarge the distance of those two hyperplanes is actually equivalent to minimize $\|\mathbf{w}\|$. Therefore, this optimization problem can be simply expressed by

$$\min_{\mathbf{w},b} \frac{\|\mathbf{w}\|}{2} \text{ subject to} y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$$

The optimization problem listed above can be solved using the Lagrange multiplier approach by reformulating the problem as

$$\frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\}$$

where α_i is the Lagrange multiplier and *N* denotes the total number of data point. Since the minimal needs to be reached, the partial derivatives with respect to **w** and *b* must be zero and the following dual problem can be obtained.

$$\min_{\alpha_i} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \text{ subject to } \sum_{i=1}^{N} \alpha_i y_i = 0$$

After solving α_i and b, for any given vector *x* that needs to be classified, the object function becomes

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \text{x} - b) = \text{sgn}\{\sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} - b\}$$

2.2.3. Pedestrian Identification based on HOG and SVM

The process of pedestrian identification based on HOG and SVM is illustrated as Figure 2. The HOG is applied for features extraction and followed by the SVM classifier for pedestrian identification.

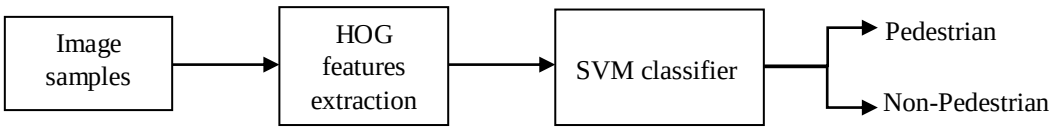


Figure 2. Flowchart of pedestrian identification based on HOG and SVM.

The positive pedestrian samples are from the INRIA pedestrian dataset, which contains 2416 pedestrian images with resolution of 64x128. Basically, the peoples in this dataset are all in standing posture, but with different appearance and clothing. Figure 3 depicts some image examples from the INRIA pedestrian dataset. There are 1218 negative pedestrian sample imageswith different sizes in the dataset. A set of 12180 negative images in total was collected by randomly cropping 10 64x128 regions from each sample. A number of negative pedestrian image samples are shown as in Figure 4. Through careful examination on those sample images, some of them, not appropriate for positive samples, were moved to the set of negative samples or even removed from the training data to enhance identification capability of the SVM classifier. Consequently, there are 1500 positive samples and 12000 negative samples in total in the training process.

Figure 3. Some positive pedestrian image samples from the INRIA dataset.

Figure 4. Some negative pedestrian image samples.

3. PEDESTRIAN TRACKING AND COUNTING

After successful foreground extraction and identification of pedestrian, pedestrian tracking and counting will be implemented by combining techniques of the Kalman filtering and BLOB (Binary large objects) analysis.

3.1. Kalman filtering [18,20]

The Kalman filtering has been an effective computational approach for tracking of a moving object. It provides a systematic recursive algorithm according to a state-space dynamic equation and an observation model including possible state estimation errors $\mathbf{w}_k$ and measurement noises $\mathbf{v}_k$, i.e.,

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k$$
$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ can be modelled as normal statistical distributions of $N(0, \mathbf{Q}_k)$ and $N(0, \mathbf{R}_k)$, respectively. $\mathbf{Q}_k$ and $\mathbf{R}_k$ denote the covariance matrices of the process noise and the observation noise.

The whole Kalman filtering scheme consisting of a prediction stage and an update stage can be illustrated as in Figure. 5. The prediction stage estimates the system's states $\hat{\mathbf{x}}_k^-$ and the covariance of the predicted error $\mathbf{P}_k^-$ before the measurement $\mathbf{z}_k$. The corresponding predicted error e_k^- is therefore defined as $\mathbf{x}_k - \hat{\mathbf{x}}_k^-$. At the update stage, the system's state and the covariance matrix of the predicted error are modified as $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ respectively when the new measurement z_k is received. $\mathbf{K}_k$ is known as the optimal Kalman gain. Both the prediction stage and the update stage are recursively executed to minimize the covariance matrix of the predicted error.

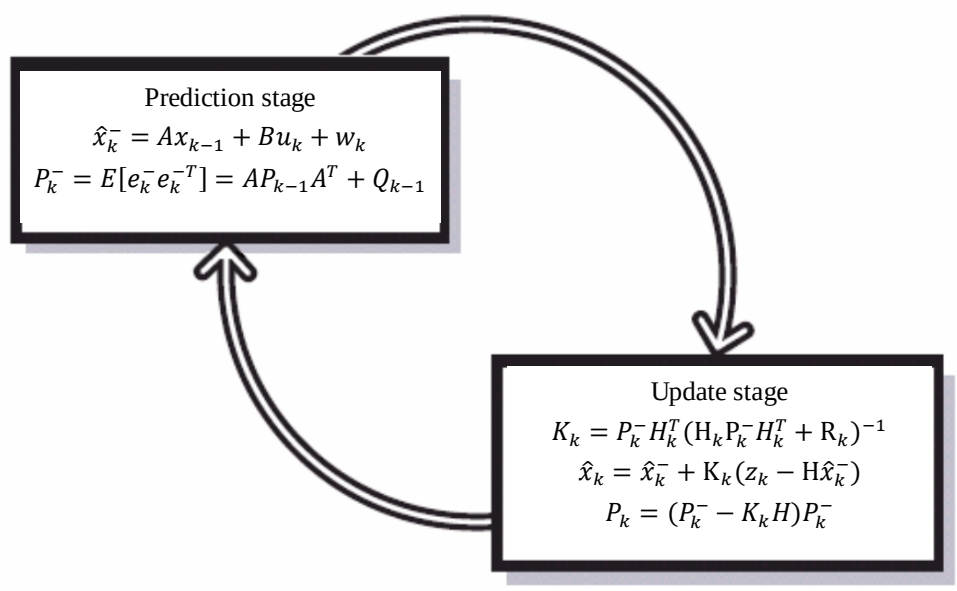


Figure 5. Two stages of the recursive Kalman filtering algorithm.

3.2. BLOB analysis

After the process of object detection, a sequence of foreground and background images are obtained. The foreground images can be represented by black-and-white binary images using the standard binarization method. As a result, the foreground consists of several connected regions. In order to acquire features of the foreground, some geometric properties such as location, size, perimeter, boundary length, and moment of inertia need to be calculated. Location and size of a connected region can be determined by its geometric center and the sum of number of pixels inside the region. The spatial moments can be applied for describing geometric characteristics of an image region because of its rotational invariant property. The general expression for the spatial moments $m_{p,q}$ can be written as

$$m_{p,q} = \sum B(x,y)x^p y^q$$

where $B(x,y)$ is the binary value either 1 or 0 at (x,y) in the image plane, and p and q stand for the order of the moment with respect to the x and y dimensions, respectively. If both p and q are all zeros, $m_{0,0}$ actually indicates the area in that image region.

The BLOB matching method takes advantage of image features such as shape, size, and the spatial moments to search for the target object and belongs to a bottom-to-top tracking approach. Apparently, this method is quite effective for rigid objects and limited number of moving objects. If the number of moving objects is large, it will be computationally expansive for the matching process. Nevertheless, the Kalman filtering technique predicts the moving object's position based on previous motion information of the moving object and is classified as a top-to-bottom tracking method. Since the Kalman filtering technique is able to predict the future position of the moving object, the BLOB matching algorithm is only applied to the nearby of the estimated location so that computational efficiency can be greatly improved. Therefore, pedestrian tracking and counting proposed in this paper will be achieved by a hybrid scheme combining both Kalman filtering and BLOB matching techniques.

Important procedures for pedestrian tracking are summarized as follows:

1. According the binary foreground generated by the improved Gaussian approach, search for connected regions. Mark each connected region as a BLOB and calculate its geometric center, contour length, and spatial moments as region features.
2. Based on current geometric centers of all BLOBs, estimate their locations for the next frame using the Kalman filtering method by defining

$$\mathbf{x}_k = [x, y, w, h, \dot{x}, \dot{y}, \dot{w}, \dot{h}]^T$$
$$\mathbf{z}_k = [x, y, w, h]^T$$

where x and y represent the coordinate of the geometric center, and w and h are the width and length of the BLOB. In addition, the control-input matrix **B** is a null matrix, and the state transition matrix **A** and the observation matrix **H** are respectively given as

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

3. Search for the matched BLOB with similar contour length and spatial moments near the estimated locations.

3.3. Pedestrian counting

Pedestrian counting is accomplished by computing the net number of pedestrian crossing a given screen line. Assume the origin of the coordinate system for the image is located at the upper left corner based on the conventional setup of the image reference frame and two end points of the screen line are defined as $P_1(a_1, b_1)$and $P_2(a_2, b_2)$. The n-th pixel along the screen line can be written as

$$P(n) = (a_1 + ncos\theta, b_1 + nsin\theta)$$

where

$$\theta = \tan^{-1}\frac{b_2 - b_1}{a_2 - a_1}$$

Let $F_r(k)$ denote the foreground image at a sampling instant. If there is a moving object on the left side of the screen line, its corresponding region can be expressed by

$$P_{fr} = \{F_r(i,j,k) | F_r(i,j,k) = 255, i_1 < i < i_2, j_1 < j < j_2\}$$

where i_1, i_2, j_1, and j_2 respectively stand for the top, bottom, left, and right limits of the region for the moving object. Once every pixel in $F_r(i,j,k)$ satisfies

$$i > a_1 + \frac{j - b_1}{\sin\theta} \cdot \cos\theta$$

It can be concluded that the moving object has successfully crossed over the screen line from left to right. Unfortunately, it would be quite complicated to judge the traveling direction of pedestrian if more than one pedestrian is involved. In order to overcome this difficulty, a screen line is extended to a screen strip with a certain width bounded by two parallel dashed line as shown in Figure 6. The vertical distance between those two dashed lines $2d$ is set to a little bit wider than the width of a normal people. In other words, those two dashed lines can be represented by

$$P(n) = (a_1 + ncos\theta \pm d, b_1 + nsin\theta)$$

When a pedestrian passes the left dashed border line of the screen zone, the current frame number will be recorded and accumulation of pixels on the dashed line for forthcoming frames will be conducted. When the sum maintains unchanged, there are two possible conditions. If the moving region is located at the left-hand side of the left dashed line, the pedestrian's moving direction is from the right to the left. However, if the moving region is within the left dashed line and the screen line, the moving direction of the pedestrian should be from left to right. Similar process can be applied to determine the moving direction of the pedestrian passes the right dashed border line. This approach solves the difficulty on determination of moving directions for pedestrians passing over the screen line from both sides at the same time.

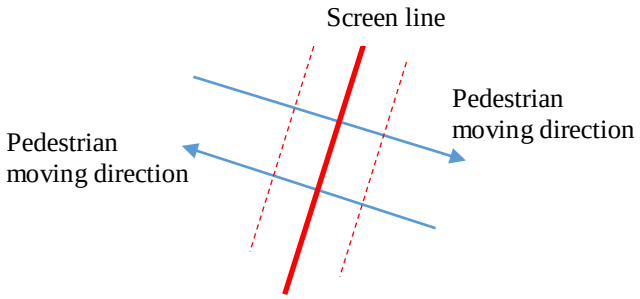


Figure 6. The screen line and two dashed border lines for pedestrian counting.

4. EXPERIMENTS

Experiments for pedestrian detection and counting were conducted by two prerecorded test videos from the internet and two live videos taken by authors. Software platform is the Visual Studio 2010 Ultimate with C++ programming language assisted by OpenCV 2.4.11 computer vision library. In addition, XviD codec was also applied for the purpose of video decoding. Before an experiment starts, a region of interest (ROI) with green line segments and a magenta screen line for pedestrian counting need to be given by users.

Two prerecorded test videos taken by monitoring cameras on the street were chosen for experiments and performance evaluation. Both test videos own 360P resolution. Test video #1 is with steady and little pedestrian flow without significant overlapping. Nevertheless, the pedestrian flow in test video #2 is much denser than that in test video #1. Besides, overlapped pedestrian frequently happens in test video #2. Figures 7 and 8 depict sample images in test video #1 and #2, respectively.

Live videos were taken by a Pantex digital single-lens reflex camera K-30 with 16.3-megapixel resolution. In order to have a sufficient height with an appropriate inclination angle downwards,

the camera was attached on a tripod standing on a table as illustrated in Figure 9. This setup was located on a hallway in campus of National Sun Yat-sen University. The height of the camera above the ground was 3.5 m and its depression angle below the horizontal level was 15 degrees so that the camera is able to capture the whole scene of the hallway. The live videos were set to 480P resolution. Sample images in live video #1 and #2 are depicted in Figures 10 and 11, respectively.

In order to evaluate the performance of pedestrian identification and counting, the accuracy η is defined as

$$\eta = 1 - \frac{|n_1 - n_2|}{n_1}$$

where n_1 and n_2 are actual and detected numbers of pedestrian, respectively. The average processing time is calculated based on the mean of computation time for five peoples by random selection.

Performance of pedestrian identification and counting for test videos is summarized in Table 1. Accuracy is a little bit reduced for test video #2 because of significant pedestrian overlaps. However, the proposed algorithm still demonstrates satisfactory performance in terms of accuracy. Without surprisingly, average processing time for test video #2 is greater than that for test video#1 due to more pedestrian involved in test video #2. A couple of errors were found in initial evaluation of pedestrian identification and counting for both live videos. The pedestrian images caused errors were therefore put into the group of negative samples for re-training. There exists two-way traveling direction for pedestrian in live video #2. The presented identification and counting strategy displays outstanding performance as shown in Table 2. Larger average processing time for live videos was mainly caused by better image resolution from 360P to 480P.

Figure 7. Sample images in test video #1.

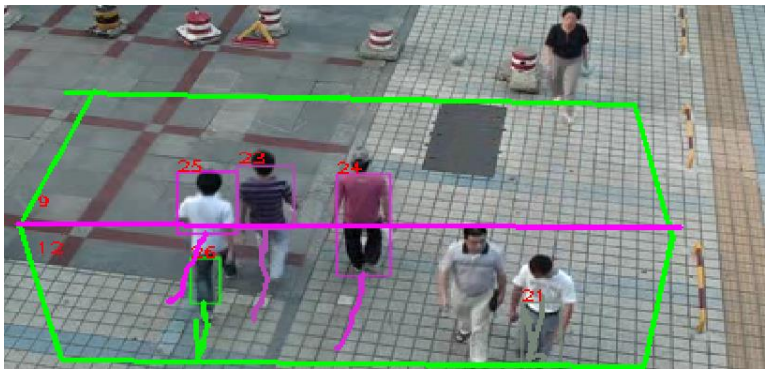


Figure 8. Sample image in test video #2.

Figure 9. Camera setup for experiments on pedestrian identification and counting.

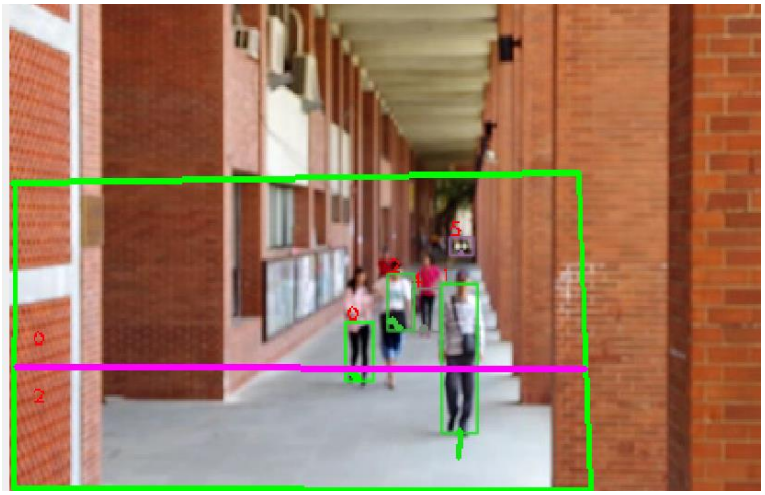


Figure 10. Sample image in live video #1.

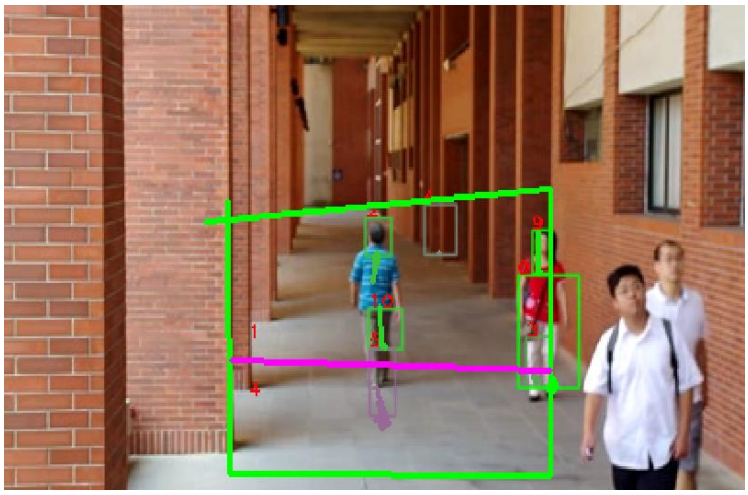


Figure 11. Sample image in live video #2.

Table 1. Performance of pedestrian identification and counting for test videos.

	Test video #1		**Test video#2**	
	Actual	Identified	Actual	Identified
Number of people entering	9	9	21	20
Number of people leaving	1	1	13	9
Accuracy	100%		82.8%	
Average processing time (ms)	35		60	

Table 2. Performance of pedestrian identification and counting for live videos.

	Live video #1		**Live video#2**	
	Actual	Identified	Actual	Identified
Number of people entering	6	6	5	5
Number of people leaving	0	0	1	1
Accuracy	100%		100%	
Average processing time (ms)	80		100	

5. CONCLUSIONS

This paper presents an effective and promising pedestrian counting scheme, which can be applied to areas required for control of people flow for the purposes of either security or marketing analysis. The proposed scheme consists of foreground extraction, pedestrian identification, pedestrian tracking, and counting of people flow. Foreground extraction is achieved by the improved mixed Gaussian model. Incorporating the HOG features detection with the SVM

classification is chosen for pedestrian identification. In order to enhance classification performance, those positive samples providing false discriminant results in the first classification run are moved to the group of negative samples. Furthermore, the Kalman filtering with BLOB analysis is employed to conduct dynamic target tracking for pedestrian trajectory prediction. Experiments on pedestrian tracking and counting for both dataset videos and live videos taken in campus environment demonstrate encouraging performance in terms of recognition rate and processing time. Above all, target misjudgment caused by overlapping can be greatly reduced and two-way counting becomes possible.

Nevertheless, further examinations on occlusion analysis and comparison with existing algorithms are required in the future work. Besides, in order to maintain portability for experimental setup, a notebook computer P770ZM with Intel® Xeon® E3-1231 v3 4x3.4 GHz was applied. If a more powerful desktop computer is chosen, real-time performance for actual applications can therefore be expected.

REFERENCES

[1] Li, F.S., Zhang, Y.C., Yang, H.C. & Wang, Y.P. (2014) "Fast pedestrians counting algorithm based on HOG", *Computer Systems & Applications*. Vol. 23, No. 5, pp. 172-176.

[2] Hsieh, J.W., Peng, C.S. & Fan, K.C. (2007) "Grid-based template matching for people counting",*Proc. IEEE 9th Workshop on Multimedia Signal Processing*, pp. 316-319.

[3] Vieren, C., Cabestaing, F. & Postarie, J.G. (1995) "Catching moving objects with snakes for motion tracking", *Pattern Recognition Letters*, Vol.16, pp. 679-685.

[4] Paviovic, V., Rehg, J., Cham, T.J. & Murphy, K. (1999) "A dynamic Bayesian network approach to figure tracking using learned dynamics models", *Proc. 7th IEEE Int. Conf. on Computer Vision*. Vol. 1, pp. 94-101.

[5] Zhao, M. (2008) "Hair-color modeling and head detection", *Proc. 7th World Congress on Intelligent Control and Automation*, pp. 7769-7772.

[6] Li, M., Zhang, Z.X. & Huang, K.Q. (2009) "Rapid and robust human detection and tracking based on omega-shape features", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2545-2548

[7] Zhao, T. & Nevatia, R. (2004) "Tracking multiple humans in complex situations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, pp. 1208-1221.

[8] Hsieh, J.W., Peng, C.S. & Fan, K.C. (2007) "Grid-based template matching for people counting", *Proc. IEEE 9th Workshop on Multimedia Signal Processing*, pp. 316-319.

[9] Yuk, J.S.C., Wong, K-Y.K., Chung, R.H.Y., Chin, F.Y.L. & Chow, K.P. (2006) "Real-time multiple head shape detection and tracking system with decentralized trackers", *Proc. 6th International Conf. Systems Design and Applications*, pp. 384-389.

[10] Zeng, C. & Ma, H. (2010) "Robust head-shoulder detection by PCA-based multilevel HOG-LBP detector for people counting", *Proc. Int. Conf. Pattern Recognition*, pp. 2069-2072.

[11] Wu, B. & Nevatia, R. (2007) "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors", *Int. J. Computer Vision*, Vol. 75, No. 2, pp. 247-266.

[12] Zhao, X., Dellandréa, E. & Chen, L. (2009) "A people counting system based on face detection and tracking in a video", *Proc. Int. Conf. Advanced Video and Signal Based Surveillance*, pp. 67-72.

[13] Yu, R (2014) "Mobile app connecting people based on personality detection and image perception analysis". *Proc. IEEE Int. Sym. Multimedia*, pp. 330-340.

[14] Brostow, G.J. & Cipolla, R. (2006) "Unsupervised Bayesian detection of independent motion in crowds", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 594-601

[15] Collins, R. (2000) *A System for Video Surveillance and Monitoring: VSAM Final Report*. Technical Report CMU-RI-TR-00-12, Carnegie Mellon University.

[16] Barnich, O. & Droogenbroeck, M.V. (2011) "ViBe: A universal background subtraction algorithm for video sequences", *IEEE Trans. Image Processin*g, Vol. 20, pp.

[17] Gavrila, D. (2003) "Pedestrian detection from a moving vehicle", *Proc. 6th European Conf. Computer Vision*, Vol. 2, pp. 37–49.

[18] Stauffer, C. & Grimson, W.E.L. (2000) "Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 747-757.
[19] Kaewtrakulpong, P. & Bowden, R. (2001) "An improved adaptive background mixture model for real-time tracking with shadow detection", *Proc. 2nd European Workshop Advanced Video-based Surveillance Systems*. pp. 149-158.
[20] Welch, G. & Bishiop, G. (2006) *An Introduction to Kalman Filter*, Department of Computer Science, University of North Carolina at Chapel Hill.

AUTHORS

Chi-Cheng Cheng was born in Taipei, Taiwan, R.O.C. He received the B.S. degree and the M.S. degree in power mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1981 and 1983, respectively, and the Sc.D. in mechanical engineering from Massachusetts Institute of Technology, Massachusetts, USA, in 1991. He is currently a Professor with the Department of Mechanical and Electro-Mechanical Engineering of the National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C. His research interests are in the areas of systemdynamics and control, machine vision, intelligent robots, and mechatronics.

Yi-Fan Wu was born in Qinhuangdao, Hebei, China. He obtained the B.S. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, Sichuan, China in 2014 and the M.S. degree in mechanical and electro-mechanical engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C. in 2016. His research interests include machine vision and automatic control.

A Novel Intelligent Image-Processing Parking Systems

Sree Veera Venkata Sai Saran Naraharisetti Benjamin Greenfield Benjamin Placzek Steven Atilho Mohamad Nassar and Mehdi Mekni

University of New Haven, West Haven, CT 06516, USA

Abstract. The scientific community is looking for efficient solutions to improve the quality of life in large cities because of traffic congestion, driving experience, air pollution, and energy consumption. This surge exceeds the capacity of existing transit infrastructure and parking facilities. Intelligent Parking Systems (SPS) that can accommodate short-term parking demand are a must-have for smart city development. SPS are designed to count the number of parked automobiles and identify available parking spaces. In this paper, we present a novel SPS based on real-time computer vision techniques. The proposed system provides features including: vacant parking space recognition, inappropriate parking detection, forecast of available parking spaces, and directed indicators toward various sorts of parking spaces (vacant, occupied, reserved and handicapped). Our system leverages existing video surveillance systems to capture, process image sequences, train computer models to understand and interpret the visual world, and provide guidance and information to the drivers.

Keywords:
Smart Cities, Car parking, Image Processing, Edge detection, Object Recognition

1 Introduction

By 2050, with the urban population more than doubling its current size, nearly 7 of 10 people in the world will live in cities [1]. Cities are major contributors to climate change. According to UN Habitat, cities consume 78% of the world's energy and produce more than 60% of greenhouse gas emissions. Yet, they account for less than 2% of the Earth's surface [2]. With regard to the United States, about 82.66% of the total population lived in cities and urban areas in 2020 [3]. As an impact of the growth of urban population, the number of land transportation vehicles in US has been increasing significantly.

Along with the increase of urban population, traffic jam and the number of parking spaces in many densely populated cities in the US become more

* corresponding author

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 151-167, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121212

problematic. Particularly in public places, the limited parking slots lead car drivers to slowly cruise the city, generating large amounts of exhaust emissions and creating traffic congestion. In addition, 86% of drivers face difficulty in finding a parking space in multilevel or geographically distributed parking lots [3]. Insufficient car park spaces lead to traffic congestion and driver frustration. Improper parking is also another parking-related issue that occurs when a driver parks on or outside of the lines of a parking space. This matter annoys other drivers and most of the time a driver who wants to park in a small leftover slot will give up and feel frustrated.

Although Closed-Circuit TeleVision (CCTV) systems help monitor parking, manually inspecting videos to recognize unauthorized parking behaviors is tedious and inefficient. Not only does it obstruct traffic and cause inconvenience, illegally parked vehicles pose economic risks [4]. Moreover, safety in parking and the need for real-time parking monitoring require the employment of applications and tools that track and record continuous activities including traffic and occupancy.

Potential solutions, such as adding more infrastructure and parking spaces, are not feasible due to the high cost and limited supply of commercial real estate in cities. Therefore, it is crucial to leverage technology advances and develop smart solutions to help drivers quickly locate unoccupied parking spaces. Implementing such smart systems will help resolve the growing problem of traffic congestion, wasted time, money, and energy. It will provide better public service, reduce car emissions and pollution, improve city visitor experience, increase parking utilization, and prevent unnecessary capital investments.

In this paper, we propose a novel image-processing Parking System, called Smart Park, to meet the critical short-term parking demand. Smart Park aims to convert traditional parking lots equipped with video surveillance systems into smart ones. The contributions of our work are: (1) monitor parking space utilization, (2) improve driver experience while decreasing drivers' frustration, (3) enhance parking lots security through number plate recognition, (4) collect valuable data for efficient parking management and informed decision making, and (5) assist drivers and recommend parking lots with respect to the spatio-temporal characteristics of an activity.

To validate and verify Smart Park, we modeled, analyzed, and experimented it on a selection of parking lots. Considering the complex spatial distribution of the parking lots, academic, administrative, service, health, and housing buildings, there is a critical need to provide a user-friendly platform to monitor, secure and efficiently navigate the campus for commuting students, faculty, staff, and visitors.

The reminder of the paper is organized as follows: Section 3 provides an extensive literature review of existing SPSs. Section 5 details Smart Park underlying software requirements engineering and software architecture and design models. Section 6 provides an overview of the obtained results. Finally, Sections 7 and 8 discuss the proposed SPS, highlight its advantages and limitations and conclude with future work.

2 Problem Statement

2.1 Difficulty in Finding Vacant Spaces

Quickly finding a vacant space in a multilevel parking lot is difficult if not impossible, especially on weekends or public holidays. A recent study shows that 86% of drivers face difficulty in finding a parking space in multilevel or geographically distributed parking lots [5]. Finding spaces during weekends or public holidays can take more than 10 minutes for about 66% of visitors. Stadiums or shopping malls are crowded at peak periods, and difficulty in finding vacant slots at these places is a major problem for customers [6]. Insufficient car park spaces lead to traffic congestion and driver frustration [7].

2.2 Improper Parking

If a car is parked in such a way that it occupies two parking slots rather than one, this is called improper parking. Improper parking can happen when a driver is not careful about another driver's rights. Sometimes improper parking occurs when a driver parks on or a bit outside of the lines of a parking space. The driver may notice his improper parking after leaving his car, but may not be willing to unlock his car, restart it, and adjust it to be inside the lines. This matter annoys other drivers and most of the time a driver who wants to park in a small leftover slot will give up and feel frustrated. Figure1 presents an improper parking situation.

2.3 Parking Fee Processing

Parking fee payment can be a time consuming activity for people. Since many current payment machines just accept small notes and coins, finding the exact amount and queuing for payment is not pleasant for drivers. Therefore, providing services that make payment convenient is desirable. One survey showed that queuing up for payment and finding coins for parking fee payment is troublesome. Moreover, most respondents agreed that using Touch'n'Go (a system that allows simply swiping a card and deduct fees from inside credit) is useful and will decrease queue up time[6].

2.4 Unauthorized Parking

Unauthorized or illegal parking is a ubiquitous problem in urban areas. Although many public areas have installed video surveillance systems (also knows as Closed-Circuit TeleVision (CCTV)) to help monitor the traffic conditions, manually inspecting these videos to recognize unauthorized parking behaviors is extremely tedious and inefficient [8]. Not only does it obstruct traffic and cause inconvenience to other drivers - particularly to those who need handicapped parking (HP) - illegally parked vehicles pose great economic risks [4].

2.5 Real-Time Parking Monitoring

Real-time monitoring is the employment of applications and tools that track and record continuous snapshots of your network's overall performance. Organizations use real-time monitoring to track network activity, improve network security, and identify potential problems as soon as they arise. Every business, regardless of size, can benefit from monitoring their network in real time.

2.6 Secure Parking

Secure Parking means an area to which cars can be parked in a secured place where there is on doubt of theft or stealing. Secured Parking should be interpreted by most as either being enclosed within walls and locked – or having security guards on permanent surveillance to protect the vehicles in an enclosed area.

3 Overview on vehicle detection in parking systems

The quality, efficiency, cost, and complexity of a smart parking system depends on the adopted detection technology [9]. Three types of detection technology, vision-based, sensor-based, and artificial intelligence-based are discussed in this section. Vision-based methods use CCTV — usually one camera is responsible for more than one parking space — and image processing software to detect parking space status. Sensor-based methods use one sensor for each individual parking space and might involve Global Positioning Systems (GPS), Global System for Mobile (GSM), or Bluetooth technologies. Finally, Artificial intelligence-based include solutions that involve multi-agent systems, machine learning, deep learning, fuzzy logic and neural network technologies.

3.1 Sensor Based SPS

Wireless Sensor Network (WSN) based SPS WSN can be defined as a network of wirelessly connected sensor nodes that are spatially dispersed and are dedicated to monitoring different environmental aspects such as sound, temperature, pressure, etc. WSN based sensor node comprises various sensors connected to monitor different aspects of the environment. In WSN, all the sensor nodes are connected to a sink node via wireless connection [10]. Nowadays, WSN has received outstanding traction among the SPS developers for flexibility, scalability, and low deployment cost.

Vehicular Ad-Hoc network (VANET) based SPS VANET is based on the Mobile Ad Hoc Network (MANET), where a wireless network of mobile devices is used. SPS utilizing VANET has three main components: Parking Side Unit (PSU), Road Side Unit (RSU), and On-Board Unit (OBU) [11]. The OBUs are installed on the vehicles, PSUs are installed on parking areas, and RSU's are

installed beside the roads near the parking areas. This type of system requires a trusted authentication authority that authorizes the vehicle's OBU. If a vehicle is parked inside of a smart parking facility, the OBU of the vehicle provides information to the PSU that the parking lot is booked. Then, this information is transferred to the RSU from the PSU. The vehicles traveling by that road where the RSU is placed can get the information of parking lot occupancy through their OBUs. VANET based smart parking systems are commonly deployed in both closed and open parking lots.

Internet of Things (IoT) based SPS IoT is the buzzing technology of the current era, where all devices are interconnected with one another through the internet. These devices can be computational devices, mechanical devices, and digital devices. They can transfer data to without human-to-human or human-to-computer interaction [12]. IoT technology acts as one of the primary key technologies that developers use for SPS. In IoT-based SPS, all the sensors and computational devices are connected through the internet and can transfer data without any human intervention. The internet connection among sensors, computational devices, and storage units can be either through a wired connection or through a wireless connection.

Global Positioning System (GPS) based SPS GPS is an essential component of different smart parking approaches. But GPS alone is unable to gather parking lot occupancy status and provide other smart parking facilities. However, GPS can provide a vehicle guidance facility for the user to drive towards vacant parking lots. From GPS data, many systems can forecast parking lot occupancy and road traffic congestion using CNN or DL algorithms [13]. The accuracy of GPS depends on the number of receivers it has. For a single frequency receiver GPS, the accuracy is around 7.8 m. On the other hand, a two-frequency receiver provides around 0.715 m of accuracy. The GPSdata is also prone to error if operated inside of a closed parking area. Thus, smart parking systems that use GPS are suitable for open parking lots [14].

Global System for Mobile (GSM) based SPS GSM is a standard for second-generation (2G) digital cellular networks. GSM standard provides a subsidiary service called SMS. SPS, based on GSM, uses SMS service to reserve parking spots at different parking spaces. Some system also generates unique codes for the users during the reservation process, which are used to authenticate the reservation and ensure that only the designated persons get to park [15].

Bluetooth based SPS Bluetooth is a wireless communication technology standard that enables data transfer within a short-range. A smart parking system that is wholly based on Bluetooth technology usually has automated valet parking installed. Regular SPS, which does not deploy an automated valet parking

facility, requires additional sensors and approaches to get different smart parking facilities [16]. Many smart parking systems use the Crowd-sensing method to gather information about available parking spots in an area. The method uses smartphone sensors (such as Accelerometer, Gyroscope, Magnetometer, and GPS) and applications to gather parking lot information [17].

3.2 Artificial Intelligence Based SPS

Multi-Agent System (MAS) based SPS MAS is a self-organizing computer-based system accumulating multiple intelligent agents to solve problems that are pretty difficult for any single system to solve [18]. To develop SPS, various researchers have deployed MAS due to its effectiveness in both closed or indoor and outdoor or open parking lot areas. A significant portion of MAS-based SPS provides computing facilities to the agents, which reduces the data transmission head of the whole system. As a result, the power consumption rate decreases.

Machine learning (ML) based SPS ML is a subset of AI that provides a system the ability to learn and improve on a particular task from the datasets or experiences without explicitly programming the system [88]. A machine learning-based SPS analyses the parking lot of data to extract the parking lot status. More- over, ML and AI-based SPS can predict parking lot occupancy status of the upcoming days, weeks, or even months and provide a dynamic pricing scheme. ML-based systems can monitor traffic congestion of particular roads and offer a smart solution to smart parking spaces [19].

Deep learning (DL) based SPS DL is a subset of ML and a function of AI which mimics the human brain in terms of data processing and feature extraction to make decisions [20]. DL algorithms detect vacantly occupied and special parking lots in an SPS instead of regular sensors, which reduces the number of sensors and cameras required by the system. DL is also used to predict parking lot occupancy.

3.3 Neural Network (NN) based SPS

NN is a combination of algorithms that extracts features and under- lying relationships from sets of data through a process that mimics human brain function [21]. In SPS, NN is used for license plate recognition using real-time video data. CNN and machine vision are implemented to detect parking lot occupancy status. CNN's are also capable of providing road traffic conditions of different routes [22].

Fuzzy logic based SPS Fuzzy logic is a reasoning method that resembles human reasoning. It uses multi-valued logic, which means there is no absolute truth

or ab- solute false value in fuzzy logic [23]. Fuzzy logic is used in SPS for predicting parking lot occupancy status. But the accuracy of the prediction model based on Fuzzy logic would not be that high without validating the prediction result with the real-time data. Therefore, Fuzzy logic, along with machine vision or sensors, improves the accuracy of the overall system [24].

3.4 Computer Vision/Image Processing based SPS

Computer vision/Image processing based SPS uses different types of camera networks to use image data to extract different information such as parking lot occupancy status, license plate recognition (LPR) and face recognition for billing, security issues, and to provide road traffic congestion report [25]. The systems based on computer vision/image processing technologies usually have a high data transmission rate from the camera network to the processing units because these systems are dependent on real-time parking lot video data for feature extraction [26]. These sorts of SPSs are usually suitable for open parking areas because a single camera can capture a significant area in the parking lot. However, these systems are prone to occlusion, shadow effects, distortion, and changing of light.

4 Methodology

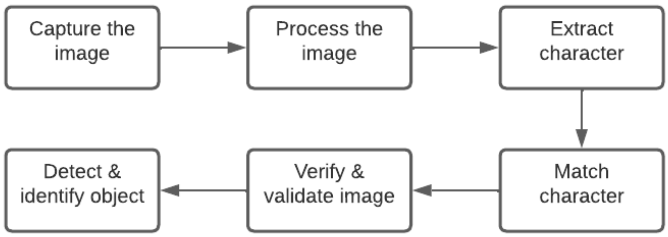


Fig. 1: Methodology Diagram

The process of availability detection relies on the following multiple-step methodology as shown in Figure 1.

1. *Capture the image*: To continuously get the parking lot status, the proposed system takes the live feed from pre-installed monitoring cameras. The cameras are connected to FTP Servers.
2. *Process the Image*: The live feed is then pre-processed frame by frame to reduce the noise using Gaussian blur combined with grayscale filtering techniques.

3. *Extract characters*: To recognize the characters on license plate, the proposed system applies Image Segmentation technique. It extracts the value channel from the HSV format of the plate's image. HSV is a cylindrical colour model that remaps the RGB primary colours into dimensions that are easier for humans to understand. Next, it applies adaptive thresholding on the plate's value channel image to reveal the characters of the image.
4. *Match characters*: After extracting the characters separately for character matching, the proposed system uses Optical Character Recognition (OCR) [27] to recognize the character one by one. OCR comes with pre trained model to classify the characters from the images.
5. *Verify and validate images*: After classifying characters, the predicted characters are validated with the test data. Such a validation is done by taking the output of an OCR run for an image and comparing it to the original version of the same text. The verification and validation include character level accuracy as well as word level accuracy.
6. *Detect and identify objects*: After Verifying and validating the image, the proposed system identifies the object using colour processing which are use as primary filtering to eliminate the unrelated colour or object in the image. Besides that, shape detection use the edge detection (Circular Hough Transform (CHT)). Finally, to detect objects, the system uses Deep learning techniques (Convolutional Neural Networks) to automatically learn an object's inherent features and characteristics.

5 System implementation

In this section, we detail the followed steps to support the Software Development Life-Cycle (SDLC) of the proposed Smart Parking System. First, we present the software requirement engineering process and highlight the key system requirements. Next, we provide an overview on the system design and architecture.

5.1 Software Requirements Engineering

The proposed Smart Parking System has been designed from the perspectives of requirements identified by two key actors; (1) User (Car Driver) and (2) Public Safety (Police). Moreover, with respect to the complexity of SPS, several services are also involved for image processing and data analysis and prediction. The use case diagram depicted in Figure 2 shows how different users with different roles interact with the system.

Requirements describe the characteristics that a system must have to meet the needs of the stakeholders. These requirements are typically divided into functional and non-functional requirements. Functional Requirements [FR] describe how a software must behave and what are its features and functions. Non-Functional Requirements [NFR] describe the general characteristics of a system. They are also known as quality attributes. The following is a selection of functional requirements:

Table 1: Business Use Case Scenarios

Use case Name: View Current Availability
Preconditions: Image processing has been completed and data has been sent to the app for formatting.
Main Sequence:

1. Images are collected from parking lots
2. Images are analyzed using object detection libraries to determine availability
3. Results are sent to the app and are displayed as a heat map
4. User opens the app and views parking information

Exceptions:

1. There is no parking data to be viewed **Solution:** User is prompted to login instead of register
2. The app cannot connect to the server
   Solution: Alert the user that there is an error with communicating to the server

Outcomes: The user is able to open the app and view parking availability for different parking lots.

Use case Name: View Future Predictions
Preconditions: Image processing has been completed and data has been fed through prediction algorithms.
Main Sequence:

[topsep=0pt]Images are collected from parking lots Images are analyzed using object detection libraries to determine availability Data is recorded and fed to prediction algorithms Prediction algorithms use past and present data to create parking availability predictions Forecast results are displayed

Exceptions:

[topsep=0pt]User opens the app and views future predictions **Solution:** Alert the user that there are currently no predictions The app cannot connect to the server
Solution: Alert the user that there is an error with communicating to the server

Outcomes: The user is able to open the app and view future predictions of parking availability for different parking lots.

Use case Name: Receive Parking Violation Notices
Preconditions: Automatic Number Plate Recognition (ANPR) scans license plates as cars drive into parking lots.
Main Sequence:

[topsep=0pt]ANPR localizes plates in images and performs Optical Character Recognition (OCR) to read license plate numbers Captured plate numbers are compared to the parking permit database Observed violations trigger a message to the campus police department and to the owner of the vehicle

Exceptions:

[topsep=0pt]No parking violation occurs **Solution:** No action is taken The app cannot connect to the server
Solution: Alert the user that there is an error with communicating to the server

Outcomes: Campus police and users committing violations are notified of said violations.

Use case Name: Report Closures or Other Issues
Preconditions: The user is in the app.
Main Sequence:

[topsep=0pt]User navigates to the report section of the app to chose type of issue and leave a comment Report is received by the server and updates parking conditions based on the information The app is updated to reflect changes

Exceptions:

[topsep=0pt]The user could report false information **Solution:** Check for similar reports to confirm information and/or manually review the report The app cannot connect to the server
Solution: Alert the user that there is an error with communicating to the server

Outcomes: Users are able to report issues related to parking on campus to cause the app to update information.

Use case Name: Parking Assistant
Preconditions: Image processing has been completed and there is parking availability information in the app to use.
Main Sequence:

[topsep=0pt]User navigates to parking planner in the app and chooses where their class is located. Based on parking availability, user's parking pass type, and parking lot locations, and class location, the app determines optimal parking location. The user is shown a recommendation on where to park.

Exceptions:

[topsep=0pt]Parking availability information is not available. **Solution:** The user is alerted of this error, and is recommended a parking spot based on location The app cannot connect to the server.
Solution: Alert the user that there is an error with communicating to the server

Outcomes Users are able to find a parking spot closest to their destination based on multiple factors.

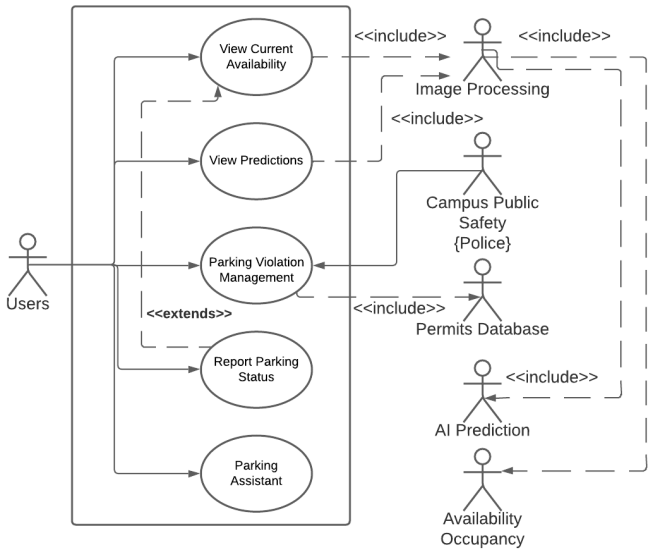


Fig. 2: SPS Use Case Diagram

1. [FR1] SPS shall allow users to view current availability in a parking lot.
2. [FR2] SPS shall compute and display future predictions to users.
3. [FR3] SPS shall detect illegal and unauthorized parking, issue Parking Violation Notices, and allow public safety and users to view and process them.
4. [FR4] SPS shall track parking traffic and identify vehicles for parking safety purposes.
5. [FR5] SPS shall assist users to identify the best parking considering the location of the scheduled activities.

The above listed functional requirements (FR) have been analyzed and validated with stakeholders and the following set of quality attributes non-functional requirements (NFR) has been derived:

- [NFR1]Performance: SPS will collect CCTV images, process, analyze, and store them every minute to keep parking information relevant.
- [NFR2]Security: SPS will use identification and authentication techniques to read Parking permit registration data and encryption techniques to securely store extracted vehicles identification data.
- [NFR3]Portability: SPS will be mobile accessible through apps and web browsers.

5.2 Software Architecture

The software architecture style of the proposed SPS relies on a service-oriented architecture. The interactions between services are depicted in Figure 3. The components of SPS architecture are:

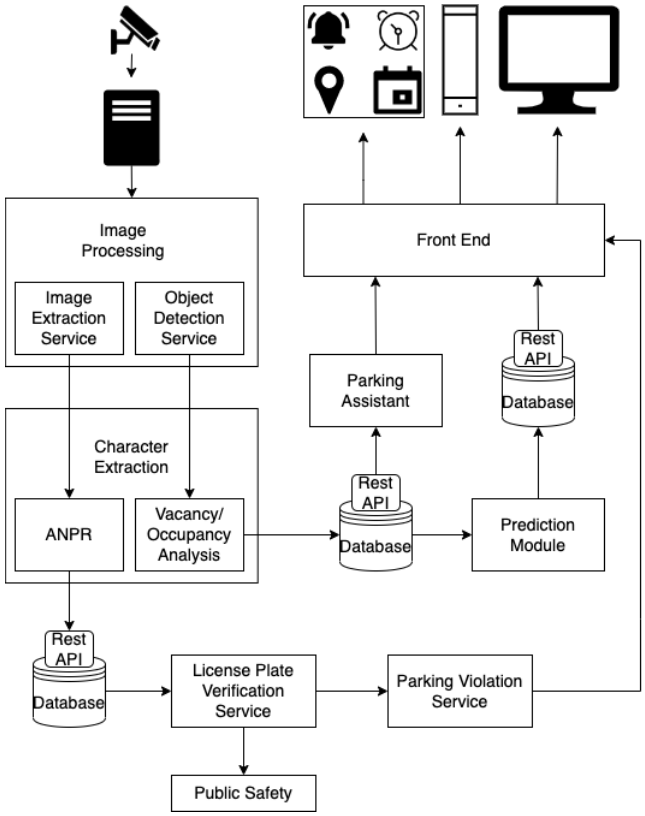


Fig. 3: SPS Software Architecture Diagram

– FTP Server: It gathers parking images from surveillance cameras which are installed in the parking lots.
– Image Processing: The image processing module collects, analyzes and processes data extracted from the FTP Server. This module has two key services:
  • The Image Extraction Service: This service acts as medium to extract the images from the recorded videos which are installed in parking lots.
  • The Object Detection Service: This service is used for the detection of objects. It generates a number n of things and their locations. This service detects and captures the vehicles in the parking spots. Objects include vehicles, parking lines (to detect illegal parking), handicap symbols (to detect unauthorized parking).
– Character Extraction: This module focuses on interpreting images and detected objects.
  • Automatic Number Plate Recognition (ANPR): The ANPR algorithm is applied to extract the plate numbers of the parked cars (See Figure 7). The data collection process consists of accumulating of images and bounding-boxes for training the machine learning model. This module

uses ML/AI based learning techniques using the collected dataset of plate images taken from different positions. The Optical character Recognition (OCR) approach is used to convert images of text into machine-encoded text. The goal is to build a model that can recognize and localize the plates.

- Vacancy/Occupancy Analysis: This service provides information on vacant spots and updates this list as cars enter or leave the parking.
- License Plate Verification: The goal of this phase is to verify and validate plate numbers using the public safety database.
- Parking Violation Service: This service is responsible to issue parking violation notices when improper (illegal or unauthorized) parking is detected.
- Prediction Module: This service extracts the information from the database and predicts the occupancy/availability of the parking based on current status and historical data.
- Parking Assistant: This service provides recommendation about the best parking lot to use while maintaining a trade off between availability and distance/time to a specific location.

6 Results

We designed, implemented, and tested our SPS in 2 parking lots . The existing CCTV deployed system uses FTP compatible cameras. We implemented a web-based application to support our students, staff, faculty, and visitors.

Our SPS successfully detected and reflected the count of parking spots at a rate of 92% accuracy (See Figure 6).

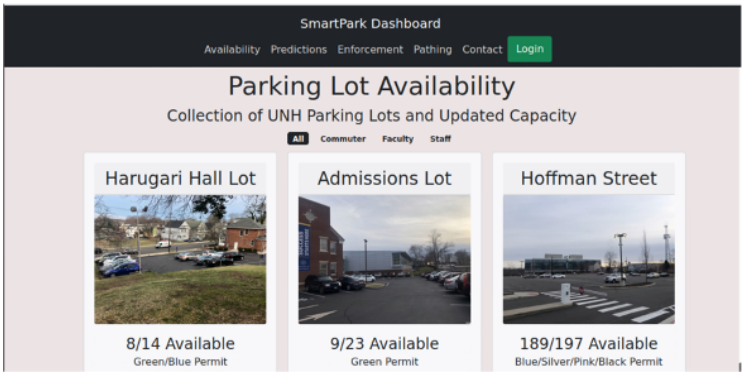


Fig. 4: Smart Park Dashboard: Parking lots availability

We built a prediction model for each tested parking lot to provide a live-feed, compute, and visualize occupancy (Figures 6a and 6b) and a dashboard for

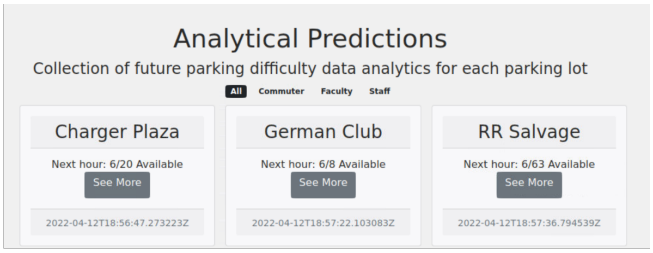


Fig. 5: Smart Park Analytical Predictions Portal

each parking lot which has its own prediction model and Predicted graphs (See Figures 5 and 8).

As far as parking lot monitoring and vehicle identification, we are able to identify and extract local Connecticut license plates information at a 81% success rate (Figure 7).

Thanks to the vehicle identification information, SPS sends a query to the public safety services to verify if a valid parking permit has been issued. We have created a dashboard where we can see all the availability of parking lots (See Figure 4).

(a) (b)

Fig. 6: Smart Park Object Detection and Identification

Finally, our SPS proposes a unique feature: *Parking Assistant*. Considering an activity start time, day of the week, location, and permit type, SPS recommends a parking lot with a convenient walking distance and a high-level confidence of availability.

To illustrate this feature, let us consider the scenario of a course taking place at Kaplan Hall and starting at 8:20am for a user holding a student permit type. First, Smart Park identifies, the parking lots that are at a reasonable walking distance from Kaplan Hall. Next, it computes the availability predictions associated with these lots.

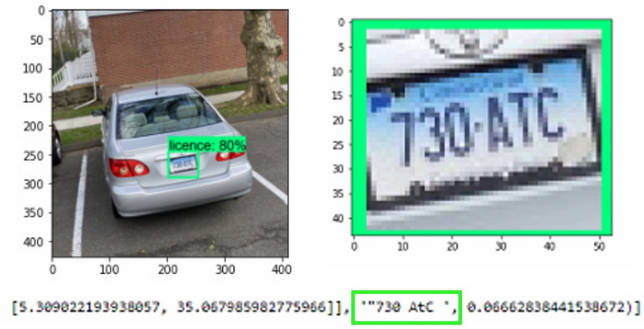


Fig. 7: Smart Park Automatic Number Plate Recognition

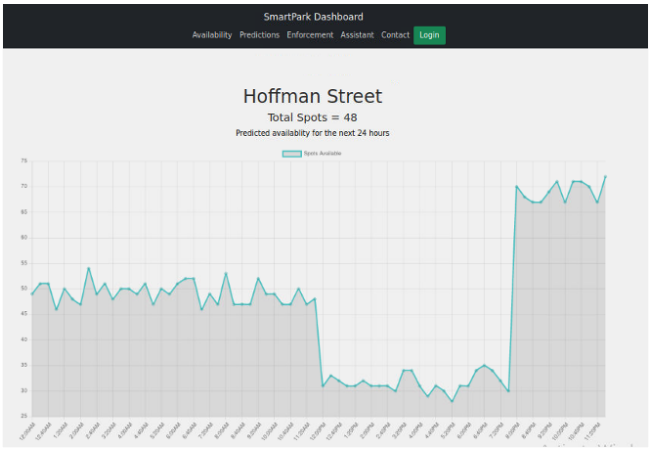
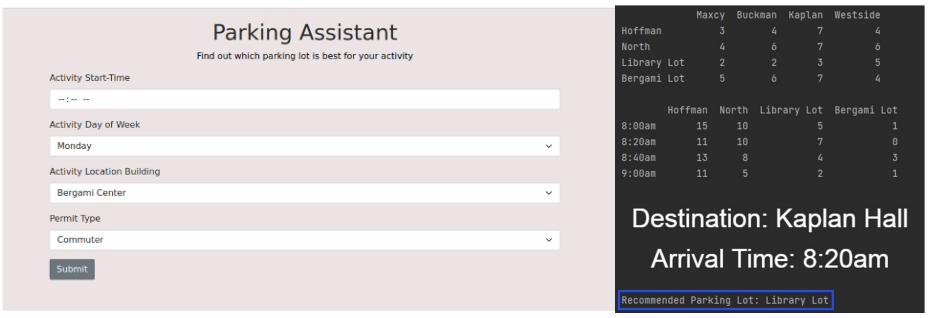


Fig. 8: Smart Park: Predicted Availability Graph

7 Discussion

A number of experiments was conducted and it was confirmed that the image processing algorithms and character extraction techniques work better under good weather conditions, daylight, high contrast between identified objects and their background, and from certain angles of camera view. In fact, under snow or heavy-rain weather conditions, the object detection accuracy can drop as low as 32%. Similarly, the precision of the license plate recognition algorithm drops to 18%. Additional verification and validation will be required to assess the scalability of the proposed solution. Moreover, the machine learning algorithm used to read license plates is trained over a data set of 100's of license plates. it is important to extend the training data set to allow SPS to comprehensively recognize and read US license plates. Although the implemented features of the proposed SPS are quite important and mainly focus on parking lot traffic

monitoring, occupancy analysis, vehicle identification and tracking, illegal and unauthorized parking, and illegal parking notice management, there are other value-added features that are still missing such as accident detection in parking lots.



(a) (b)

Fig. 9: Smart Park :(a) Parking Assistant Dashboard, (b) Recommended Parking Lot

8 Conclusion and Future Work

In this paper, an image processing based smart parking system was presented. The proposed image processing based smart parking system has four advantages. First, special hardware infrastructure is not necessary because a CCTV camera can cover large parking spaces. Second, the system can provide an accurate occupancy prediction that is essential for finding the vacant parking. Third, camera-based system can also be applied to the parking lot in the street or residential area. Fourth, the assistant parking feature is a time-space optimization solving that improved parking users experiences. However, such camera-based parking are still vulnerable to accidents that may occur. In this respect, the proposed SPS needs additional features to detect an accident and avoid hit and run situations. Future research following this project will focus on accident detection and processing of parking violation notices. In addition, research to improve the detection accuracy and the processing speed will be performed.

References

1. The World Bank, "The World Bank," 2022, [Online; accessed January 01, 2022]. [Online]. Available: https://www.worldbank.org/en/topic/urbandevelopment/overview1

2. United Nation : Climate Action, "Cities and Pollution," 2022, [Online; accessed January 01, 2022]. [Online]. Available: https://www.un.org/en/climatechange/climate-solutions/cities-pollution
3. Statista, "Degree of urbanization in the United States from 1970 to 2020," 2022, [Online; accessed January 01, 2022]. [Online]. Available: https://www.statista.com/statistics/269967/urbanization-in-the-united-states/
4. I. H. Chowdhury, A. Abida, and M. M. H. Muaz, "Automated vehicle parking system and unauthorized parking detector," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*. IEEE, 2018, pp. 542–545.
5. A. Kianpisheh, N. Mustaffa, J. M. Y. See, and P. Keikhosrokiani, "User behavioral intention toward using smart parking system," in *International Conference on Informatics Engineering and Information Science*. Springer, 2011, pp. 732–743.
6. D. Bong, K. Ting, and K. Lai, "Integrated approach in the design of car park occupancy information system (coins)." *IAENG International Journal of Computer Science*, vol. 35, no. 1, 2008.
7. A. Kianpisheh, N. Mustaffa, P. Limtrairut, and P. Keikhosrokiani, "Smart parking system (sps) architecture using ultrasonic detector," *International Journal of Software Engineering and Its Applications*, vol. 6, no. 3, pp. 55–58, 2012.
8. W. Chen and C. K. Yeo, "Unauthorized parking detection using deep networks at real time," in *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2019, pp. 459–463.
9. K. Yamada and M. Mizuno, "A vehicle parking detection method using image segmentation," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 84, no. 10, pp. 25–34, 2001.
10. S. Rani, R. Maheswar, G. Kanagachidambaresan, and P. Jayarajan, *Integration of WSN and IoT for smart cities*. Springer Nature, 2020.
11. D.-B. Nguyen, C.-R. Dow, and S.-F. Hwang, "An efficient traffic congestion monitoring system on internet of vehicles," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
12. S. Kumar and S. Swaroop, "Collateral development of invasive pulmonary aspergillosis (ipa) in chronic obstructive pulmonary disease (copd) patients," in *Recent Developments in Fungal Diseases of Laboratory Animals*. Springer, 2019, pp. 111–118.
13. J. Guo, Y. Liu, Q. Yang, Y. Wang, and S. Fang, "Gps-based citywide traffic congestion forecasting using cnn-rnn and c3d hybrid model," *Transportmetrica A: transport science*, vol. 17, no. 2, pp. 190–211, 2021.
14. K.-C. Lan and W.-Y. Shih, "An intelligent driver location system for smart parking," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2443–2456, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417413007987
15. Y. Rahayu and F. N. Mustapa, "A secure parking reservation system using gsm technology," *International Journal of Computer and Communication Engineering*, vol. 2, no. 4, p. 518, 2013.
16. M. Lewandowski, B. Płaczek, M. Bernas, and P. Szymała, "Road traffic monitoring system based on mobile devices and bluetooth low energy beacons," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
17. C. Wan, J. Zhang, and D. Huang, "Scpr: secure crowdsourcing-based parking reservation system," *Security and Communication Networks*, vol. 2017, 2017.
18. S. Belkhala, S. Benhadou, K. Boukhdir, and H. Medromi, "Smart parking architecture based on multi agent system," *Int. J. Adv. Comput. Sci. Appl*, vol. 10, pp. 378–382, 2019.

19. S. J. Kamble and M. R. Kounte, "Machine learning approach on traffic congestion monitoring system in internet of vehicles," *Procedia Computer Science*, vol. 171, pp. 2235–2241, 2020, third International Conference on Computing and Network Communications (CoCoNet'19). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920312321
20. M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," *IEEE Transactions on Intelligent transportation systems*, vol. 21, no. 8, pp. 3152–3168, 2019.
21. G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, "Car parking occupancy detection using smart camera networks and deep learning," in *2016 IEEE Symposium on Computers and Communication (ISCC)*, 2016, pp. 1212–1217.
22. T. Pamula, "Road traffic conditions classification based on multilevel filtering of image content using convolutional neural networks," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 3, pp. 11–21, 2018.
23. H. M. Kammoun, I. Kallel, J. Casillas, A. Abraham, and A. M. Alimi, "Adapt-traf: An adaptive multiagent road traffic management system based on hybrid ant-hierarchical fuzzy model," *Transportation Research Part C: Emerging Technologies*, vol. 42, pp. 147–167, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X14000692
24. X. Xie, C. Wang, S. Chen, G. Shi, and Z. Zhao, "Real-time illegal parking detection system based on deep learning," in *Proceedings of the 2017 International Conference on Deep Learning Technologies*, 2017, pp. 23–27.
25. J. Zheng, "Augmented deep representations for unconstrained still/video-based face recognition," Ph.D. dissertation, University of Maryland, College Park, 2019.
26. R. S. Dangi, A. Kuvelkar, S. K. Maity, and S. Wandhekar, "Efficient and robust indian number plate recognition through modified and tuned lprnet," in *ICDSMLA 2020*. Springer, 2022, pp. 115–129.
27. J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten optical character recognition (ocr): A comprehensive systematic literature review (slr)," *IEEE Access*, vol. 8, pp. 142 642–142 668, 2020.

DETECTION OF ROAD TRAFFIC CRASHES BASED ON COLLISION ESTIMATION

Mohamed Essam, Nagia M. Ghanem and Mohamed A. Ismail

Department of Computer Engineering,
Alexandria University, Alexandria, Egypt

ABSTRACT

This paper introduces a framework based on computer vision that can detect road traffic crashes (RCTs) by using the installed surveillance/CCTV camera and report them to the emergency in real-time with the exact location and time of occurrence of the accident. The framework is built of five modules. We start with the detection of vehicles by using YOLO architecture; The second module is the tracking of vehicles using MOSSE tracker, Then the third module is a new approach to detect accidents based on collision estimation. Then the fourth module for each vehicle, we detect if there is a car accident or not based on the violent flow descriptor (ViF) followed by an SVM classifier for crash prediction. Finally, in the last stage, if there is a car accident, the system will send a notification to the emergency by using a GPS module that provides us with the location, time, and date of the accident to be sent to the emergency with the help of the GSM module. The main objective is to achieve higher accuracy with fewer false alarms and to implement a simple system based on pipelining technique.

KEYWORDS

RTCs, ViF, SVM, Deep Learning, Collision Estimation.

1. INTRODUCTION

Nowadays, the usage of vehicles increases with the corresponding rise in population. Consequently, accidents are increasing as well due to different reasons. The world health organization (WHO) states that RTCs are in the top 10 reasons that cause death; there are more than 1.35 people die and 50 million injuries each year because of RTCs [1][2]. In addition, in Egypt, there are nearly 12000 Egyptians die, and thousands of people injure because of RTCs. RTCs will increase continuously to become the top cause of death by 2030 [3]. The main causes of accidents are due to over speed, driver inattention, blown tire, and wrong passing as shown in Figure 1. Also, the delay in reporting the accident and the delay in reaching the ambulance to the accident location are considered to be one of the main reasons [4]. In addition, RTCs that happen in remote places are very difficult to be traced. So, it's a challenge for the emergency services to get to the exact location of the accident which results in death. Fortunately, governments in developed and upper-middle-income countries installed a large number of CCTV cameras on roads. For example, in China, 200 million CCTV cameras have been installed [5]. In London city, there are 500 thousand CCTV cameras [6] and with the availability of the huge processing power of computers nowadays, such as cloud computing services and relatively affordable hardware. This paper aims to detect RTCs depending on CCTV cameras installed, report to the emergency services in real-time in order to recover victims, and allow them to monitor the accident using a client-server architecture and an interactive GUI.

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 169-179, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121213

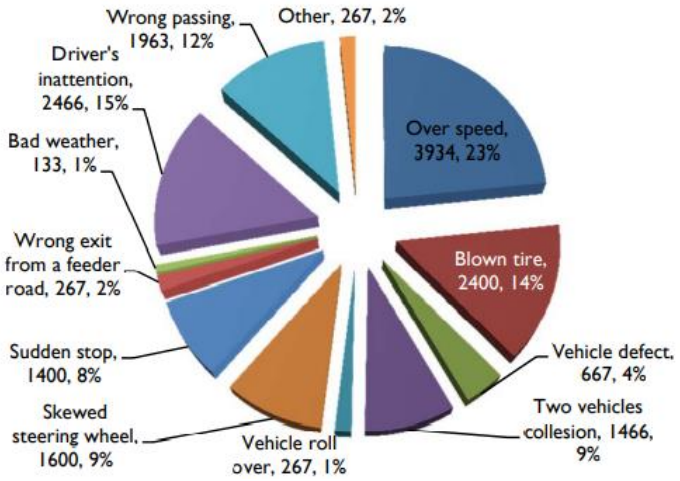


Figure 1. Number & Percentage of main causes of crashes in Egypt

2. RELATED WORK

Many authors have worked on the topic of RTCs. Some of them worked by using deep learning like this one [7]. However, the deep learning method is not the best solution because of the shortage of video crash datasets. In [7], they used a very low number of videos which makes it difficult to solve a complex problem like RTCs detection. Moreover, the accident datasets need to cover different types, different scenarios, and different ways that need to be fed to the neural networks, the network can be considered a good solution. There are also some papers that depend on feature extraction like this paper [8]. The researchers describe three stages of systems to detect RTCs, starting with the first stage which detects the car then the damaged texture detector with SVM which recognizes damaged parts, and finally, the car parts detector which detects the car's parts, the accuracy is about 81.83%. Also, in [9], the authors discussed single-vehicle traffic accident detection which consists of an automated traffic region detection method, a traffic direction estimation method, and a first-order logic traffic accident detection method, their proposed methods achieve good performance in real-time RTCs. Also, in [10], the authors proposed a three stages framework to detect RTCs. They start with a car detection method and then a tracker to focus on each car then in the final stage, they used the ViF descriptor that was introduced in this paper [11]. They got 89% accuracy. Their approach succeeded to be a general solution as it achieved high true alarms. But unfortunately, they got a lot of false alarms which made the system unreliable to detect the accidents. The bad accuracy was because of a lack of dataset of accidents and the accidents are stochastic.

The last approach was the most promising one, and we continued on their work with a method to find a way to maintain their high true alarms while achieving very low false alarms, and that is why we introduced our system for detecting accidents based on collision estimation.

After introducing the previous approaches to solving the problem, we found that they depend more on the clarity of the CCTV cameras and assuming that all CCTV cameras have exact quality resolution and are installed at a specific height and capture the vehicles at the same scale, which is not in real life.

Our approach seeks to work well on most CCTV cameras on roads. Moreover, they depend on the availability of a dataset of road accidents in the future which will increase the accuracy of

their approaches, but what we propose here does not depend on the dataset as we solved the problem as if it was a 2D car game in a 3D world and dealing with the problems of different resolution and different scales as we will explain later.

3. PROPOSAL

The framework consists of 5 modules; it starts with the vehicle detection module using YOLO neural network. In the second module, we track each vehicle using the MOSSES tracker. A third module is a new approach to detecting accidents based on collision estimation. Then the fourth module for each vehicle, we detect if there is an accident or not by using ViF descriptor with an SVM classifier to detect vehicle crashes. Finally, if there is an accident, it will report the emergency using GPS and GSM modules. Figure 2 illustrates the flow of the system.

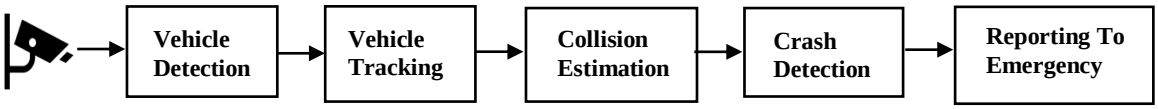


Figure 2. Proposed framework for detecting accidents

3.1. Vehicle Detection

For vehicle detection, we have used You Only Look Once (YOLO) network proposed in [12] because of its high accuracy and very low time processing compared to different networks as shown in Figure 3. In this paper, we have used YOLOv3 [13]. Yolo is considered faster than other convolutional networks because it looks at the image once and derives the bounding box and class probability from each object. A grid of SxS boxes has been used to divide the image. Then compute the confidence score for each box to see if the bounding box contains an object or not. So, the higher the confidence score gets the higher probability that the bounding box contains an object. When there is a single object surrounded by multiple bounding boxes. non-maximum suppression is applied to keep the most robust detection around a single object.

So, the results of this stage are the bounding box and labels that passed to the tracking module where the tracking module uses these bounding boxes as a starting point to track the vehicles.

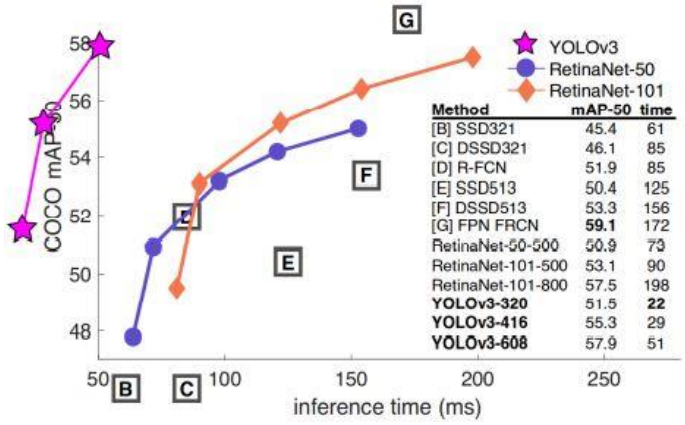


Figure 3. Comparison between the performance of YOLOv3 and other networks [13].

3.2. Vehicle Tracking

There are many tracking algorithms. But we need the type of tracking algorithm that fits our problem, that acts well on rotations, occlusion, and other distractions, that depends on correlation filters, so we used an algorithm called Minimum Output Sum of Squared (MOSSE) [14]. MOSSE depends on the frequency domain. It is capable of quickly obtaining the frequency transformation of the vehicle image using fast Fourier transform. With each new frame, this correlation filter is updated online. MOSSE filter is a correlation filter and that's why it can track complex objects. In addition, it produces a stable and reliable correlation filter when initialized using a single frame. It is also strong to variations in illumination, scale, position, and non-rigid deformations as well.

We are not only focusing on increasing the accuracy but also focus on performance. Vehicles are tracked every frame for 30 frames. And due to the congested road or traffic lights, some vehicles may not move or move at a slow speed. This is causing overprocessing on the tracking module to track the vehicle in every frame. So, we propose track compensated frame interpolation (TCFI) by estimating the vehicle's speed. If the vehicle's speed in the last three frames was less than the minimum speed, rather than tracking it in every frame, we track it in one frame and estimate where it will be in the following frame without using the tracking technique. So, the TCFI algorithm should be reevaluated in every frame since the vehicle may move higher than the minimum speed in some frames. Hence, in this situation, we will move it to the tracking algorithm then.

Applying TCFI improves the framework performance a lot, especially on congested roads, as the tracking module will use approximately half of the performance and will have the same accuracy as before.

3.3. Collision Estimation

In the beginning, we have an initial set of detected vehicles, then we create a unique id for each detected vehicle, and then we track each of these vehicles and maintain their id as they move through frames in a video. Now we have an object for each vehicle. The vehicle's position is saved for 30 frames in that object. The idea of the new approach is to limit the trackers entering the ViF descriptor using various features.

The collision estimation module might be used as a classifier to identify whether a crash happened or not, using the following algorithms.

First, we need to estimate the vehicle's speed in the video. Having a tracker on every vehicle, we can easily estimate the average speed of the vehicle by pixel unit, However, the camera angle differs from one CCTV camera to another as well as the camera position, height, and resolution. So, we have to keep in mind that we are dealing with a variety of CCTV cameras on the road. Thus, another unit instead of a pixel unit is needed to estimate the average speed of a tracker. To solve the camera resolution problem, every input feed must be resized to fixed width and height (480,360). However, if the majority of the CCTV cameras have a resolution of (1920, 1080), it is better to resize the input feed to that resolution. Another problem arises when CCTV cameras are hung at various heights or the feed is captured on various scales. As the area of the vehicle varies in the video, the vehicle will appear small if the camera is placed at a distant height. However, if the camera is set at a low height, it will appear that the vehicle has a large area; and because we estimate speed using pixel units, the small area will appear that moves slower than the larger area. So, the height or scale problem can be solved by multiplying the average speed of a vehicle

by a speed coefficient parameter stated in Equation 2. This parameter has an inverse relation to the area; a larger vehicle has a lower speed. But we also take into consideration the different types of vehicles as the motorcycle area will appear less than the truck area. This is solved by α parameter that is set according to each type of vehicle. For instance, a car has (α = 4). We calculate the coefficient as shown in Equation 1. sum_dx in Equation 2 represents the vehicle's horizontal movement during the last 10 frames.

$$\text{Coefficient} = \frac{Area\ of\ video\ frame}{\alpha \times \text{Area}\ of\ vehicle} \tag{1}$$

$$Speed = \text{Coefficient} * \frac{\sqrt{sum\_dx^2 + sum\_dy^2}}{10} \tag{2}$$

The angle of the camera plays an important role in the collision estimation process. The best camera angle is the one with the top view since the focus is on the x and y plane where the vehicles move. In the real world, it is impossible to capture such a top view image unless you use a drone as a CCTV camera. this makes the perfect possible perfect CCTV angle is the one in Figure 4 instead of Figure 5. If you can't modify the camera angle, the vehicle's average speed estimation and future position will suffer, reducing the system's accuracy; however, this may be fixed later in the next module.

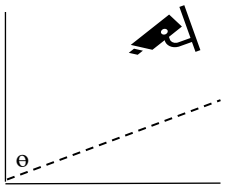


Figure 4: CCTV camera capture the view with an angle 30° to 90°

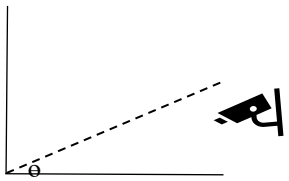


Figure 5: CCTV camera capture the view with an angle 0° to 30°

Second, compare every two vehicles with each other in different frames. If the average speed of the two vehicles is less than the speed limit, which is a hyper-parameter, then the two vehicles will not collide. As they are both moving slowly, So, we can reduce the possibility of their colliding at this stage. However, if one or both of them are moving faster than the speed limit, there is a risk of a collision, and the next step should be taken. This process of discarding some results is considered an excellent optimization of the computational power, especially when dealing with congested roads. So, only the high-speed cases will be investigated.

Third, measure the estimated centers in future frames for both vehicles after 10 frames of their current frame by calculating the average speed of the last 10 frames, as well as their angle, and predict the vehicle's position after ten frames from the current frame.

Fourth, measure the distance between the two estimated centers vehicles for both vehicles. We do that as we need to limit the cases where the vehicles will not meet close to each other in the future. So, if the distance between the two estimated centers is greater than half of the distance between the vehicle's center and its corner, plus the distance between the other vehicle's center and its corner, then they will not collide because they are apart from each other. But if the

distance is below, then the next step should be taken. So, only high speed and close to each other cases will be investigated.

At this stage, high-speed vehicles and close range to each other will report an accident but if the camera angle is not perpendicular to the road as shown in Figure 5 which is what most CCTV cameras are. Then an occlusion will occur and it will show as if they crash into each other. To solve this, we will calculate the difference between the estimated future and actual position for every two vehicles and get the maximum distance. If the maximum distance between two vehicles' actual and estimated centers is greater than half the distance between their two estimated centers in the future, it may be a crash, but if it is less, it is not a crash and they are limited out.

The results of detecting RTCs using collision estimation only without the next module crash detection were excellent. However, we also tried to add the next module crash detection to the pipeline to evaluate the performance of each module and determine the best technique to detect RTCs.

3.4. Crash Detection

We added another module to increase accuracy by tackling the camera angle problem mentioned in the collision estimating module after filtering out bad candidates and categorizing accidents. As a result, a feature vector is obtained, which is then used as input to a support vector machine (SVM) model that identifies whether there has been an accident or not. The feature vector for a single vehicle's sequence of frames is obtained using the violent flow (ViF) descriptor [11]. As an optical flow algorithm, we used ViF descriptor and Horn-Schunck. We used it for its good accuracy and minimal computing cost, as explained in this [10]. ViF descriptor is based on statistics of change in the magnitude of the optical flow vectors.
Finally, an SVM model is trained on the obtained feature vector from the ViF descriptor and classifies if there is an accident or not.

3.5. Reporting to the Emergency

In the last module in our pipeline, if the system detected an accident, it will not only send an SMS to the emergency services with the help of the microcontroller, GPS, and GSM modules but it will also send a notification in nearly real-time and allow them to monitor accidents using client-server architecture and an interactive GUI. In addition, the system saves accidents in the database indicating the location, date, time, and video of the accident where they can inspect it at any time as shown in Figure 6.

In Figure 7, we show the architecture of our system where it begins with the backend which collects the frames from the CCTV camera and sends them to the model passing through the four modules to determine whether there has been an accident or not. And once an accident happens, it will send an SMS to the emergency services and a notification to display the accident to check the criticality of the accident and take an urgent action. This helps to save many lives by reporting them in time.

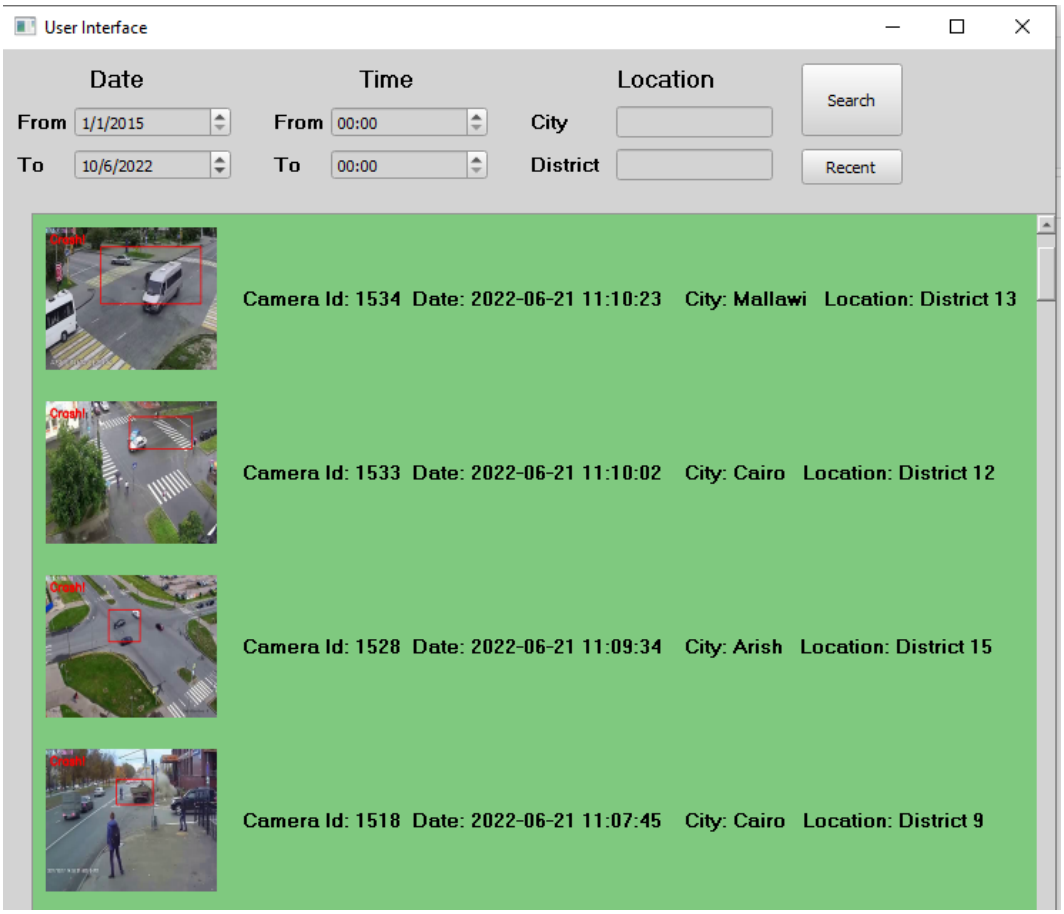


Figure 6. The system saves the accidents indicating date, city and location.

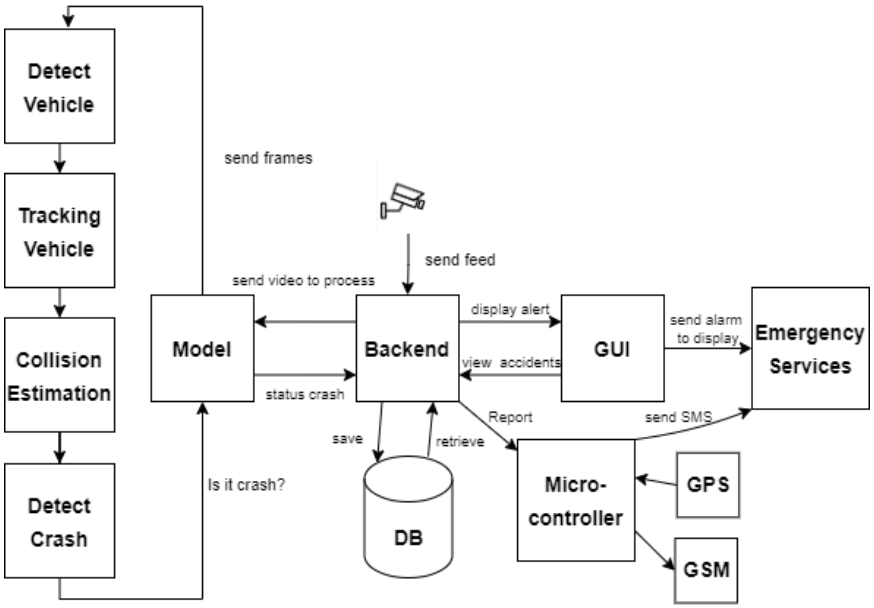


Figure 7. Architecture of the system

4. RESULTS

The experiments were done on a computer running Windows 10, Intel Core i7-10510U CPU @ 2.30 GHz with 16 GB of RAM and NVIDIA GeForce 920MX, and the system runs on python 3.6.

4.1. Datasets

We had to collect two types of the dataset; the first for training the ViF descriptor, which will be as described in 30 frames per second for the crashed vehicle only, and the second dataset for testing the whole system and the length of its video are about 20 seconds to 120 seconds.

First, We collected the training dataset for the ViF descriptor in three steps, the first was searching for an available dataset till we found those two papers [10][15], filtering them to obtain a high-quality resolution for vehicle crashes and the last step was executing our system and getting the output, which we then fed the training dataset with it. So, in total after those steps, we got 200 videos.

Second, we gathered the test datasets by downloading them from different YouTube sources, cutting and editing them, and getting 75 videos to test the system.

4.2. Results

We proposed two approaches, one depends only on collision estimation, and the other depends on both collision estimation and ViF descriptor. The system which depends only on collision estimation gives higher recall but lower accuracy. While the system which depends on both collision estimation and ViF descriptor have a lower recall but better accuracy as shown in Table 1.

In Figure 8; shows a comparison in processing performance. Our system is better in performance, especially on congested roads.

In Figure 9, we show our results in detection crashes using collision estimation. we mark the exact position of car crashes with red boxes.

Also, we measure the processing time of our system from the vehicle detection stage till emergency notification stage to get 3.04 seconds only for processing a video of 5 seconds as shown in Table 2.

Table 1. Accuracy and recall of our system compared to others

Method	**Accuracy**	**Recall**
Deep Spatio-Temporal Model [7]	79%	77%
ViF Descriptor [10]	75%	80%
Collision Estimation	91%	**94%**
Collision Estimation + ViF Descriptor	**93%**	78%

Using two techniques, our system performs better, especially on congested roadways. First, by applying TCFI to our tracking algorithm where the tracking algorithm applies only to the vehicles that move fast and apply TCFI to vehicles that move slowly or stopped. Second, we managed to limit the number of trackers entering the ViF descriptor using collision estimation which results

in very fast and high performance compared to using only ViF.

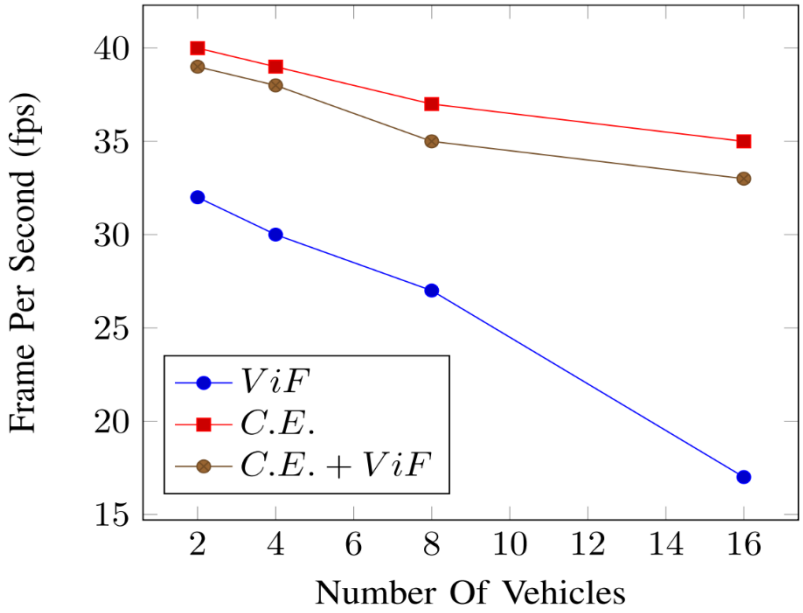


Figure 8. Comparison of our system and others in performance of frames per second with various numbers of vehicles

Figure 9. Car crash detection using collision estimation

Table 2. Processing Time of our system using collision estimation

	Video duration in sec.	Time processing in sec.
Collision Estimation	5	3.04

5. CONCLUSION

We presented a system that detects RTCs using installed CCTV cameras in real-time based on collision estimation. We also proposed a new technique track-compensated frame interpolation (TCFI) to track vehicles in a more efficient manner, especially on congested roads. The system achieves excellent results with low processing and performed better than other systems, with a 94% recall rate and 93% accuracy rate.

REFERENCES

[1] "The top 10 causes of death", WHO, Dec 2020, [online] Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[2] "Road traffic injuries", WHO, June 2021, [Online] Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[3] "Egypt Road Safety", WHO, 2012, [online] Available: http://www.emro.who.int/images/stories/cah/fact_sheet/road_safety.pdf.

[4] Vaishnavi Ravindran, Lavanya Viswanathan, and Shanta Rangaswamy, "A Novel Approach to Automatic Road Accident Detection using Machine Vision Techniques", International Journal of Advanced Computer Science and Applications, vol. 7, Nov 2016.

[5] Coco Feng. "China the most surveilled nation? The US has the largest number of CCTV cameras per capital". In: (Dec. 2019). URL: https : / / www . scmp . com / tech / gear / article / 3040974 / china - most - surveilled - nation - us - has - largest - number-cctv-cameras-capita.

[6] Jonathan Ratcliffe. How Many CCTV Cameras in London? URL: https://www.cctv.co.uk/how-many-cctv-cameras-are-there-in-london/

[7] Singh and C. K. Mohan. "Deep Spatio-Temporal Representation for Detection of Road Accidents Using Stacked Autoencoder". In: IEEE Transactions on Intelligent Transportation Systems 20.3 (2019), pp. 879–887. DOI: 10.1109/TITS.2018. 2835308.

[8] Ravindran, L. Viswanathan, and S. Rangaswamy, "A novel approach to automatic road-accident detection using machine vision techniques," INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, vol. 7, no. 11, pp. 235–242, 2016.

[9] Ghahremannezhad, Hadi & Shi, Hang & Liu, Chengjun, "A Real-Time Accident Detection Framework for Traffic Video Analysis". In 16th International Conference on Machine Learning and Data Mining(2020).

[10] V. Machaca Arceda and E. Laura Riveros. "Fast car Crash Detection in Video". In: 2018 XLIV Latin American Computer Conference (CLEI). 2018, pp. 632–637. DOI: 10.1109/CLEI. 2018.00081.

[11] Vicente Machaca, J.C. errez, and K. n. "Real Time Violence Detection in Video". In: Jan. 2016, 6 (7.) –6 (7.) DOI: 10. 1049/ic.2016.0030.

[12] J. Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[13] Joseph Redmon and Ali Farhadi. "YOLOv3: An Incremental Improvement". In: CoRR abs/1804.02767 (2018). arXiv: 1804. 02767. URL: http://arxiv.org/abs/1804.02767.

[14] David S. Bolme et al. "Visual object tracking using adaptive correlation filters". In: June 2010, pp. 2544–2550. DOI: 10. 1109/CVPR.2010.5539960.

[15] A. P. Shah et al. "CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis". In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2018, pp. 1–9. DOI: 10.1109/AVSS.2018. 863

AUTHORS

Mohamed Essam Ahmed
Teaching Assistant of computer science, Alexandria University,
Faculty of Engineering.

Nagia M. Ghanem
Associate professor of Computer Science, Alexandria University,
Faculty of Engineering.

Mohamed A. Ismail
Professor of Computer Science, Alexandria University,
Faculty of Engineering.

MULTI-VIEW HUMAN TRACKING AND 3D LOCALIZATION IN RETAIL

Akash Jadhav

Noque.store, India

ABSTRACT

In recent years, retail stores have seen traction in bringing online shopping experience to offline stores via autonomous checkouts. Autonomous checkouts is a computer vision-based technology that needs to understand three human elements within the store: who, where, and doing what. This paper addresses two of the three elements: who and where. It presents an approach to track and localize humans in a multi-view camera system. Traditional methods have limitations as they: (1) fail to overcome substantial occlusion of humans; (2) suffer a lengthy processing time; (3) require a planar homography constraint between camera frames; (4) suffer swapping of labels assigned to a human. The proposed method in this paper handles all the aforementioned limitations. The key idea is to use a hierarchical association model for tracking, which uses each human's clothing features, human pose orientation, and relative depth of joints, and runs at over 23fps.

KEYWORDS

Multi-view, Data Association, Tracking, Localization.

1. INTRODUCTION

There has been significant research work done in estimating and tracking the pose of a human from a single view [4, 2, 16] and multi-view images [5, 12, 17, 18], yet few existing methods have been crafted to tackle the problem of tracking and localizing humans in a retail store. Problems faced in tracking and localizing humans in the retail stores are especially challenging for computer vision algorithms due to: significant heavy occlusion of humans, similarly dressed humans, swapping of labels assigned to humans, and having an extensive baseline between cameras. However, the potential application of estimating and tracking the 3D pose of humans is in retail stores. It provides an autonomous checkout experience to the customer and gathers a lot of analytics, which helps increase the sales and helps understand customer behavior, which brings the potential of online retail to offline retail.

Estimating 2D poses of humans in the images is a well-researched problem in deep learning [1, 9]. With further optimization [26] of these deep learning models, they can run in real-time [29]. The task of tracking humans in a single view and multi-view camera system [8, 11, 5], further combining the 2D poses from multiple perspectives to generate 3D skeletons, has also been explored [12]. However, none of these methods work robustly for the application of tracking and localizing humans in retail stores due to: (1) substantial occlusion of humans; (2) lengthy processing time; (3) requirement of a planar homography constraint between camera frames; (4) swapping of label assigned to a human. This paper proposes a hierarchical association model to find correspondences between 2D poses in multiple views, employing them to generate and track 3D skeletons. By maintaining the exact label of a human throughout the journey in the 3D space,

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 181-192, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121214

the resulting method provides a significant improvement over previous methods and corrects the errors associated with multi-view human tracking. Example results can be seen in Figure 1.

(a) (b)

Figure 1. Results of the proposed method. The two images on the left of (a) and (b) are from the dual-view camera system, which shows humans detected with a particular color bubble (label).

The right of (a) and (b) shows the localized 3D skeletons in the 3D space with the same color. (a) Results show tracking and localization of humans in an occluded scene, where the blue bubble is the predicted state of the occuluded human in the upper view; (b) Results show tracking and localization of humans in case of missed detection, where the blue 3D skeleton is the predicted state of the occuluded human with missed detection in the upper view.

This association model helps in estimating the 3D skeleton in 3D space for a prolonged time under substantial occlusions and missed detections of humans in a multi-view system, which enables to understand some complex actions like picking an object, placing the object back on the rack, dropping the object in the bag or cart, passing or throwing an object to another human, and catching an object by another human.

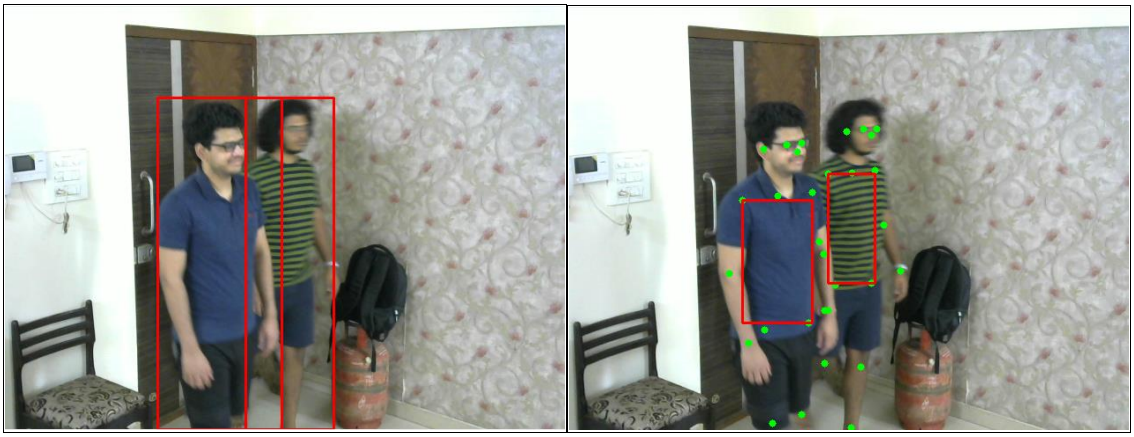
The proposed algorithm is experimented on the generated dataset [31], which is generated using two Logitech C270 cameras with two-to-three people maneuvering in the surrounding. This dataset consists of a readme.pdf file that defines the content in the dataset. Unfortunately, there is no similar dataset available that includes rigid camera poses in 3D space with a fixed number of humans maneuvering in the same 3D space. The following section provides litrature for some previous approaches and comapers them to solving this problem to the proposed algorithm.

2. LITERATURE REVIEW

Estimation of 2D human poses from a monocular image can be categorized into two categories: (1) single-person 2D pose detection [20, 22, 23]; (2) multi-person 2D pose detection [1,9, 21]. Toshev and Szegedy [22] provide a regressor to directly estimate the 2D joint coordinates in an image. An end-to-end deep learning model that learns spatial models for 2D pose estimation was presented in [23]. Deep convolutional network-based pose estimators result in a significant increase in accuracy and provide a basis for more difficult pose estimation tasks such as multi-person 2D pose estimation [1, 9]. Cao et al. [1] presents a fusion of joint confidence map and a learned vector that defines the relationship between the joints, and estimates the 2D poses.

Some researchers have proposed a single-view and multi-view human detection and tracking algorithm to overcome the limitations of humans being occluded in the camera scene. Cai and Aggarwal [10] extend a single-camera tracking system by switching to other views when the system predicts that the current camera will no longer have a clear view of the object. Krumm et al. [13] combine information from multiple stereo cameras in the 3D space. They perform background subtraction and then detect human-shaped blobs in 3D space. For detecting and tracking the same human in numerous images, each person is assigned a color histogram. Back-projection in 3D space estimates the 3D points guaranteed to lie inside the detected objects. Wojke et al. [8] tries to detect and track objects in a single view by extracting a feature vector from the bounding box assigned to the detected objects and compares this feature vector with other feature vectors. Even though this method attempts to resolve occlusions, the underlying problem of using such features is that the overlapping bounding boxes might get corrupted, as shown in Figure 2 (a). In the approach proposed in this paper, a bounding box is computed using the 2D joint coordinates to extract the clothing features, as shown in Figure 2 (b).

Khan et al. [5] requires a planar homography constraint between multiple cameras to track multiple humans. This constraint creates a dependency that the cameras need to see the floor with the human feet pixels always visible to project them in all other frames. Then a clustering algorithm associates these projected pixels to a particular human. The floor and human feet would not always be visible from the camera feed in a retail store as shown in Figure 2. Bridgeman et al. [3] provides a method which only depends on 2D joint association without any appearance features to track and generate a 3D pose. This approach results in the swapping of labels when two humans pass close by with the same orientation.



(a) (b)

Figure 2. In computer vision, estimatimating the six degree-of-freedom camera pose in the world frame from n 3D-to-2D point correspondences is a fundamental and well understood problem. This pose could be estimated with minimum 6 correspondences in 3D sapce and image pixels, using the well known Direct Linear Transform (DLT) algorithm. To improve the accuracy of the DLT, Perspective-n-Point with Ransac is used [27]. In the proposed framework, to estimate the orientation of human in the image, rough estimates of the 3D coordinates are taken for the 2D joint pixels. For rough estimates of the 3D coordinates of an object and its corresponding 2D joint pixels, a scaled translation of object origin is obtained, but the orientation remains within limits of 3 to 5 degrees across all axis.

Depth is an useful representation for actions in the physical environments. Monocular depth estimation remains a challenging problem that is heavily underconstrained. To solve it, one must exploit many visual cues, as well as long-range context and prior knowledge. This calls for

learning-based techniques [6, 7, 19]. Ranftl et al. [19] proposes a robust training objective that is invariant to changes in depth range and scale, uses principled multi-objective learning approach to combine data from different sources, and highlights the importance of pretraining encoders on auxiliary tasks. Ranftl et al. [19] provides a pretrained model [28] which estimates the relative depth information from a monocular image and runs in real time. In the proposed framework, relative depth information of neck joint is used for the betterment in data association when the clothing features and human orientations are approximately same.

The scene with a higher number of occluded objects would be challenging to resolve for any of the previous methods. Not only are there cases of near-total occlusion, but similarly dressed people would also be a challenge. Using just the color distributions, or full human bounding box features, or 2D joint coordinates for region matching across cameras would lead to the incorrect association and result in swapping of labels or assigning a new label to the human. A hierarchical association model is proposed in the approach in order to overcome these limitations, which uses clothing features, human orientation data which is computed using the 2D joint pixels [27], and relative depth estimate [19,28] of the neck joint.

With the advancement in deep convolutional neural networks, it is easy to extract many features from an object. How well these features are associated with objects across all the frames defines a valuable tracking system. Feature vectors are computed from the bounding boxes of detected objects using similar convolutional neural networks. Wojke et al. [8] employs a feature extraction model from bounding boxes as shown in Figure 2(a), trained on a reidentification dataset [24]. This dataset contains over 1,100,000 images of 1,261 pedestrians, making it well suited for deep metric learning in a people tracking context. The method proposed in this paper employs the bounding box computed from the 2D joints detected as shown in Figure 2 (b), hence a clothing feature extraction model is trained on the reidentification dataset [24].

Single view and multi-view tracking algorithms adopt a single conventional hypothesis tracking methodology with recursive Kalman filtering and frame-by-frame data association. In the approach proposed in this paper, Kalman filter-based tracking is employed in both image coordinates and 3D space coordinates. In image coordinates, the bounding boxs, as shown in Figure 2 (b), are tracked, and in 3D space coordinates, the 3D joints of the skeleton are tracked with Kalman Filter. Tracking in both image coordinates and 3D coordinates provides robust and continuous 3D pose estimation in the 3D space. This technique enables the possibility to understand some complex actions performed by the 3D skeleton of humans in the 3D space.

3. METHODOLOGY

The proposed framework takes as input images from multiple cameras, camera calibration parameters, and the rigid pose of cameras in the 3D space. The multi-view images are passed through a pose detector [29] and a monocular depth estimator [19, 28], providing 2D pose estimations and relative depth information in each frame. Two successive processes are applied to the 2D pose data: the first step computes the bounding box from the 2D poses, and the second step computes the orientation of humans from their 2D poses in the images. From the bounding boxes, clothing features are computed using a feature extracting neural network. The relative depth of neck joints is estimated in each image using the relative depth information and 2D joint pixels. Further, the relative depth of neck joints, human orientation data, and the human clothing features are fed as input to the hierarchical data association model, which assigns a label to every human, ensuring consistency between multiple views. This set of labels with humans is used to track bounding boxes in the images and 3D joints in the 3D space. 3D joints are estimated if the human 2D joints are seen in two or more views using robust triangulation. In each image,

bounding boxes are tracked instead of 2D joints to reduce the time complexity during runtime. Tracking of the bounding boxes ensures the tracking of the 2D joints, as mapping is stored between the 2D joints and the bounding box. If 2D pose measurements are available for a label assigned to a human, they are passed to the 3D joints estimation and tracking block. If the 2D pose measurements are not available, the predicted state of 2D joints from the tracked bounding boxes is passed to the 3D joints estimation and tracking block. A system overview is presented in Figure 3.

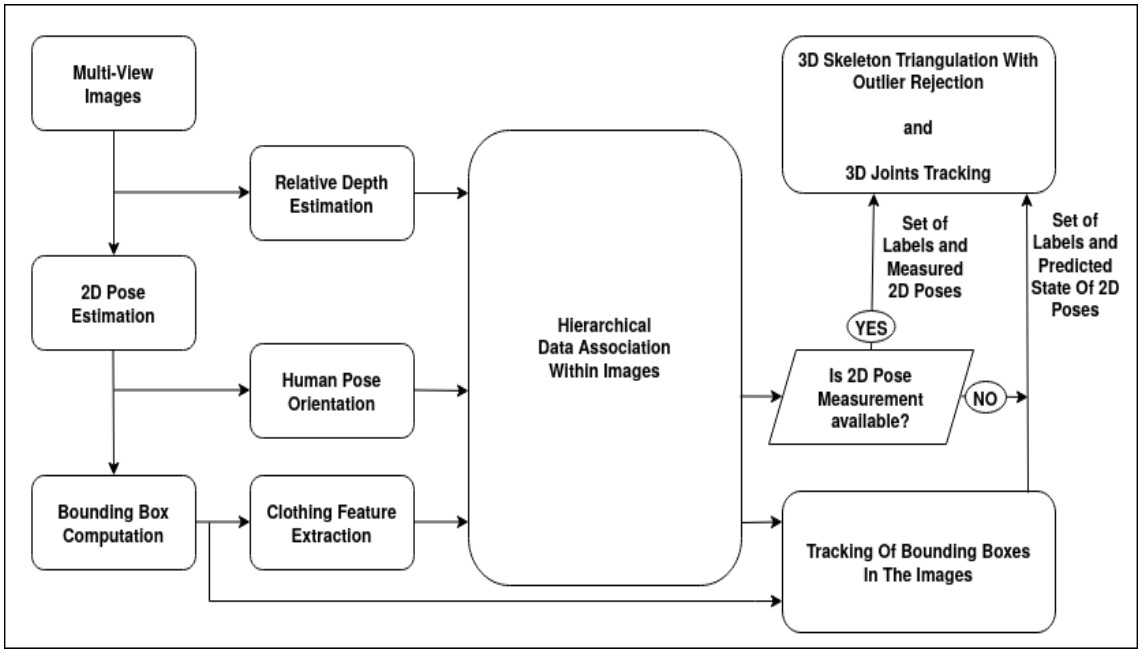


Figure 3. An overview of the pipeline.

The pose detector [29] estimates a total of 18 joints on a human in an image. To avoid the corruption in features due to overlapping bounding boxes, as shown in Figure 2 (a), a bounding box is computed using the left shoulder joint, right shoulder joint, right hip joint, and left hip joint as shown in figure 2 (b). This bounding box is passed to two functional blocks of the pipeline: (1) a feature extraction block to compute a feature vector from the clothing; (2) the bounding box tracking block.

Once the 2D pose of a human is detected, six joints are used to compute the orientation 27] of human in an image: neck joint, left shoulder joint, right shoulder joint, right hip joint, left hip joint, and the joint between the right hip joint and the left hip joint. These six joints always lie on a plane in the 3D space, the distances between these six joints always remain the same, and a reference frame could be attached to one of these joints. The 3D coordinates used for the neck joint, left shoulder joint, right should joint, right hip joint, and left hip joint are (0.0, 0.0, 0.0), (-250.0, 0.0, 0.0), (250.0, 0.0, 0.0), (250.0, -500.0, 0.0), (-250.0, -500.0, 0.0), and(0.0, -500.0, 0.0), where the readings are in milli-meters, and the neck joint is considered to be the origin. Relation between image pixels (u,v), their corresponding 3D points (X, Y, Z), intrinsic camera matrix (K), and the pose (rotation (R) and translation (T)) is given in Eq. 1, where K and RT are of size 3x3 and 3x4 respectively.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} r11 & r12 & r13 & t1 \\ r21 & r22 & r23 & t2 \\ r31 & r32 & r33 & t3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

To extract a feature vector from the bounding box a convolutional neural network has been trained on a re-identification dataset [24]. This dataset contains over 1,100,000 images of 1,261 pedestrians, making it well suited for clothing feature extraction in a people tracking context. The convolutional neural network architecture of my network is shown in Table 1. Table 1 (a) displays the architecture of the encoder network. It consists of 5 layers with 5x5 filters and a stride of 2 using ReLU as an activation function. The output size from each layer should be read as height x width x channels. Table 1 (b) displays how the last activation map from the encoder network is transformed into the latent vector z, where the latent dimension is set to 10. Table 1 (c) displays how the output vector of size 1x768 is transformed into a tensor of shape 3x2x128. Note that the network is symmetrical with the equal number of layers, filter size, stride length, and activation as the encoder network. The training process of this network architecture is out of scope for this paper. One forward pass of 32 bounding boxes takes approximately 30ms on an Nvidia GeForce GTX 1070 GPU. Thus, this network is well suited for online tracking.

Table 1. The network layers and their respective output sizes are shown. The terms "C", "S" and "F" stand for channels, stride and filter size respectively. (a) Encoder Network architecture displaying activation function, number of channels, stride and filter for each layer in the encoder network. (b) Latent space architecture displaying the flow of data received from the encoder. Note that z-mean and z-std are separate fully connected layers. They are placed in the same row since they are computed in parallel. (c) Decoder Network architecture displaying activation function, number of channels, stride and filter for each layer in the decoder network.

Encoder Layers	**Output Size**
Input image	96x64x3
Conv1 - ReLU, C: 8, S: 2x2, F: 5x5x3	48x32x8
Conv2 - ReLU, C: 16, S: 2x2, F: 5x5x8	24x16x16
Conv3 - ReLU, C: 32, S: 2x2, F: 5x5x16	12x8x32
Conv4 - ReLU, C: 64, S: 2x2, F: 5x5x32	6x4x64
Conv5 - ReLU, C: 128, S: 2x2, F: 5x5x64	3x2x128
Reshape - Flatten	1x768

(a)

Latent Space Layers	**Output Size**
Flattened vector	1x768
Fully connected z-mean and z-std	1x10 + 1x10
Generated latent vector	1x10
Fully connected - ReLU	1x768

(b)

Decoder Layers	**Output Size**
Activation:ReLU, Fully-connected	1x768
Reshape - Tensor form	3x2x128
Deconv1 - ReLU, C: 64, S: 2x2, F: 5x5x128	6x4x64
Deconv2 - ReLU, C: 32, S: 2x2, F: 5x5x64	12x8x32
Deconv3 - ReLU, C: 16, S: 2x2, F: 5x5x32	24x16x16
Deconv4 - ReLU, C: 8, S: 2x2, F: 5x5x16	48x32x8
Deconv5 - ReLU, C: 3, S: 2x2, F: 5x5x8	96x64x3

(c)

From the relative depth information obtained from the mono-depth estimation model [19, 28] and 2D poses obtained from the pose detection model [29], humans are ranked according to their relative depth of the neck joint in the image. Later, for these relative depth points, a scaled distance is assigned as per their rank to create a scaled 3D coordinate. Finally, to associate humans to a particular label, these 3D coordinates are transformed onto other camera frames and checked with the neck's rank of relative depth point in the other camera frames. Refer to Figure 4 for reference; in view 1, the rank of cylindrical objects with respect to their relative depth points are [P1, P2]. A scaled distance 1D1 and 1D2 (1D1<1D2) is assigned for P1 and P2, in view 1. When these depth points are transformed to view 2, 2D1 and 2D2 are obtained (2D1<2D2). In view 2, the rank of cylindrical objects with respect to their relative depth points are [P1, P2], so 2D1 is associated with P1, and 2D2 is associated with P2.

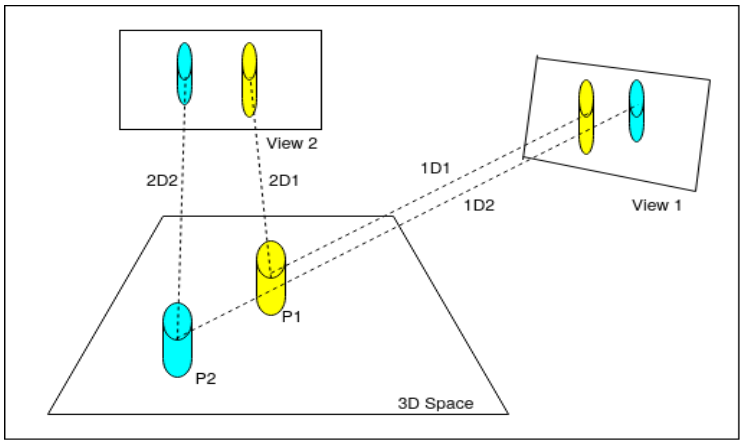


Figure 4. The figure shows two cylinderical objects standing in the 2D space, and the scene being viewed by two cameras.

The Hierarchical data association model takes as input a human vector which consists of: (1) the extracted clothing features; (2) human pose orientation; (3) relative depth of the neck joint. The hierarchical data association model computes an association mapping of labels with 2D poses and association mapping of labels with bounding boxes. As the N number of people inside the store are known, first the human data vectors are clustered into N labels using the clothing feature vector and human orientation data. This clustering is done by comapring the clothing features and human orientations data in the same image and later with the clothing features and human orientations data in every other image. The clothing appearance features are associated using a metric which computes a scalar value by comparing two feature vectors of size 1x768. This

metric measures the cosine distance d between two normalized clothing feature vectors v1 and v2 as shown in Eq. 2. Again, a binary variable is introduced as shown in Eq. 3, to indicate if an association is admissible according to the above metric, where b is the binary value and t is a suitable threshold computed on the dataset [24].

$$d(v1, v2) = (1 - v2.T * v1) \tag{2}$$

$$b(v1, v2) = 1[d(v1, v2) \leq t] \tag{3}$$

If N different clusters are not formed by associating just the clothing features, the human orientations which are obtained with respect to the camera coordinates are transformed to the 3D space coordinates and compared with the other available human orientations. A binary variable, similar in Eq. 3 is used to compare if the difference between two human orientations is within a threshold. If still N different clusters are not formed by associating the clothing features and human orientations, humans are orientated using the relative depth information of neck joints. Just the relative depth information alone is not used to associate humans to labels, as this information is usable only upto a particular distance. Once N different clusters are formed, the multi-view tracking system is initialized and creates a set of labels and human vectors. Later, this set of labels and human vectors is used to associate human vectors from the multi-view system to a particular label. If a human assigned to a particular label is seen in two or more views, a 3D skeleton is estimated using robust triangulation. The 3D location of each joint s_{Ij} in a skeleton with label I is optimised using Eq. 4.

$$\underset{s_{ij}}{argmin} \sum_{c} \sum_{I} alpha_{ij}{}^{c} \| P_c(s_{Ij} - p_{ij}{}^{c}) \|, \{p_i \in I, c \in C\} \tag{4}$$

where, $P_c(s_{Ij})$ is the projection of s_{Ij} in a camera c. This results in a set of 3D skeletons per frame, and RANSAC is used to eliminate the outlier pose detections. The bone lengths of the resultant skeleton are thresholded to remove any remaining outlier 3D joints.

To improve the localization of joints in the 3D space, the proposed framework implements Kalman filter-based tracking of bounding boxes in each image and 3D joints. The 2D joints are mapped with the bounding boxes. As the bounding boxes are tracked, the 2D joints get tracked by default. Tracking of bounding boxes is done instead of 2D joints to reduce the time complexity of the pipeline during runtime. If the 2D pose measurements are available, they are used to compute and track the 3D joints. If the 2D pose measurements are not available, the 2D joints obtained from the predicted state of bounding boxes compute and track the 3D joints. Kalman filter is applied in tracking to smoothen the final results, to compensate for the missed detections, and substantial occlusions of humans for a few frames.

4. EVALUATION

The method proposed in this paper is evaluated on the generated dataset [31], which consists of two cameras with two-to-three people maneuvering in the surrounding. Limitations of some previous methods [8, 5] on this dataset are described in this section. The specification of the dataset used for evaluation is shown in Table 2.

Table 2. Properties of the datasets used for qualitative evaluation: number of cameras (C); number of people (P); camera resolution (R);and number of frames (F).

Dataset [31]	C	P	R	F
2person.zip	2	2	480p	130
3person.zip	2	3	480p	330

The proposed method's feature extraction process is compared with the state of the art tracking method's [8] feature extraction process. The feature extraction model of the proposed method and the state-of-the-art method [8] is tested on Figure 2 (b) and Figure 2 (a) respectively. The comparison is made by calculating the cosine distance(CD) from Equation 2 between two human's bounding box feature vectors in each image, where, Lower cosine distance indicates more similarity between the two feature vectors.. In Figure 2 (a), the bounding boxes are generated by detecting humans using the yolov4 deep learning model [25], and in Figure 2 (b), the bounding boxes are generated by the 2D joints detected [29]. The cosine distances computed between two humans in Figure 2 (a) and Figure 2 (b) are shown in Table 3. The cosine distance between two humans in Figure 2 (a) is less than that in Figure 2 (b) due to the overlapping of bounding boxes which leads to corruption in the feature vectors. When the number of humans maneuvering in the scene increases, there would be more overlapping in bounding boxes, adding errors in the human-label association process.

Table 3. Metric values between two feature vectors.

Figure	CD between two clothing feature vectors
Figure 2 (a)	0.3489
Figure 2 (b)	0.9427

To assess the quality of the proposed system, the number of label switches are computed, a metric commonly used in multi-object tracking [15] that counts the number of times a tracked object is assigned a new identity. The results are shown in Table 4. Previous method [8, 30] which tracks humans in a single view using just the bounding box features results assigning new labels to humans under substantial occlusion, as shown in Figure 5 (a). Wojke et al. [8] fails to maintain the label assigned initially to a human throughout the journey in the 3D space. It fails to track a human in a single view and cannot be scaled to a multi-view tracking system. The method [5] does not work on the provided dataset [31] as there is no planar homography constraint between the camera frames, and in a retail store, it is tough to make every camera view the floor.

Table 4. The number of frames (F), tracked people (TP), and Label switches (LS) for out and previous methods [8, 5] in each dataset.

Dataset [31]	F	TP	LS - Ours	LS - [8]	LS - [5]
2person.zip	130	2	0	8	N.A.
3person.zip	330	3	0	21	N.A.

In the 2person.zip and 3person.zip dataset, all humans maintained their label for the duration of the sequence, even during close contact, substantial occlusion, and missed detections, which are observable in Figure 5 (b). The proposed method in this paper succeeds in maintaining the label assigned initially to a human throughout the journey in the 3D space. It also succeeds in tracking a human in a single view and a multi-view system.

The method is tested on a system with an Intel i7 2.2GHz processor. 16GB of RAM, and 12GB Nvidia GeForce GTX 1070 GPU. The deep neural networks run over GPU and tracking

algorithms run over CPU. The parallelized implementation runs at over 23fps on the dataset [31]. The 2Dpose association stage is the most computationally expensive, and the time taken to track the 2D bounding boxes and 3D skeletons adds a little latency. The methods which use pictorial structure models [12, 14] for association, run at approximately 1fps and 10fps respectively.

5. CONCLUSIONS

This paper presents a new method for computing and tracking 3D skeleton of humans in a multi-view camera system. The proposed hierarchical model for associating humans to labels compensates for errors in overcoming substantial occlusions of humans, does not have any constraints like consistency in planar homography between the cameras, and can identify correspondences between humans in different viewpoints. Moreover, the algorithm is capable of running at over 23fps, and tracks the 3D skeletons for a prolonged time. In future, the association algorithm could be made more robust by researching some deeplearning models which predict the human skeleton depth from monocular cameras instead of the relative depth which is currently used in the proposed approach. Tracking of 3D skeletons for a prolonged time enables to understand better some complex actions done by humans in the retail store.

(a)

(b)

Figure 5. Left images show results on frame before occlusion, center images show results on frame during occlusion, and right images show results on frame after occlusion. (a) Results on Wojke et al. [8, 30] method. (b) Results on the proposed method.

REFERENCES

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, (2017) "Realtime multi-person 2d pose estimation using part affinity fields", IEEE CVPR.

[2] M. Andriluka, S. Roth, and B. Schiele, (2010) "Monocular 3d poseestimation and tracking by detection", IEEE CVPR.

[3] Bridgeman, Lewis and Volino, Marco and Guillemaut, Jean-Yves and Hilton, Adrian, (2019) "Multi-Person 3D Pose Estimation and Tracking in Sports", IEEE CVPR (Workshops).

[4] Rochette, Guillaume and Russell, Chris and Bowden, Richard, (2019) "Supervised 3D PoseEstimation from a Single Image using Multi-View Consistency", BMVC.

[5] Khan S.M., Shah M., (2006) "A Multiview Approach to Tracking People in Crowded Scenes Using a Planar Homography Constraint", ECCV.

[6] D. Hoiem, A. A. Efros, and M. Hebert, (2005) "Automatic photo pop-up", ACM Transactions on Graphics.

[7] A. Saxena, M. Sun, and A. Y. Ng., (2009) "Make3D: Learning 3D scene structure from a single still image", IEEE PAMI.

[8] Wojke, Nicolai and Bewley, Alex and Paulus, Dietrich, (2017) "Simple Online and Realtime Tracking with a Deep Association Metric", IEEE International Conference on Image Processing.

[9] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, andB. Schiele, (2016) "Deepercut: A deeper, stronger, and faster multi-person pose estimation model", ECCV.

[10] Cai, Q. and Aggarwal, (1998) "Automatic tracking of human motion in indoorscenes across multiple synchronized video streams", ICCV.

[11] Wojke, Nicolai and Bewley, Alex, (2018) "Deep Cosine Metric Learning for Person Re-identification", IEEE Winter Conference on Applications of Computer Vision (WACV).
[12] V. Belagiannis, S. Amin, M. Andriluka, B. Schiele, N. Navab, and S. Ilic. (2016) "3d pictorial structures revisited: Mul-tiple human pose estimation", IEEE PAMI.
[13] Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., and Shafer, S., (2000) "Multi-camera multi-person tracking for easy living", IEEE International Workshopon Visual Surveillance.
[14] J. Dong, W. Jiang, Q. Huang, H. Bao, and X. Zhou, (2019) "Fast androbust multi-person 3d pose estimation from multiple views", IEEE CVPR.
[15] A. Milan, L. Leal-Taix, I. D. Reid, S. Roth, and K. Schindler, (2016) "Mot16: A benchmark for multi-object tracking", CoRR.
[16] D. Tome, C. Russell, and L. Agapito, (2017) "Lifting from thedeep: Convolutional 3d pose estimation from a single image", IEEE CVPR.
[17] H. Joo, T. Simon, X. Li, H. Liu, L. Tan, L. Gui, S. Baner-jee, T. S. Godisart, B. Nabbe, I. Matthews, T. Kanade,S. Nobuhara, and Y. Sheikh, (2017) "Panoptic studio: A massively multiview system for social interaction capture", IEEE Trans-actions on Pattern Analysis and Machine Intelligence.
[18] L. Sigal, M. Isard, H. W. Haussecker, and M. J. Black, (2011) "Loose-limbed people: Estimating 3d human pose and mo-tion using non-parametric belief propagation", international Journal of Computer Vision.
[19] Ranftl and Katrin Lasinger and David Hafner and Konrad Schindler and Vladlen Koltun, (2020) "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer", IEEE PAMI.
[20] P. F. Felzenszwalb and D. P. Huttenlocher, (2004) "Pictorial structures for object recognition", International Journal of Computer Vision.
[21] M. Kocabas, S. Karagoz, and E. Akbas, (2018) "Multiposenet: Fastmulti-person pose estimation using pose residual network", ECCV.
[22] A. Toshev and C. Szegedy, (2014) "Deeppose: Human pose esti-mation via deep neural networks", IEEE CVPR.
[23] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, (2016) "Convolutional pose machines", IEEE CVPR.
[24] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, (2016) "MARS: A Video Benchmark for Large-Scale Person Re-Identification", ECCV.
[25] Bochkovskiy, Alexey and Wang, Chien-Yao and Liao, Hong-Yuan Mark, (2020) "YOLOv4: Optimal Speed and Accuracy of Object Detection", IEEE CVPR.
[26] Nvidia: Torch2TRT, https://nvidia-ai-iot.github.io/torch2trt/v0.2.0/index.html
[27] Perspective-n-Point, Wikipedia, https://en.wikipedia.org/wiki/Perspective-n-Point
[28] MIDAS v21 Small Model, https://github.com/AlexeyAB/MiDaS/releases/download/midas_dpt/midas_v21_small-70d6b9c8.pt/
[29] Nvidia: PoseTRT, https://github.com/NVIDIA-AI-IOT/trt_pose
[30] Wojke et al. [8] Implementation, https://github.com/nwojke/deep_sort
[31] Dataset, https://drive.google.com/file/d/11OqvWwXXqnR8KP_QmVk07Pgwg20rw2H3/view?usp=sharing

AUTHORS

Akash Jadhav
Founder, Noque.store
akash.jadhav@noque.store
https://www.linkedin.com/in/akash-jadhav-2201/

AN ONLINE GRAPHICAL USER INTERFACE APPLICATION TO REMOVE BARRIERS IN THE PROCESS OF LEARNING NEURAL NETWORKS AND DEEP LEARNING CONCEPTS USING TENSORFLOW

Justin Li[1] and Yu Sun[2]

[1]Troy High School, 2200 East Dorothy Ln, Fullerton, CA 92831
[2]California State Polytechnic University,
Pomona, CA, 91768, Irvine, CA 92620

ABSTRACT

Over the years, neural networks have become increasingly important and complex due to the rising popularity of artificial intelligence technologies. It allows for complex decision prediction making, and is an essential part in the modern AI industry. However, due to the complex nature of neural networks, a lot of complex math and logic has to be well understood along with a proficiency in programming in order for one to make anything practical with this technology. This is unfortunate, however, that many do not have the required high level math skill, or the proficiency in coding, blocking a lot of people from reaching and experimenting with this technology. My method attempts to eliminate the complexity that developing neural networks bring, and bring a clearer picture of what the user may be creating and working with. With the help of modern web technologies such as JavaScript and tensorflow.js, I was able to create a GUI program that can create, train, and test a neural network right on a browser, and without writing any code with a comparable result [13].

KEYWORDS

Neural network, deep learning, CNN.

1. INTRODUCTION

From self-driving vehicles to advances in healthcare applications, Deep Learning has been revolutionizing today's society [2]. For its ability to learn from large amounts of unstructured and unlabeled data, Deep Learning possesses the capability to perform complicated tasks such as driving, translating, and even performing image recognition [4]. And as the goal of Deep Learning is to simulate the process of a learning human brain, a multi-layered Neural Network is used at its core [5]. It functions as universal function approximators, which allows it to be trained for any circumstances given an appropriate set of input and output data-set. This is also the main reason for the popularity and potential of Deep Learning, as its flexibility allows for it to adapt to real world situations [8]. Such flexibility enables Deep Learning to automate jobs never thought was possible before, such as self-driving and image-colorization [3]. Fields such as Healthcare even started adopting this technology to create diagnosis for Breast Cancer based on related data [1].

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 193-200, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121215

However, all these benefits come with a catch, and that is that Deep Learning involves complex math knowledge such as Multivariate Calculus, Linear Algebra, and Statistics. Knowledge such as Algorithms and Programming are also required in order to implement Deep Learning [7]. However, as the data shows in an article, in 2019, there are only 23.9 million out of 7.71 billion who are programmers and software developers worldwide, which is less than 1% of the population. Because technologies such as weather predictions depend on this sort of technology, the result of the incompetence in this field will result in less innovation in this field, thus technologies such as smart stock trading will cease to improve [11].

Some preexisting softwares and applications have allowed users to create and view their neural networks visually <evidence needed>, but most have confusing layouts, complicated user interfaces, or lack the ability to create complex neural networks with custom data entries. Tools such as the Tensorflow Playground <source needed>, only allows the user to choose from a defined set of training data. This prevents the user from testing the architecture's efficiency on real world data that the architecture may be used on. These types of implementation pose a limit for the user on their data variability, neural network complexity, and usefulness in general. A second problem with these software is that they generally have very complex and hard to understand user interfaces, making it very hard for beginners to experiment with and use. Static user interfaces, such as the one from Tensorflow, limits the user to the type and complexity of neural networks they can create, and thus making it impossible for more advanced users to create more sophisticated architectures [14].

Taking the pros and cons of previous methods into consideration, our goal is to create a simple, easy to use, modular, and highly expandable software for creating neural networks visually. With that in mind, our implementation features a minimalism design, a highly expandable and easy to use layout, as well as the ability to run directly on a web browser. Compared to existing methods, our method is a lot cleaner and organized, while being highly functional and expandable.

In two application scenarios, we demonstrate how the above combination of features increases the experiences and speed for users to use this app. First we show the ease of use and functionality of our method through a comprehensive application test. Second, we compared the speed at which a user can create a functional and well performing neural network with our method, previous methods, and traditional method.

The rest of the paper is organized as follows: Section 2 provides the details on challenges that I encountered during design and development; Section 3 focuses on the details of my solution and well as the solution to the problems described in Section 2; Section 4 presents the relevant details about the experiments regarding the solution, comparing it to older methods and alternate methods; Section 5 gives more details on the alternate methods that was used to compare to my method. Finally, Section 6 gives the conclusion as well as planned future works on this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Parsing custom data

Letting the user choose their own data to train the neural network on is a very useful feature, as it allows for a wider and more expansive training environment [9]. However, it is not always clear if the user's data structure will match the neural network's input structure, or is even a valid

dataset. Therefore checks will have to be performed on both the network and the neural network to ensure no failure occurs during training.

2.2. Creating a usable, minimalistic user interface

User interface is a crucial part of our method, as it is aimed towards beginners. And while minimalistic user interfaces are great at being easy to use and simple to understand, they often lack functionality, or take up too much space. The way of expression is also important, as we need to represent an abstract idea of a neural network, a system of matrix dot products and vectors, into a graphical visual that is accurate, easy to understand and customizable. Other aspects such as tutorials and controls have to all be as intuitive as possible.

2.3. Making a custom architecture that is trainable

Training a neural network well can sometimes be the hardest part of AI development. Things such as training hyper-parameters and activation have to be chosen wisely for a good performing network, and those are dependent on the data. These hyper-parameters can be chosen through calculations and parameters, but are often left to the user to decide, which can be quite daunting to beginners. Other things like training optimizer and loss also influences the network's performance, and using the wrong ones can have devastating effects on the training's outcome.

3. SOLUTION

Through the interaction with the graphical neural network representation, the different aspects and features of the network are read in and interpreted to create a tensorflow neural network that depicts exactly what is shown on screen. During the network creation, the user interacts with the user interface, manipulating their network architectures and uploads their training input and outputs. The graphical network is then interpreted, and a tensorflow neural network of equivalence is created automatically. During training, the program checks the validity of the input and output shapes of both the neural network and the training data. After the check is complete, training begins, and the program reports the training progress to the user graphically. After training, the final loss for the neural network is shown.

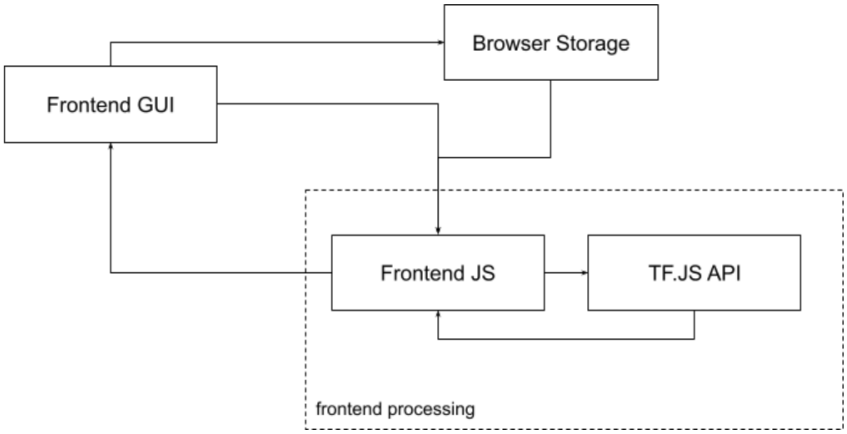


Figure 1. Overview of the system

```
// test model
document.querySelector('#testBtn').addEventListener('click', () => {
    if(model == null || trainIter == 0) return

    helpWanted = false

    // select a random input
    let length = tX.shape[0]
    let index = Math.floor(Math.random() * (length-1))

    let dataPoint = tf.tensor2d([tXarr[index]])

    let ys = model.predict(dataPoint)
    ys.print();

    let pred =  ys.arraySync()
    pred[0].forEach((e) => {
         Math.round(e*1000)/1000

    })

    for(let i = 0; i < pred[0].length; i++){
        pred[0][i] = Math.round(pred[0][i]*1000)/1000
    }

    changeStatus("Input Data: <b>" + tXarr[index] + "</b><br>Prediction: <b>" + pred + "</b>")
})
```

Figure 2. A segment of the code

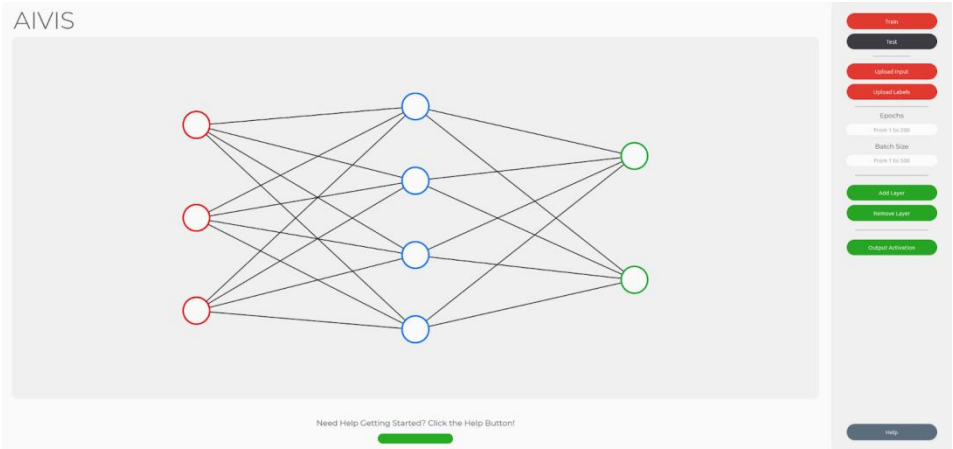


Figure 3. Screenshot and UI

In order to create the graphical interface to be scalable and user friendly, I used HTML CSS to style the interface, as well as JavaScript to provide its functionality. The user interacts with the HTML + CSS site, and uploads their input and label datasets. When the datasets are uploaded, they are stored in the browser's local storage for the ease of access and modification. When the train button is clicked, the application attempts to create and compile the neural network in TFJS based on what the user defined on the front-end. It then automatically partitions the data-set into training set, validation set, as well as testing sets. The program then uses the TFJS library to train the neural network, and sends analytic info to the front-end after every epoch, things like training and validation losses. After training is complete, the user can then choose to test the network by pressing test, which then the program will select a random data-point from the testing data-set, and run it through the neural network, and send back the output.

4. EXPERIMENT

4.1. Experiment 1

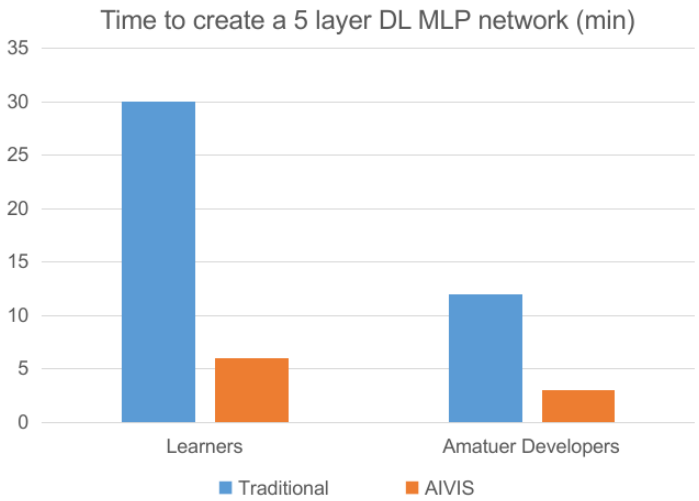


Figure 4. AIVIS is better for speed prototyping for both amateurs and beginners

My solution proves that the output resulting from this method is comparable, and even sometimes better than traditionally made neural networks.

```
Epoch 50/50
18/18 [==============================] - 0s 447us/step - loss: 0.4759 - accuracy: 0.5596
26/26 [==============================] - 0s 370us/step - loss: 0.4580 - accuracy: 0.5854
Accuracy: 58.54
```

Figure 5. Final loss of traditional network after 200 epochs and 25 batch size

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 5)                 20
_________________________________________________________________
dense_1 (Dense)              (None, 4)                 24
_________________________________________________________________
dense_2 (Dense)              (None, 5)                 25
_________________________________________________________________
dense_3 (Dense)              (None, 4)                 24
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 5
=================================================================
Total params: 98
Trainable params: 98
Non-trainable params: 0
```

Figure 6. Neural network structure of traditional network

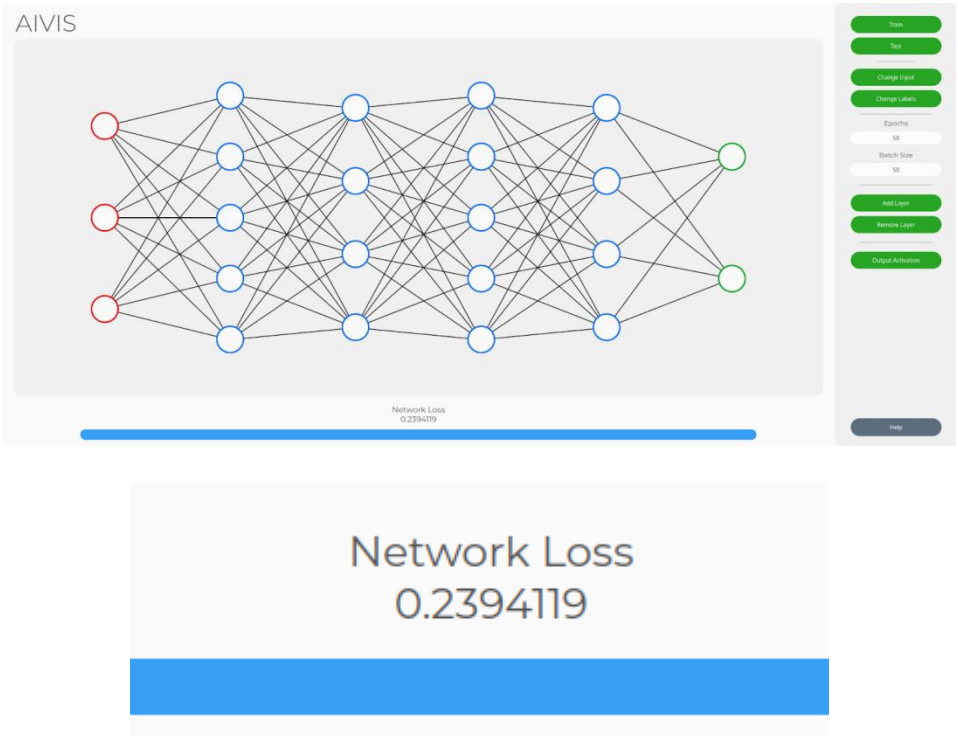


Figure 7. Final loss of the model trained with AIVIS, with the same hyper parameters

After about 50 training cycles with the same training parameters, data, and architecture, the average loss of the model trained with our method was about 30% lower than traditional. The optimizer, loss, and activation in each model is the same, but somehow the training results show that the visual made neural network performs better overall than the traditional one. Testing the same neural network out, it seems that AIVIS was better at avoiding overfitting, as the output of the neural network trained in AIVIS had more variety than that trained traditionally.

4.2. Experiment 2

Our solution not only dramatically decreases the amount of time to create neural networks, it also simplifies the steps needed to get a network working.

experiment settings: indicate that experiment design is scientific.

1. 1 participant
2. Has knowledge in neural networks and coding

The average amount of steps required to create and train a neural network in our solution is around 20-30 steps, depending on how complex the neural network is. However, when coding in python with Tensorflow, the user needs around at least 70 lines of code, and also needs to prepare the data, including partitioning and prepossessing it. Overall, the much simpler design and nature of our solution results in a faster development speed, as well as less room for technical errors.

4.3. Experiment 3

Even though our solution can offer a wide variety of neural networks, AIVIS can only create fully connected layers as of the moment, with many limitations on it due to graphical inadequacy.

Our solutions can solve simple problems such as predicting the best color of text to go above a colored background, XOR classifications, and other simpler problems. However, when more neurons are needed for a specific problem such as MNIST, our solution is currently incapable of doing so, as our graphical user interface does not allow for such a big network.

The experiments showed that AIVIS is a viable way to create neural networks graphically, with a much faster prototyping time and a better result than traditional neural networks. This is due to the minimalist nature of the program, where the user can easily and quickly change hyper parameters and train over and over again.

The result was surprising, as I was not expecting the graphical method to outperform the traditional method with the same architectures and training parameters. A possible source to this difference may be in how the code was run, and also how the weights are initialized. But currently with my code analysis, there is no difference between the two methods' way of weight initialization.

5. RELATED WORK

Neutron is a program that allows users to visually see their neural network in a node based fashion. It has a simple user interface and works with a variety of neural network types. It's way of representation is modular and dynamic, which is something our method needs to improve on. However, Neuron lacks the ability to modify or create neural networks.

Tensorflow Playground is an online application that allows users to tinker with neural networks of different sizes without hassle.

Neuron is a program which aims to solve regression, time series, binomial and multi-nominal classification problems for businesses and professional needs. It has a way to visualize data that the user is working with, and is suited towards companies which need rapid prototyping. However, Neuron is not an open source software, and is more aimed towards big businesses and professional work, and therefore does not suit an average user or student very well. Even though AIVIS may not perform as well in either functionality, efficiency, or scalability as Neuron, it is more education oriented and meant for small scale development. And while the free version of Neuron lacks the ability to export trained neural networks, AIVIS is planned to add that feature in a later version.

6. CONCLUSIONS

The goal is to create a simple, light weight, and easy to use GUI application for creating neural networks in order to mitigate the flaws of traditional neural network development. And when comparing the result of our method and traditional methods, our method has shown to be over three times faster in creating and training neural networks, while retaining network performance and accuracy. Our method shows that neural networks can be created and trained faster with comparable, and sometimes better, performance than traditional methods. As the use of a graphical interface provides a more intuitive and simple usage experience.

The current limitation of our method is that neural network training is not optimal due to the lack of customizability in hyper-parameters [15]. Users don't have much choice and selections over how the networks get trained, when to stop training (early stopping), what gradients to use, etc… Our method also lacks scalability due to its limiting UI design, making it more suitable as a demonstration tool than a real developer tool. The training speed and network complexity is also extremely lacking, as the neural network is trained on the client's side instead of on a dedicated server. This compromises the amount of complexity the neural network can be, as well as limiting the training speed by a wide magnitude.

As the user interface is not very well designed in this solution, our future work will feature a better, more modular, and more scalable interface; allowing for more complex neural networks to be made. Training will also be moved to a back-end dedicated server, which can potentially increase training and compiling speed by 300 to 700% depending on the network's complexity. Features such as exporting an integrate neural network file, such as TensorFlow's h5 files, will also be included for a more integrated and streamline workflow [6].

REFERENCES

[1] Steiner, David F et al. "Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer." The American journal of surgical pathology vol. 42,12 (2018): 1636-1646. doi:10.1097/PAS.0000000000001151

[2] Daniels, Norman. "Justice, health, and healthcare." American Journal of Bioethics 1.2 (2001): 2-16.

[3] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." European conference on computer vision. Springer, Cham, 2016.

[4] Honneth, Axel, and Avishai Margalit. "Recognition." Proceedings of the Aristotelian society, supplementary volumes 75 (2001): 111-139.

[5] Piramuthu, Selwyn, Michael J. Shaw, and James A. Gentry. "A classification approach using multi-layered neural networks." Decision Support Systems 11.5 (1994): 509-525.

[6] van Der Aalst, Wil MP, et al. "Workflow patterns." Distributed and parallel databases 14.1 (2003): 5-51.

[7] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[8] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

[9] Hu, Yu Hen, and Jeng-Neng Hwang, eds. "Handbook of neural network signal processing." (2002): 2525-2526.

[10] A. G. Salman, B. Kanigoro and Y. Heryadi, "Weather forecasting using deep learning techniques," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015, pp. 281-285, doi: 10.1109/ICACSIS.2015.7415154.

[11] Molina, Gabriel. "Stock trading with recurrent reinforcement learning (RRL)." CS229, nd Web 15 (2016).

[12] Feindt, M., and U. Kerzel. "The NeuroBayes neural network package." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 559.1 (2006): 190-194.

[13] Smilkov, Daniel, et al. "Tensorflow. js: Machine learning for the web and beyond." arXiv preprint arXiv:1901.05350 (2019).

[14] Abadi, Martín. "TensorFlow: learning functions at scale." Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming. 2016.

[15] MacKay, David JC. "Hyperparameters: optimize, or integrate out?." Maximum entropy and bayesian methods. Springer, Dordrecht, 1996. 43-59.

AN INTELLIGENT VIDEO EDITING AUTOMATE FRAMEWORK USING AI AND COMPUTER VISION

Haolin Xie[1] and Yu Sun[2]

[1]Northwood High School, 4515 Portola Pkwy, Irvine, CA 92620
[2]California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

At present, many video editing software have been created, but what they all have in common is that they require manual work to edit. And it takes a lot of time and the user needs to watch each frame before editing. In this paper, we have developed a program about AI intelligence. The most important point of this software is that it can automatically focus the face of a person and edit only selected clips of the person to make a complete video. Users only need to prepare the video they want to edit and a photo of the main character. Then, upload both to the software and AI Intelligence will automatically edit it, providing the user with a way to download and save it after editing the main character they need. we applied our application to an example video using the Marvel character Hawkeye as he appears in End Game, and tried many experiments with the clip, eventually we tried many experiments with the clip and finally got a video of our selected character. The results show that this software saves the user a lot of time and is highly efficient. All operations are carried out by AI.

KEYWORDS

Software, API, Face recognition.

1. INTRODUCTION

At present, many video editing software has been created, but what they all have in common is that they require manual work to edit. And it takes a lot of time and the user needs to watch each frame before editing. In this paper, we have developed a program about AI intelligence. The most important point of this software is that it can automatically focus on the face of a person and edit only selected clips of the person to make a complete video. Users only need to prepare the video they want to edit and a photo of the main character. Then, upload both to the software and AI Intelligence will automatically edit it, providing the user with a way to download and save it after editing the main character they need. The results show that this software saves the user a lot of time and is highly efficient AI carries out all operations. The purpose of the theme of this software is to provide the user with the opportunity to edit a video of a specific person of their choice, and the edited video will only feature the person chosen by the user.

There is no software or web page similar to our software available to users. But there are many methods and systems that can perform face recognition. Some of the face detection techniques and systems that have been proposed to improve surveillance and help track criminals and terrorists, allow the user to enhance personal security and also extend to convenience. For example, many Apple products are available with Face ID, which eliminates the need for users to

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 201-209, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121216

enter a password, and also enhances user privacy and security. MediaPipe provides an ultra-fast face detection solution with six landmarks and multi-face support [3][4]. It is based on BlazeFace, a lightweight and well-performing face detector designed for mobile GPU inference. Facial recognition technology doesn't always work as well as it should [1]. Facial recognition systems can be affected by poor lighting or low image quality, which can prevent accurate recognition, and this improves the computer's algorithm and error rate [2][5]. Or the data may not match the person's nodes because the camera angle is obscured; this creates errors when matching facial prints cannot be verified in the database. Because all faces are physically similar to each other, each person's facial parts are in the exact location [6].

In this paper, we follow the same line of research by face recognition. Our goal is to improve the accuracy and degrade the error rate of face recognition and provide the user with a video with only the parts that the user needs, and do so with as little error rate as possible to save the user's time and improve productivity. First, identify and locate the facial features to get the coordinates of the eyes, ears, cheeks, nose, and mouth of each detected face. And get the contours of the detected face and its eyes, eyebrows, lips, and nose. Third, the images and videos are grayed out to avoid overexposure or other elements that cause lighting problems. The whole system will not affect the success of the final result after ensuring that the lighting in the original image is uniform and that the facial features can be accurately detected. This is also a great help to our software, as we believe that the system can accurately identify the people in the video and edit the video.

We use the Face Detection API that is now available to help our software a lot and apply it to our software. We practice two different methods to detect the success of the software. First, we used the Marvel character Hawkeye and the Avengers: Endgame trailer for this test. After we uploaded both of them, the video generated by the software system after editing was Hawkeye, and after several attempts, the system generated the same video as the previous ones. We then used the same photo to test all the videos of Hawkeye in Marvel, and our software was able to generate the same video with only Hawkeye. Secondly, we used different characters from the Endgame trailer to test the same software system to edit the video successfully, and there were no errors, which shows the usefulness of our method and system.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the concluding remarks, as well as points out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Making the Software Easier to Use

When I chose the software to make this theme, I was confused because I would struggle with what to add to make the software look better or easier to use. When designing the interface of the software, I struggled with the layout, such as where to place the icons for uploading photos and videos. I wasn't sure how to use the GUI and the information it required to exchange between software and the amount of memory and processing power they needed to use. The challenge is to use the GUI to make the App look aesthetically better and organize its features.

2.2. Passing Every Frame

After executing the Face Detection API for face recognition, the next challenge is to pass every frame and every frame in which the person appears to the next code for the operation. And how to stitch together the images of each appearance of the character, and then need to avoid is the character after a flash and then no picture, so also need to edit out the character appeared before and after each video left a second or two, so that the character will not be a flash and thus can not see the appearance of the character, these are the challenges and difficulties to face.

2.3. Developing Software into Android and IOS

After the system and AI part of the software is solved, it needs to be developed into Android and IOS software [7]. And maintenance and building are the issues that need to be faced. The development of Android and IOS applications involves a lot of complexity. For example coding multiple files for one screen, problem-solving and bug fixing, and compatibility with all Android devices including IOS as well [8]. Each step is a new challenge and after solving one problem another new one may appear, so this is the biggest difficulty and one of the most time-consuming parts of developing this software.

3. SOLUTION

We needed to solve the problem of how to use the GUI to make the application look more beautiful and organize its functions, how to stitch together the images of the characters each time they appear, and then we needed to avoid having no images after the characters flicker and we needed to develop the software for Android and IOS and make sure that most of the IOS and Android devices are compatible [9]. We want to save more time for users by providing software for users to edit their videos to improve efficiency. And we want to demonstrate through this research that face recognition is widely available and can be integrated into people's daily lives, just like our software, which uses face recognition to edit a complete video.

Python GUI project is a simple API for developers to create user interfaces using native elements of Python applications [10]. As a lightweight API, not much code is required between the application and the target platform, which makes it more efficient than many other frameworks on this list. Our software needs to use the GUI to make the application look good and to lay out our interface. Swift is a general-purpose, multi-paradigm, compiled programming language developed by Apple Inc. and the open-source community.

```
_loadHistory() async {
  SharedPreferences prefs = await SharedPreferences.getInstance();
  List dates = [];
  print(prefs.containsKey('date'));
  if (prefs.containsKey('date')) {
    dates = prefs.getStringList('date')!;
    for (var date in dates) {
      _history[date] = json.decode(prefs.getString(date)!);
    }

    print(_history.keys);
    setState(() {});
  }
}
```

Figure 1. Code of how history is stored

Here is an example of code I wrote for my project: The code on the right shows how the video history is accessed/stored.

```
Future<void> _setVideoController() async {
  File file = File(widget.videoFilePath);
  if (file != null && mounted) {
    VideoPlayerController controller;
    print('play video ');
    if (kIsWeb) {
      controller = VideoPlayerController.network(file.path);
      print('network:' + file.path);
    } else {
      controller = VideoPlayerController.file(File(file.path));
      print('file:' + file.path);
    }
    setState(() {
      _controller = controller;
    });
  }
}
```

Figure 2. Code of loading video information

The code on the right shows the class in charge of loading video information to be edited by the programmed AI.

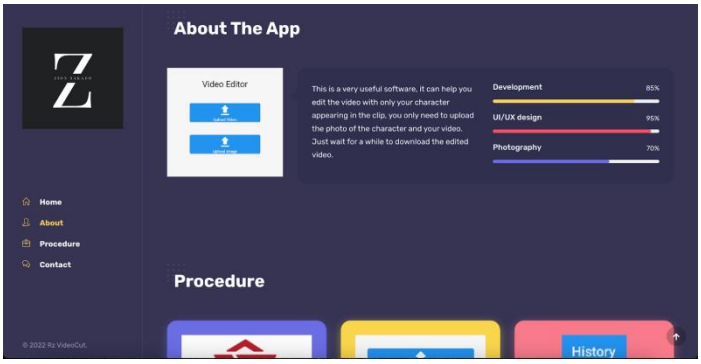


Figure 3. Screenshot of the App

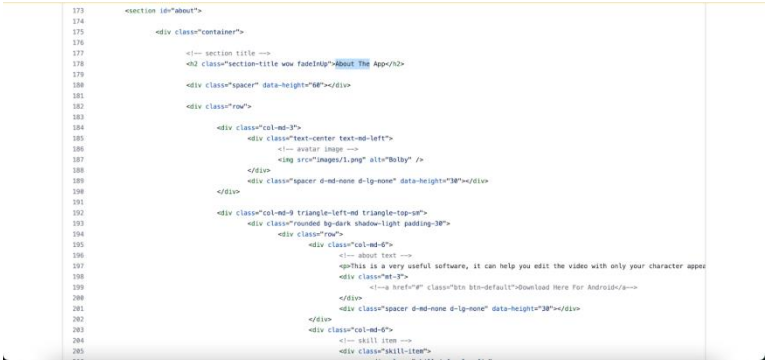


Figure 4. Screenshot of code

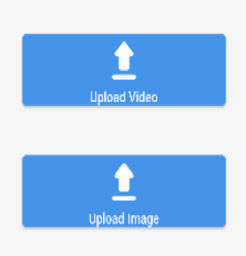


Figure 5. Screenshot of upload page

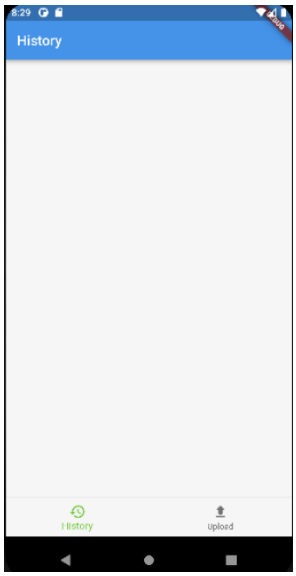


Figure 6. Screenshot of history page

There are many free panels or templates available on the web for the general public, and it is easy to reposition or re-color each button using the templates; there is not much code required between the GUI application and the target platform. This makes it more efficient than other programming languages or methods of creating frameworks, and it is good for solving the aesthetics of our software and organizing its functionality. Then the face recognition will scan one side of the video uploaded by the user, edit it and send it back to the port so that the user can download and save the edited video. AWS is a widely adopted cloud platform that provides a variety of on-demand operations such as computing power, database storage, content delivery, etc. This step solves the problem of receiving videos as well as photos uploaded by users and then proceeding to the next step. Developing software for IOS and Android both require different programming languages. development for IOS can be more cumbersome and only requires improving the readability of constants. The best way to use structures in Swift is to create a file for all constants in the application.

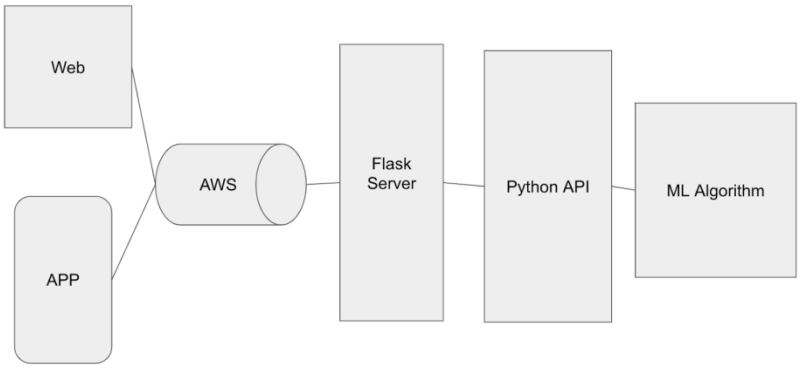


Figure 7. Overview of the system

The development of this software uses a GUI for aesthetic interface and functionality, AWS for computation and database storage, APIs to use it with any other programming language, and ML algorithms to allow the machine to follow a set of instructions to perform a task; the algorithms also help the machine to choose and decide which set of instructions will produce better results.

Mr. Evan Gunnell helped me and guided me together to solve the problem of getting the frame after face recognition and each frame and composing it into a video, as well as solving the problem that there is no more video when the character flashes by. And Mr. Armando helped me with the aesthetics and layout of the software interface and developing it for Android and Apple. We use the same character and test it several times on the same video to see if the edited video is the same and alive with errors. Similarly, using different characters in the same video for multiple tests, the clips came out with the corresponding characters.

It ensures the correctness of the software and the multifaceted connection of the tools to make the whole software work properly. We use the current face detection to help our software a lot and apply it to our software.

There are many methods as well as tips available on the internet when it comes to problems, but sometimes these may not be useful and we may need to incorporate all the tips and methods in order to function.

4. EXPERIMENT

4.1. Experiment 1

My solution was to use the same person in the same video for 5-8 tests and also use different photos of the person but use the same video for multiple tests. This way we can test the correctness of this software as well as the error rate. First, we prepared the trailers of Hawkeye and Avengers: Endgame with Marvel characters, and uploaded them to the software. AWS will start the operation to store the photos and videos and then do the face recognition and edit them to download them and repeat the above steps to ensure the accuracy rate. The final result is that the edited videos are identical and no other characters appear inside.

4.2. Experiment 2

My second solution was to use the same character but change different photos but still use the same video for multiple tests. This way I could also test the correctness and error rate of the

software [11]. The other solution is to use different characters but the same video for the test. First, again, we prepare many photos of Hawkeye and upload them to the software for testing, use each photo to edit the video, and then repeat the test with different photos. After several tests, the edited video was the same as the other edited video and there was no difference, so the experiment was successful.

The results of the experiment are up to expectations because after uploading new photos and videos, the face recognition will re-identify one side of the photos and videos instead of existing in one data, it will not overlap with the previous data, so it will not lead to the next editing of the video and face recognition and thus will not affect the results of the test. The results of the test show that the face recognition and editing of the composite video are successful [12]. We tried many different factors, but the software achieved the desired goal.

5. RELATED WORK

Limin Wang and his team study appearance and relational networks for video classification, focusing on learning video representation in an end-to-end manner [13]. Their team also focuses on the goal of each frame, but unlike our software, they connect the appearance for spatial modeling. And their team has gained significant improvements over 3D convolution in Spatio-temporal feature learning.

S Janhavi and Chandra Sekhar Malepati use video classification for real-time human activity recognition, which is very similar to our project because we need to edit a video after face recognition [14]. And theirs is the ability to recognize the activity of others, and again, both are in a moving state. I think their project is interesting because it can detect what kind of movement people are doing, which will help in the future.

Paul Viola & Michael J. Jones has created powerful real-time face detection, which is similar to my software project, and also sports face recognition [15]. But the difference is that their project is real-time while my software requires user uploads. And they used the AdaBoost learning algorithm (Freund and Schapire, 1995) to build a simple and efficient classifier for selecting a small number of key visual features from a large number of potential features, which greatly improves the speed of face detection and the accuracy of real-time detection.

6. CONCLUSIONS

This software is a program about AI intelligence. The most important point of this software is that it can automatically focus on a person's face and edit only selected clips of that person to create a complete video. Users only need to prepare the video they want to edit and a photo of the main character. Then, both are uploaded to the software and AI Intelligence automatically edits them, providing a way for users to download and save them after editing the desired protagonist. And AWS was used to store the data, with the goal of providing users with a way to upload as well as download the edited video. The software was tested repeatedly to ensure accuracy and downgrade the error rate. The results show that the software saves users a lot of time and is efficient, with all operations performed by artificial intelligence.

This software is a program about AI intelligence. The most important point of this software is that it can automatically focus on a person's face and edit only selected clips of that person to create a complete video. Users only need to prepare the video they want to edit and a photo of the main character. Then, both are uploaded to the software and AI Intelligence automatically edits them, providing a way for users to download and save them after editing the desired protagonist.

And AWS was used to store the data, with the goal of providing users with a way to upload as well as download the edited video. The software was tested repeatedly to ensure accuracy and downgrade the error rate. If there is more time, I will improve the interface and content of the software and add more information for the And adding details makes the interior of the software look more concise. If I were to continue to work on this project, I would add these three things: 1. add subtitles. 2. Improve the GUI. 3 .Background Music.

And this software is currently available for download on the Google Play Store, link: https://play.google.com/store/apps/details?id=com.codingminds.video_editor_flutter

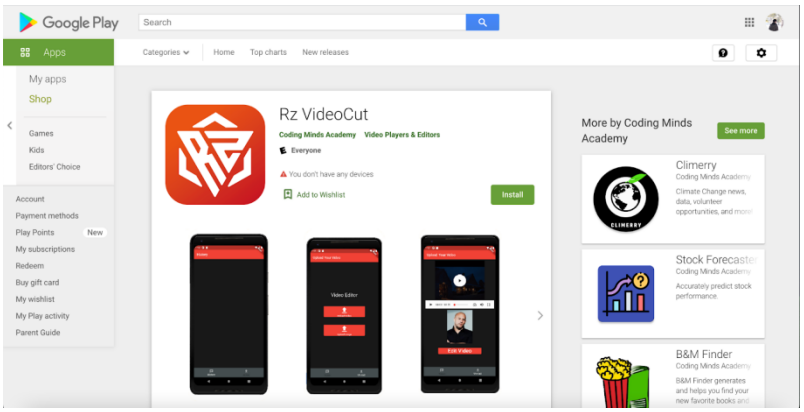


Figure 8. Screenshot of App

REFERENCES

[1] Kaur, Paramjit, et al. "Facial-recognition algorithms: A literature review." Medicine, Science and the Law 60.2 (2020): 131-139.
[2] Gorodnichy, Dmitry O. "Facial recognition in video." International Conference on Audio-and Video-Based Biometric Person Authentication. Springer, Berlin, Heidelberg, 2003.
[3] Lugaresi, Camillo, et al. "Mediapipe: A framework for perceiving and processing reality." Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR). Vol. 2019. 2019.
[4] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).
[5] Duchaine, B. "Individual differences in face recognition ability: Impacts on law enforcement, criminal justice and national security." Psychological Science Agenda (2015).
[6] Liu, Haowei. "Face technologies on mobile devices." Facial Detection and Recognition on Mobile Devices (2015): 11-38.
[7] Enck, William, et al. "A study of android application security." USENIX security symposium. Vol. 2. No. 2. 2011.
[8] Chin, Erika, et al. "Analyzing inter-application communication in Android." Proceedings of the 9th international conference on Mobile systems, applications, and services. 2011.
[9] Liu, Jianye, and Jiankun Yu. "Research on development of android applications." 2011 4th International Conference on Intelligent Networks and Intelligent Systems. IEEE, 2011.
[10] Podrzaj, Primoz. "A brief demonstration of some Python GUI libraries." Proceedings of the 8th International Conference on Informatics and Applications ICIA2019. 2019.
[11] Madhuri, T., and P. Sowjanya. "Microsoft Azure v/s Amazon AWS cloud services: A comparative study." International Journal of Innovative Research in Science, Engineering and Technology 5.3 (2016): 3904-3907.
[12] Real, Esteban, et al. "Automl-zero: Evolving machine learning algorithms from scratch." International Conference on Machine Learning. PMLR, 2020.

[13] Wang, Limin, et al. "Appearance-and-relation networks for video classification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[14] Janhavi, S., and Chandra Sekhar Malepati. Real Time Human Activity Recognition with Video Classification. No. 7377. EasyChair, 2022.
[15] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154.

CRYPTOGRAPHIC ALGORITHMS IDENTIFICATION BASED ON DEEP LEARNING

Ruiqi Xia[1], Manman Li[2] and Shaozhen Chen[2]

[1]Department of Cyberspace Security,
Information Engineering University, Zhengzhou, China
[2]State Key Laboratory of Mathematical Engineering and Advanced Computing,
Kexue Avenue, Zhengzhou, China

ABSTRACT

The identification of cryptographic algorithms is the premise of cryptanalysis which can help recover the keys effectively. This paper focuses on the construction of cryptographic identification classifiers based on residual neural network and feature engineering. We select 6 algorithms including block ciphers and public keys ciphers for experiments. The results show that the accuracy is generally over 90% for each algorithm. Our work has successfully combined deep learning with cryptanalysis, which is also very meaningful for the development of modern cryptography and pattern recognition.

KEYWORDS

Deep learning, Cryptography, Feature engineering, Residual neural network, Ciphers identification.

1. INTRODUCTION

1.1. Motivation

Cryptography is widely used in privacy protection and communication security with the development of computer techniques [1]. The main target of cryptanalysts is to recover the keys from the ciphertext. However, they can only acquire ciphertext through public channels most of the time. Cryptanalysts determine the scopes of encryption algorithms through monitors, reverse analysis or Side Channel Attack of ciphers. They will know which several possible ciphers are used at the moment only based on ciphertext. Knowing the encryption algorithms assists cryptanalysts in recovering the keys. Therefore, it is very important to identify the encryption algorithms in advance for cryptanalysis.

1.2. Related Work

Thanks to the development of artificial intelligence, we can use machine learning techniques to solve the problems of cryptanalysis. Some researchers have already studied the identification of cryptographic algorithms. In 1998, Ramzan proposed that neural networks could be used for identifying ciphers [2]. Subsequently, Dileep, et. al [3-5]successfully identified DES, Blowfish and some other algorithms by Support Vector Machine and Decision Trees. However, the results became unsatisfactory when the keys were changed. Recently, Mishra, et. al [6,7] applied PART, C4.5 to the ciphers identification and the accuracy reached over 80%.

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 211-219, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121217

Although deep learning has become very popular in many subjects, we notice that few scholars consider identifying the algorithms by deep neural network. In 2021, Sandeep, et. al tried to apply convolutional neural network to the identification of several block ciphers. However, the work did not show a detailed scheme[8]. On the other hand, the cryptosystem often uses random keys for safety while the previous work was unsatisfactory when the keys were unfixed. This makes the work less valuable in application. Therefore, there are many challenges for us to investigate further.

1.3. Our Contribution

It is necessary to investigate how to improve the accuracy of identification in the conditions of random keys and give a detailed and executable scheme based on deep learning. Hence in this paper we construct a novel model of cryptographic algorithms identification based on feature engineering and residual neural network. We select 6 algorithms including block ciphers and public keys ciphers for experiments. The accuracy is generally over 90% for each algorithm in the conditions of random keys. Compared with the former work, not only do we successfully apply the neural network to ciphers identification, but also improve the results of experiments in the conditions of random keys. Such technique assists cryptanalysts in recovering the keys and obtaining the plaintext. Our work also provides a new direction for the development of pattern recognition.

1.4. Arrangement

The arrangement of the paper is shown as follows. The first section introduces the background of our work and the main contribution. The second section briefly describes the cryptographic algorithms. We illustrate the model of identification in the third section, which includes feature engineering, residual neural networks and so on. The fourth section is the experiments of identification based on our approach. The last section is the conclusion.

2. CRYPTOGRAPHIC ALGORITHMS

Modern Cryptography could be divided into symmetric cryptography and asymmetric cryptography. These cryptographic algorithms make remarkable contribution to information security and privacy. In this work we select 4 block ciphers and 2 public keys ciphers for experiments. All of them are commonly used in reality.

2.1. Block Ciphers

Block ciphers divide the plaintext into fixed-length blocks and then encrypt or decrypt the encoded block sequences using the same keys. These algorithms are widely used in the protection of hardware, digital signature and so on. Nowadays, lightweight block ciphers become one of the most useful applications in IoT devices [9].

AES(Advanced Encryption Standard)[10]. The construction of AES is SPN (Substitution Permutation Network) structure. The block length is 128 bits. The key length is 128/192/256 bits. The numbers of rounds are 10/12/14. Here we use AES-128.

KASUMI [11]. The construction of KASUMI is Feistel structure with 64 bits block length and 128 bits key length. The number of rounds is 8. KASUMI algorithm was designed for the basis of the 3GPP (3rd Generation Partnership Project).

3DES (Triple Data Encryption Standard)[12]. 3DES was developed to overcome the shortages of DES. The block length is 56 bits and the key length is 168 bits.

PRESENT [13]. PRESENT belongs to lightweight block ciphers. The construction is SPN structure. The block length is 64 bits and the key length is 80/128 bits. The number of rounds is 31.

2.2. Public Keys Ciphers

Public keys ciphers encrypt the plaintext with the public key and decrypt with the private key. Public key cryptography was designed by Whitfield Diffie and Martin Hellman in 1976[14]. We use RSA and ElGamal algorithms for the research.

RSA [12]. RSA algorithm is based on the decomposition of large numbers. The private keys are computationally difficult to require from the public keys. The key length is usually 1024 bits or 2048 bits.

ElGamal[15]. ElGamal algorithm is another public key cryptography. It is based on calculating the discrete logarithm over a finite field. This algorithm is broadly used in digital signature.

3. MODEL OF IDENTIFICATION OF CIPHERS

3.1. Design for the Model

Figure 1 shows the construction of our model. The model of identification mainly consists of three parts with two stages: obtaining the original datasets, feature engineering, and deep neural network classifier. The ciphertext is encrypted by each algorithms in the conditions of random keys. After obtaining the ciphertext the feature engineering extracts the feature indices of it. Each feature vector is attached with the corresponding labels. Finally the feature files are inputted into the model for training and testing. The whole process includes training phase and testing phase. We package the whole process to form an end-to-end framework for application.

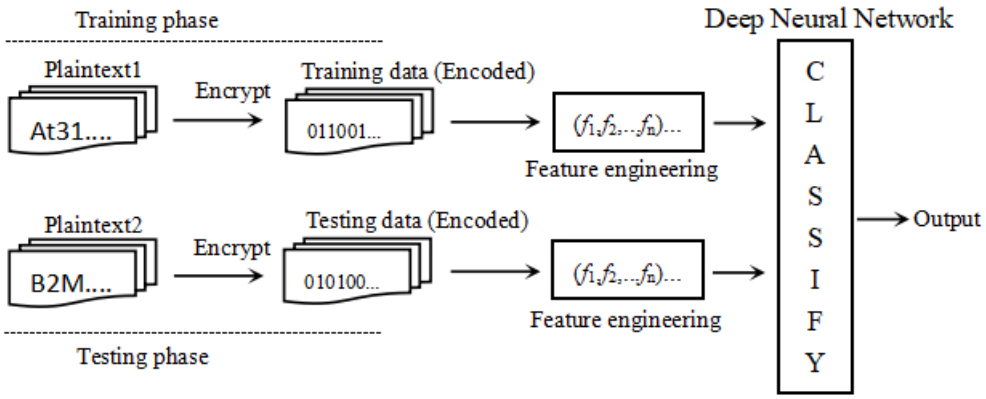


Figure 1. Model of Identification

In our experiments, 6 algorithms are encrypted by random keys. Meanwhile, the block ciphers are encrypted in CBC (Cipher Block Chaining) mode for safety. Compared with [6,7], the conditions of our experiments are far more strict, which make our work more meaningful.

3.2. Feature Engineering

Ciphertext seems diffused and random, especially when keys are unfixed. Feature engineering helps find out the characteristics of the data and helps the models work better. Although the deep neural networks could automatically extract the features of the data, we can make it in advance to help the networks understand the target and work more effectively. Feature engineering is one of the most essential steps in our work. We select three randomness indices published by NIST(National Institute of Standard and Technology) [16] in our experiments.

Frequency within blocks index. The frequency within blocks index collects the proportion of 0 or 1 in each sub-block divided from the ciphertext blocks.

Runs index. The index collects the sum of each length of run in the sequence. A run of length k consists of exactly k identical bits and is bounded before and after with a bit of opposite value.

Serial index. The index gets the sum of all sub-sequences of the ciphertext. A sequence of length m has 2^m sub-sequence. If there is no such sub-sequence, note it with 0.

Feature engineering extracts such three feature indices to make up the feature vectors. The corresponding labels are attached to each feature vectors. Then we input the feature vectors into the neural network for training and testing. The end-to-end framework we construct for application will package the feature engineering so it is more convenient to use the model.

3.3. Residual Neural Network

Thanks to the development of artificial intelligence, deep learning technology has been successfully combined with cryptography such as recovering the keys and simulation encryption [17]. In 2019, Gohr applied residual neural network to cryptanalysis [18], which improved the traditional cryptanalysis significantly. Inspired by his work, we also use such neural network for ciphers identification.

Residual neural network introduces a residual tower which helps the model works better when the depth of network increases. Such networks avoid degradation by using identity mappings [19]. Here we choose ”ReLU” as the activation and ”Conv1D” as the basic convolution layer. Most importantly, we use cross entropy as the loss function. It is shown as follow(y_i means the real value and a_i means the prediction value).

$$\text{cost} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln a_i + (1-y_i)\ln(1-a_i)\right] \tag{1}$$

Figure 2 shows the structure of residual neural network.

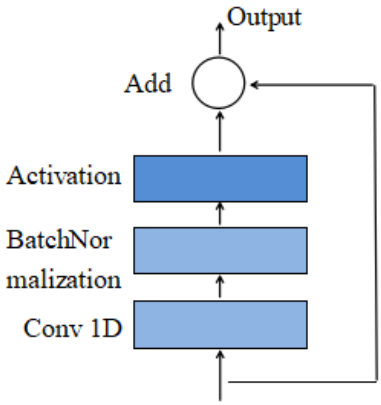


Figure 2. Residual neural network

4. EXPERIMENTS AND RESULTS

The hardware configurations of the experiments are Windows 10 system with 2 Intel Core i7 processors having 16 cores each and one NVidia GEFORCE RTX 2080Ti GPU, 256 RAM. The software we used is Python 3.7 with Keras 2.3.1.

The plaintext is selected from the texts in Open American National Corpus (OANC). We select some commonly used texts or sentences as the plaintext such as idioms. The size of plaintext is around 1.2 GB, which is slightly more than previous work. Then the plaintext files are divided into 1000 parts, which are about 1.1 MB. The keys are changed at each time of encryption.

After obtaining the ciphertext, we extract the feature indices of each part, which has about 2600 feature vectors. The feature vectors are attached with the corresponding labels. We use 0 to 5 to represent such 6 algorithms. Then the feature vectors are inputted into the neural network for training and testing. We set the epochs 200. The proportion of training sets and testing sets is 6:4. The learning rate is 0.01 and the batch size is 500.

4.1. Results

The results of the classifier are expressed by accuracy, precision and recall [20]. *TP* (True Positive) represents the number of right examples which are sentenced to right ones. *TN* (True Negative) represents the number of right examples which are sentenced to wrong ones. *FP* (False Positive) represents the number of wrong examples which are sentenced to right ones and *FN* (False Negative) represents the number of wrong examples which are sentenced to wrong ones. Hence accuracy means the ratio of all samples which are correctly sentenced in the entire dataset.

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{2}$$

Precision means the ratio of *TP* in the samples sentenced to be right.

$$precision = \frac{TP}{TP + FP} \tag{3}$$

Recall refers to the proportion of *TP* in the whole right samples.

$$recall = \frac{TP}{TP + FN} \tag{4}$$

The results are shown in Table 1 and Figure 3.

Table 1. Results of identification.

Algorithm(Label)	Precision	Recall	Accuracy
AES-128(0)	89.23%	72.1%	90.05%
KASUMI(1)	92.36%	66.38%	91.46%
3DES(2)	90.6%	69.05%	93.65%
PRESENT(3)	93.12%	65.28%	95.72%
RSA(4)	88.45%	74.51%	88.79%
ElGamal(5)	89.77%	74.21%	90.93%

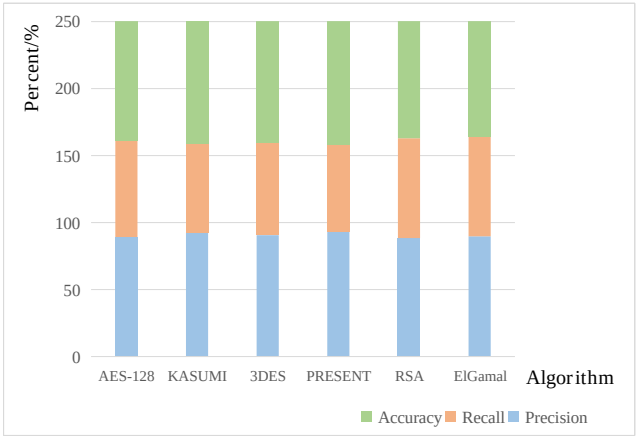


Figure 3. Percentage Stack histogram of identification

We find that the model works effectively in the experiments. Each algorithm's accuracy is higher than 90% generally. Even the worst group's accuracy is also higher than 85%. The precision of identification is also satisfactory, which is a little lower than the accuracy.

On the other hand, the recall is relatively lower, which is between 65% and 75%. According to the definitions of the three indices, when precision or accuracy becomes higher the recall becomes lower.

The average accuracy is 100/6 ≈ 17%. Therefore, our experiments' results are better than the average. This indicates that the model we construct is efficient indeed. Such technique can be used in practice.

The results of identification of block ciphers are better than public keys ciphers, which is around 5% higher in accuracy. The phenomenon means that block ciphers are more easily to be distinguished compared with public keys ciphers.

4.2. Discussion

The algorithms influence the results of identification significantly. Based on the results, we conclude that the block ciphers are more easily to be identified by classifiers. The ciphertext of

block ciphers may have more characteristics or differences which can be identified by neural networks. So we recommend that people should avoid using the algorithms which have many characteristics in the ciphertext for privacy safety.

Although we use random keys for encryption and CBC mode for block ciphers, the results of identification are still remarkable and stable. The gap between the highest accuracy and the lowest accuracy is less than 10%, which is more stable than previous work. It indicates that although deep neural network model needs more data, the model has stronger generalization ability which can be applied to more algorithms.

However, the recall seems unsatisfactory. Recall means the ability to find the correct samples in all correct sets, while precision means the ability to judge all correct samples. Hence our model ought to improve the ability in finding correct samples further. In addition, the model cannot give a correct classification when the ciphertext is not encrypted by the included algorithms. These shortages are worth improving in the future.

5. CONCLUSIONS

We study the identification of cryptographic algorithms in this work. The model of identification is based on the residual neural network and the feature engineering. The neural network classifier trains and tests the 3 feature indices extracted from ciphertext encrypted by 6 ciphers and random keys. Our experiments have successfully applied deep neural network to ciphers identification in detail. Compared with the previous work, not only do we improve the accuracy by around 10% in the conditions of random keys, but also investigate more complex ciphers including block ciphers and public keys ciphers.

Identifying cryptographic algorithms is one of the essential steps for keys recovery and it is useful in the application of cryptanalysis. According to Kerckhoffs' s assumption [21], cryptanalysts ought to know the encryption algorithms as well as other details. Therefore it is meaningful to identify the algorithms effectively in reality. Our work helps cryptanalysts know the encryption algorithms at the moment. So they will find out the most efficient method for recovering the keys more easily. It is also a novel application in pattern recognition, which provide some new directions for the development of deep learning.

In the future, we will possibly consider improving our model further. First, it is necessary to investigate whether we can reduce the size of data compared with the traditional machine learning approach. Second, the ability to judge all correct samples is still need to be improved. Meanwhile, it is necessary to apply our model to the identification of more cryptographic objects such as the modes of operation of block ciphers, etc. Hence there is much more research for identification of cryptographic algorithms.

ACKNOWLEDGEMENT

Thanks my fellows in State Key Laboratory of Mathematics and Advanced Computing! This paper is supported by Open Fund Project of the State Key Laboratory of Mathematical Engineering and Advanced Computing (No. 2019A08).

REFERENCES

[1] Coron, J. S. (2006). What is cryptography?. *IEEE security & privacy*, 4(1), 70-73.
[2] Ramzan, Z. (1998). On using neural networks to break cryptosystems. Manuscript.

[3] Dileep, A. D., & Sekhar, C. C. (2006, July). Identification of block ciphers using support vector machines machines. In The *2006 IEEE International Joint Conference on Neural Network Proceedings* (pp. 2696-2701). IEEE.

[4] Manjula, R., & Anitha, R. (2011, January). Identification of encryption algorithm using decision tree. In *International Conference on Computer Science and Information Technology* (pp. 237-246). Springer, Berlin, Heidelberg.

[5] Chou, J. W., Lin, S. D., & Cheng, C. M. (2012, October). On the effectiveness of using state-of-the-art machine learning techniques to launch cryptographic distinguishing attacks. In *Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence* (pp. 105-110).

[6] Barbosa, F., Vidal, A., & Mello, F. (2016). Machine learning for cryptographic algorithm identification. *Journal of Information Security and Cryptography (Enigma)*, 3(1), 3-8.

[7] De Mello, F. L., & Xexeo, J. A. M. (2016). Cryptographic algorithm identification using machine learning and massive processing. *IEEE Latin America Transactions*, 14(11), 4585-4590.

[8] Pamidiparthi, S., & Velampalli, S. (2021). Cryptographic algorithm identification using deep learning techniques. In Evolution in Computational Intelligence (pp. 785-793). Springer, Singapore.

[9] Hatzivasilis, G., Fysarakis, K., Papaefstathiou, I., & Manifavas, C. (2018). A review of lightweight block ciphers. *Journal of cryptographic Engineering*, 8(2), 141-184.

[10] Abdullah, A. M. (2017). Advanced encryption standard (AES) algorithm to encrypt and decrypt data. *Cryptography and Network Security*, 16, 1-11.

[11] Kim, H. W., Park, Y. J., Kim, M. S., & Ryu, H. S. (2002). Hardware implementation of the 3GPP KASUMI crypto algorithm. In *Proceedings of the IEEK Conference* (pp. 317-320). The Institute of Electronics and Information Engineers.

[12] Singh, G. (2013). A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. *International Journal of Computer Applications*, 67(19).

[13] Bogdanov, A., Knudsen, L. R., et. al. (2007, September). PRESENT: An ultra-lightweight block cipher. In *International workshop on cryptographic hardware and embedded systems* (pp. 450-466). Springer, Berlin, Heidelberg.

[14] Diffie, W., & Hellman, M. E. (2019). New directions in cryptography. In *Secure communications and asymmetric cryptosystems* (pp. 143-180). Routledge.

[15] ElGamal, T. (1985). A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE transactions on information theory*, 31(4), 469-472.

[16] Soto, J. (1999, October). Statistical testing of random number generators. In *Proceedings of the 22nd national information systems security conference* (Vol. 10, No. 99, p. 12). Gaithersburg, MD: NIST.

[17] Benamira, A., Gerault, D., Peyrin, T., & Tan, Q. Q. (2021, October). A deeper look at machine learning-based cryptanalysis. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 805-835). Springer, Cham.

[18] Gohr, A. (2019, August). Improving attacks on round-reduced speck32/64 using deep learning. In *Annual International Cryptology Conference* (pp. 150-179). Springer, Cham.

[19] Thorpe, M., & van Gennip, Y. (2018). Deep limits of residual neural networks. *arXiv preprint* arXiv:1810.11741.

[20] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint* arXiv:2010.16061.

[21] Xuejia, L. A. I. (2001). Basic Concepts of Cryptography. *Advanced Security Technologies in Networking*, 178, 21.

AUTHORS

Ruiqi Xia. Graduate student at the Institute of Cyberspace Security, Information Engineering University.

Manman Li. PhD of State Key Laboratory of Mathematical Engineering and Advanced Computing.

Shaozhen Chen. Professor of State Key Laboratory of Mathematical Engineering and Advanced Computing.

A DATA-DRIVEN MOBILE COMMUNITY APPLICATION FOR BOOK RECOMMENDATION AND PERSONALIZATION USING AI AND MACHINE LEARNING

Lulu Zha

Crean Lutheran High School, Irvine CA, USA

ABSTRACT

Knowing a movie or a book fits your flavor without finishing the whole film or the book? Although there are many ways to find a summary of a film or a book, having an app that generates needed information according to the genre will make things much more manageable. This paper develops a mobile app named Book and Movie Search that uses API or the online database to generalize data such as authors, plots, overview, and more with a few clicks. The results show that within seconds, a list of information will show according to movies and books, and a qualified way to find information using the Book And Movie Search app. For example, if one decided to buy a book named Flipped and did not have time to finish the whole book, he can enter the name on the app. It will generate a book summary that quickly gives him more information about it and help him decide whether he wishes to make the purchase.

KEYWORDS

technology, movie, book, search engine.

1. INTRODUCTION

Books and Movie Search app using the online database or the API and coded it with flutter in android studio. This app contains three split screens as well as a welcome page. The first split-screen has five pictures that list several popular books; the second split-screen is the book search, allowing people to search for books. There are four recommended books, including Animal Farm, Harry Potter, Flipped and lemons, and the author's publication date. At the very top of the screen, the search engine will gather information such as author, overview, publication date, and country of the publication as the user enters the book's name. The movie search engines contain similar content on the last and third split-screen. It has four recommended books, including Spiderman-no way home, Avengers, and Encanto, along with the director and the year it was published. As the user enters the movie's name, the director, plot, overview, and year of publication will show up. This app is designed for people to use before watching or reading a book. It will be easy for people to determine whether they wish to read/manage the content with a summary. For future books and movies, the app will automatically renew itself. It will update the new information and update the database used. The reason for me to build an app that generates books and movies summarizes that I think it will save a lot of time. To have a summary quickly determines whether someone is interested in the whole story or not.

Once I wanted to read Animal Farm, I decided to search whether it was fun. I got too much information about this book, specific to every character. Or I get a lot of related but not helpful

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 221-230, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121218

information such as the biography of George Orwell. After that, I found searching things on google, especially summaries of books, very time-consuming, and I wanted to create something easier to use. I hope this app will generate the most helpful information and provide a simple overview of the whole story, and they can choose whether to read the book or not.

One related method of searching this topic was Douban, an app that generates the most popular movies and books through commentary with the audience. Douban provides discussion groups and servers based on books and films, and through that, there will be a detailed, editable everyone information list that the users can view. However, no "official" summary exists for any books and movies; it is more like a discussion on a particular book/movie. Although they proposed to make an environment that allows movie and book lovers to share their thoughts, the accuracy of the content was limited. Douban, as users enter the names of books and films, will show descriptions of the movies and books. Although it allows for a longer description of a specific film, it will enable all users to edit the information. Due to the edibility of the information, anyone who has access to the internet can change the content, which will later result in the inaccuracy of the movie's content. For example, one wishes to have a brief overview of Charlie and the Chocolate Factory and use Douban. He will get a detailed result based on the movie, but with the risk that others might edit the content he is reading, he is likely to read something biased and might contain incorrect information. For example, if a viewer did not like Johnny Depp, the actor in the first place, he might watch the movie more on a stricter side and focus on whether Jonny Depp acted well. If he believes the opposite, he might go onto Douban and edit the content information based on how he felt. In that case, the content that people read could be biased, and that with possible false information will not be accurate and should not be used as information for either books or movies.

For this paper, I aim to explain how the app I created using API and flutter can generate information about books and movies and provide a summary of films and books in a few clicks. Douban inspires my method, and I liked how it allows people to create a discussion group to share thoughts on books and movies. However, it does not contain information about the movie or book in most cases. In the future, I want to create an app that gathers data to provide a quick overview of either books or films, which will help people decide what to read and watch quickly. There are some excellent methods for this app. First, on the first split-screen, there are four available book recommendations, and on a separate page, there are general recommendations for movies. Second, an extensive search engine allows people to input either books or films, and within a few clicks, a list of information will be shown, such as the author, the year of publication, and the summary of a book. For movies, there will be information such as the producers, authors, and an overview of the film.

As previously stated, the app's usefulness is based on the results shown for both books and movies. The results should contain various things such as authors, summary, the producers, and more. I needed to make sure the data I was getting was accurate to find information about the overview of books and movies; I could use the app Books and Movie Search. In two application scenarios, I demonstrated how the combination of accuracy and the techniques show that this app's results are accurate and can be used. In the first case study on the evolution of the consistency of results, I randomly picked twenty movies. Flip was the movie I chose to test the app, and I entered the film into the search engine. If the app gives the same result every time I search for the same thing, it shows consistent results. I tested five times and got the same feedback from the app, including the language, producer, and overall summary of the movie.

The second thing I tested was thewed accuracy of the results I got; since I was using an API, I needed to make sure that the results I was getting were correct or matched what was on the internet the most times. I again used the movie Flipped and checked the information I got from

google results; I needed to ensure that the producer, the language, and the publish time were accurate.

Based on the two tests, I concluded that the feedback was accurate and consistent.

This research paper follows the structure of an introduction paragraph, whether I showed what the app was about and how I managed to build it out. Section two was about the challenges I faced building this app named Books and Movie Search. Section three was about how I overcame the challenged I got previously stated in section two. Section four was about the details that shoed of the experiment I did as I tested and perfected mistakes. Section five was about related works that inspired me to change possible errors and learn from similar things that already existed in the market. Lastly, Section six was about the conclusion and future things I might do to improve my app.

2. CHALLENGES

Writing a research paper is a challenge, as picking a topic to write is very hard. There are many things that I wanted to include in my research paper, and I was trying to pick out a title that provides for everything. I tried to name my research paper something like how to connect the results of an AI app better and how it benefits the community, but I thought it would be too broad. After several tries, I still found it hard to decide what topic to dive into. There are many things to consider, and it took me a long time to finally start writing about the app I created, the Books and Movie Search. This app is not only a search engine. There are ideas behind it I tried to include, and I wish to develop them in the future further.

Organizing my work was also challenging, and it was hard to recall the steps I did as I did this app, and I was trying to combine things nicely. There are server things in the app, and I was trying to decide what came first and what came after, and I found it very hard to organize the language as I decided what to do. I looked through the codes, and the previously made slides helped a little. I tried to recall the steps as I wrote the research paper, but sometimes I was clueless about how I came up with some ideas, making me choose the topic. Overall, it was a challenge because it was hard for me to remember things nicely, and I could not, for the time, recall what I wanted to do as I designed the parts of the app.

The third challenge I faced was that although there are similar methods on the market, it was hard to source and compare with different things because not everyone had written a research paper on it. As I browsed through Google Scholar, I found a lot of related works but not the same things, and It was challenging for me to learn something from similar methods that there are, in fact, very few research papers that explained what they did on the plans. As I put in a book or movie search, many things showed up, but the reports discussed the pros and cons of the search engine itself instead of developing one on themselves. I had to find others who helped me get to the development of search engines, but if I put search engines, not all search engines specialized books and movies, and it was a challenge to find sources.

3. METHODOLOGY/SOLUTION

The basic idea of the Books and Movie Search was simple; the idea behind it was to find a way to use API and gather information for books and movies. This information will help users find the essential information for them and help them choose the proper film to watch or the right book to read in a few seconds by providing the summary. Since API is an online database containing such

information, the app will process and limit the search resource to only a few things, including the author, year of publication, summary, book and movie image, and more.

This app's structure connects the database with four other pages, including the splash screen, the search screen, the book detail page, and the movie detail page. As users open the mobile app books and movie search, there will first be a welcome page that will allow users to enter their names. To get started, there is a general page with four book recommendations on the page, and on the other page, there are another four movie recommendations on the top of the screen. These four recommendations provide the most popular or classic ones, allowing users to know what might appear in the search. In addition, I picked the movies and the books with friendly front pages so that it is cheerful to see as people enter the page. There are two separate tabs on the first screen: the bottom, the movie engine, and the booking engine. The first screen is a book search engine where people can enter the book's name to get a list of information, including the author and the book's image. It took me some time to decide whether I should contain the movie poster or the idea of the book; I only wished to have the essential pieces of information to limit the reading time and help people decide whether they are interested in the content. However, I found that a lot of times, it was the poster of the movie that hooked my interests, and the same thing applied to books. Therefore, I came to the idea of adding images to the app. Although it meant more coding to do, I think it worked out nicely. Whenever I was interested in some book name, it told me that for the time, I might also be hooked on the front page of the book. In addition, images would make the app pretty and pleasant; for that, I used to love comic books, and flipping through the books and scanning through the images had always been my favorite. Also, throughout this research paper, I always emphasized the importance of using the summary. However, having images along with the text would help things work out smoothly. As I was testing the app by entering random names for the booking engine, there were sometimes interesting front pages that interested me.

On the next page is the movie search screen. There will be a list of recommendations on the top, and if people enter the movie name from the movie search engine, there will also be four recommendations on the top of the movie search. Suppose people enter the name in the movie search engine. In that case, there will also be a list of information from the screen, including the producer, the language, the year of production, and an overview of the movie.

The steps are simple, the database connects everything, and there will be a separate screen that combines the information. From the app, there are only two main things to search for; one is for movies and one for books.

The app's components were the two split screens, the intro page, and the search screens. The app was supposed to make things easier; for that, it allows people to read the essential information, and it gathers stuff in a short amount of time.

Overall, the app's structure was to connect the database with two different search engines. By gathering and limiting the search result, the app will help its users find the right movies and books, provide background information, and help them decide what books and movies interest them in a short amount of time.

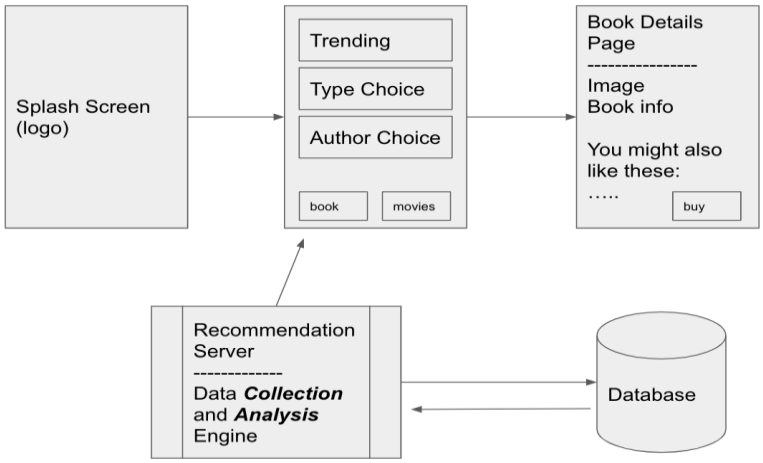


Figure 1. Books and Movies Search structure

The basic idea of the Books and Movie Search was simple; the idea behind it was to find a way to use API and gather information for books and movies. This information will help users find the essential information for them and help them choose the proper film to watch or the right book. The results given in a few seconds allow people to get the information quickly and ensure that before the users read or watch the content, they get some background information. I coded my program using Flutter in Android Studio. I used my computer, the Android Studio, and some Youtube help to create the splash screen. I used an art template to connect things with the app I started. Since the template I first used as a food delivery template, I had to change several things, including changing the images of food and calories and changing the screens from delivery and service tabs to search engines. Besides that, I tried to make sure that my app looked nice, so I drew the logo myself, as shown on the first screen. Then, I tried to make some connections between my app and the other apps I had been using, and I found out that they usually have a welcome screen and allow people to enter their names. My logo is on the top of the first welcome page, showing half of the popcorn and half of the book. The logo's meaning is to allow people to gain information about books or movies.

On the second screen, there is the book search engine. To get automatic information, I tried to research using API or the online database and include this in my app. After a few tries, I found the right one and decided to connect the API with my coding. I tried to limit the search results to only the most crucial things, including the image of the book, the author, the summary, and the language that this book is in. I wanted to use this to help the users limit their reading time and see if they are interested in the book's content.

On the other hand, I used almost the same things on different search engines, including the movie search. I tried to include the language, the producer, and the overview of the whole movie. I wanted to ensure that the users would get the most basic information, allowing them to gain some background knowledge and ensure people get the information they need in a short amount of time.

I created an app that interconnects information with API to provide information for people, allowing them to get background information about movies and books quickly.

```
var _appBar = Align(
  alignment: Alignment.centerRight,
  child: Padding(
    padding: const EdgeInsets.only(top: 37.0, left: 20.0, right: 15.0),
    child: Row(
        mainAxisAlignment: MainAxisAlignment.spaceBetween,
        children: <Widget>[
          Text(
            "Book & Movie Search",
            style: TextStyle(
                fontFamily: "Sofia",
                fontWeight: FontWeight.w800,
                fontSize: 30.0,
                letterSpacing: 1.5,
                color: Colors.white),
          ),
          Padding(
            padding: const EdgeInsets.only(top: 10.0, right: 10.0),
            child: Container(
              height: 50.0,
              width: 50.0,
              decoration: BoxDecoration(
                  image: DecorationImage(
                      image: CachedNetworkImageProvider(
                        "https://images2.imgbox.com/7d/50/GDU0vQnM_o.png",
                        errorListener: () => new Icon(Icons.error),
                      ),
                      fit: BoxFit.cover),
                  borderRadius: BorderRadius.all(Radius.circular(150.0))),
            ),
          ),
        ]),
  ),
);
```

```
Widget build(BuildContext context) {
  return Scaffold(
    body: Container(
      decoration: BoxDecoration(
          image: DecorationImage(
              image: AssetImage(
                  "assets/Template1/image/SplashScreenTemplate1.png"),
              fit: BoxFit.cover)),  // DecorationImage, BoxDecoration
      child: Center(
        child: Padding(
          padding: const EdgeInsets.only(bottom: 60.0),
          child: Row(
            mainAxisAlignment: MainAxisAlignment.center,
            crossAxisAlignment: CrossAxisAlignment.center,
            children: <Widget>[
              Image.asset(
                "assets/Template1/image/icon3.png",
                height: 45.0,
              ),  // Image.asset
              SizedBox(
                width: 10.0,
              ),  // SizedBox
              Text(
                "Movie & Book Search",
                style: TextStyle(
                    color: Colors.white,
                    fontSize: 25.0,
                    fontWeight: FontWeight.w200,
                    letterSpacing: 1,
                    fontFamily: "Poppins"),  // TextStyle
              ),  // Text
            ],  // <Widget>[]
```

```
class _bottomNavBarState extends State<bottomNavBar> {
  int currentIndex = 0;
  bool _color = true;
  Widget callPage(int current) {
    switch (current) {
      case 0:
        return new HomeScreenT1();
        break;
      case 1:
        return new BookSearch();
        break;
      case 2:
        return new MovieScreen();
        break;
      default:
        return new HomeScreenT1();
    }
  }
```

4. EXPERIMENTS/EVALUATION

In the first experiment, I tested the consistency of the result since my favorite part of the app was the images it showed; I wanted to ensure that there would be images on the results every time someone entered a name for a book or movie. To test that, I picked five films and five books. Then, I input the characters into the two different search engines to see if there will be images showing every time I search for something. I expected it to work, but it only worked for movies, not books. Then, I realized that I might have made a mistake on the app, so I went to the Android Studio and checked my codes on the book search engine. I realized I forgot to put an extra comma after the codes, resulting in a mistake. Therefore, I added another comma after the code. However, immediately after changing the codes, when I tried to reenter the thing again, the images were still not showing. I was very nervous about it and tried to ensure that nothing was wrong with the other codes. After making sure that all the other codes were correct, I thought there might be other bad things with my program. I restarted the app several times to ensure that the app worked properly, and after the third trial, images were showing again when I input the names for books.

This experiment shows that the results I am getting are consistent, and after the change in coding, the images are now displaying correctly.

The second experiment I used was to ensure that the app would not crash with many searches. I wanted to ensure that the app would work typically even after a lot of inquiries so that the app would work properly. I would not want the app to crash when the number of users increases; it will still be unknown how many people are using the app simultaneously. Therefore, I have to come up with something that works, and I need to test the effect with the app before more people use it. I want to ensure that things work correctly and that people do not need to worry about the number of users that use the app simultaneously and are worried about crashing. I wanted to ensure that the app would still work correctly even after a more significant number of searches. Therefore, I wanted to test the app and see whether the app would crash or function properly after a large number of searches. To test the app's efficiency, I decided to try twenty movies and twenty books simultaneously. To do that, I first selected the books and the movies, and I started to input the names of books and films. The number of books and movies I wanted to do was twenty, but later I thought twenty was not a more significant number, so I decided to ask more participants to join this experiment. Later, I asked my brother and my mother to enter the names of books and movies, and at the same time, after 60 searches, twenty from me, twenty from my mother, and twenty from my brother, the app still works properly. I thought that it was good to conclude that the app will be able to work correctly with an increasing number of searches. However, sixty was not a very large number, so I decided to make my grandparents and my dad test it as well. Again, after one hundred and twenty searches, the app still works properly, so I think it means that the experiment I wanted to test was on the app's effect.

I faced challenges as I tried to do the experiments; first of all, it was hard to decide which ones to test since there are multiple things the search engines will provide, and it is hard to pick a specific one for me to test. I try the images because it is always great to give some graphics in the app to make it more interesting. I wanted to create an app that others could love, and I did not want this feature to crash. So I tested it several times, surprising that the images crashed for the book search and that I needed to change the codes and ensure that it worked.

The other experiment that I did was to test whether the app would crash or not. I think it was a little challenging for the number of participants I needed to gather. I needed to ensure that the app would not crash even if multiple people were using the app at the same time. However, it was hard for my grandparents to download the app, and since we did not live together, it was hard for

me to explain things to them. It took me a long time to contact my grandparents to explain why iI needed help from them. I needed to end up making a phone call with them and explain why it was essential to have multiple people test it.

Other than that, it was nice to see that things went as planned and that it was pleasant to see the app getting done with little things left to fix.

5. RELATED WORK

One method that exists on the market is collaborative filtering, which will help the system access people's preferences on books and movies. [1] The difference between using an API and collaborative filtering is that the API does not gather personal information and recommend books and movies to the users. It is more like a machine that gives feedback when entering an input, and the output will always stay the same(with the categories or the information the app promises to provide, such as authors, summary, and producers).

Another related method on the market regarding movie search is that people use GPS to help the users connect the background information of movies and to help them locate the nearby theatres. The main difference between the movie and book search app and The Smart Movie Recommendation[2] is that my app mainly focuses on the information I will be providing. The Smart Movie Recommendation focuses more on how to help people find the movie they like and to find theatres that connect them to the theatres.

The last similar method with my app, the Movie and Book Search, is a Netflix Rest API that gathers the users' watching habits based on the last sixty hours of the content. And based on that, the API will automatically collect information that will likely be on a similar topic from the previously watched content. The difference between the Movie and Book SErach app is that my app does not collect personal data from the users and, therefore, cannot predict what people might like or not. It only shows information to provide some background information on books and movies.

6. CONCLUSION AND FUTURE WORK

The idea behind my work was simple: I wanted to find a way to minimize people's time deciding what books and movies to read or watch; for that, choosing was very hard for me. For the app Books and Movie Search, I coded with Flutter using Android Studio. I used an online API, or database, to gather information for various information for books and movies, such as authors, producers, and the summary. Once I was at the theatre and interested in the poster and the movie name, I bought the ticket and went to watch that movie. Later on, I realized the film was filled with dirty jokes and politician sacraments, and I noticed I had chosen the wrong movie. It was possible for me to google this movie and read some background knowledge of it, and at that second, something clicked in my head. What if I create an app that gathers information for both books and movies that allows people to have background knowledge of what is going on. If one wishes to watch a film and is interested in the movie name, he can use my app and enter the movie's title. Within seconds, he will be able to get a summary of the film, and instead of watching the whole film, he limits the time to a few minutes of reading time. Based on the information provided, it gives users some background idea about what the film will be about and that the app can hopefully help decide whether or not that person is interested in the content. With the help of the app, the results within the API that I included in the app will be out in a few seconds.

It is straightforward to test whether the app works or not; say I wanted to know more about the movie Flipped, I can open the app from a mobile device and select the movie tab. Then, all I need to do is to enter the movie name, in this case, Flipped, into the search engine. After a few seconds, a long list of things will appear on the screen, including the author, publication date, and a summary of this book. This app solves time-consuming deciding on what movies or books to buy that with some background knowledge of the book, it would be easy to determine whether the content is interesting or not.

Due to the fact that this app's data is gathered by an API, it could be some accuracy problem that I cannot change what information would be shown. Also, since all items that I have tested are books and movies that already exist, there might be future factors that I cannot control from the generator, and there might be a possible delay in updating new information to the app.

For the app's future development, I would like to add recommendation buttons for books and movies. The app's current version allows people to know about books and films, and I wanted to create tabs that include categories of films and books, such as trending, romance, and horror. I think that will make my app a lot neater and that If I were the user, I would love to see that there are different categories of things instead of plain search engines. If I have the chance, I would also want to add some motivational quotes, of course, either from movies or books that will update daily as people open the app.

REFERENCES

[1] KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data miningAugust 2007 Pages 550–559https://doi.org/10.1145/1281192.1281252

[2] Ko, SK. et al. (2011). A Smart Movie Recommendation System. In: Smith, M.J., Salvendy, G. (eds) Human Interface and the Management of Information. Interacting with Information. Human Interface 2011. Lecture Notes in Computer Science, vol 6771. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21793-7_63

[3] Berry, S., Fazzio, S., Zhou, Y., Scott, B. and Francisco-Revilla, L. (2010), Netflix recommendations for groups. Proc. Am. Soc. Info. Sci. Tech., 47: 1-3. https://doi.org/10.1002/meet.14504701402

A SUMMARY OF COVID-19 DATASETS

Syed Raza Bashir[1] , Shaina Raza[2], Vidhi Thakkar[3] and Usman Naseem[4]

[1]Department of Computer Science,
Toronto Metropolitan University, Toronto, Canada
[2]University of Toronto, Canada
[3]Institute of Aging, Faculty of Nursing, University of Victoria, Victoria, Canada
[4]Department of Information Technology, Sydney International School of Technology and Commerce, Sydney, Australia

ABSTRACT

This research presents a review of main datasets that are developed for COVID-19 research. We hope this collection will continue to bring together members of the computing community, biomedical experts, and policymakers in the pursuit of effective COVID-19 treatments and management policies. Many organizations, such as the World Health Organization (WHO), John Hopkins, National Institute of Health (NIH), COVID-19 open science table and such, in the world, have made numerous datasets available to the public. However, these datasets originate from a variety of different sources and initiatives. The purpose of this research is to summarize the open COVID-19 datasets to make them more accessible to the research community for health systems design and analysis. We also discuss the numerous resources introduced to support text mining applications throughout the COVID-19 literature; more precisely, we discuss the corpora, modelling resources, systems, and shared tasks introduced for COVID-19.

KEYWORDS

COVID-19, Text Mining, Public health, Risk, Public Health, COVID-19 Data; Data Science.

1. INTRODUCTION

COVID-19, a SARS coronavirus 2-related acute respiratory infectious disease, was discovered in late 2019 and declared a global pandemic by the WHO in March 2020 (4). Recent estimates show that there have been over 500,000 deaths and 400 million jobs lost due to the COVID-19 pandemic (5), causing widespread socioeconomic, political, and health policy implications as well as increases in physical and mental health conditions (6). Many empirical and theoretical studies have looked into COVID-19's morbidity and mortality rates since its discovery (7). The effects of COVID-19 have lasted for over a year since it was first reported. Many people's mental health has been severely impacted by the COVID-19 epidemic. The long-COVID effects are also being studied in the latest research (8). To summarize, the amount of information on the epidemic that would define a generation is becoming nearly overwhelming.

A few projects, such as CoronaNet [1]; CORD-19 [2] and a United States (US) county-level dataset [3], among a few, that presents a collection of data from different sources. However, the nature of these datasets are different, for example CoronaNet [1] and county-level dataset [3] are repositories of government responses, while the CORD-19 [2] is a collection of academic literature. Nonetheless, our approach to this work is unique in that we also focus on multiple data

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 231-242, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121219

sources and data types, such as epidemiology, literature, government policies, lab's data, mortality and other miscellaneous data. We believe that having a summary of COVID-19 datasets is a necessary first step in the innovation and research landscape for understanding the COVID-19 pandemic and mitigating its effects.

To deal with the overwhelming amount of COVID-19 literature, the computer community has developed text mining corpora, modelling resources, systems, and community-wide shared tasks. Text mining uses corpora, which are collections of documents pre-processed to extract machine-readable text. Corpora are collections of documents that have been pre-processed to extract machine-readable text and are used for text mining; in this example, the corpora contain scientific publications. Text mining practitioners can include modelling resources into production systems. These resources include text embeddings, data annotations, pretrained language models, and knowledge graphs.

In this article, we discuss some of the most common types of COVID-19 data sources available. We also outline prominent text mining systems that have been developed/implemented to assist text mining in the COVID-19 literature. Our goal is to make it easier for researchers, scientists, and other interested parties to get most up-to-date data for their research. We believe that a list of the most relevant sources of COVID-19 data is needed for the purpose of health systems design and health services research. It is our hope that this resource will continue to bring together the computing community, biomedical experts, and policy makers in the pursuit of effective treatments and management policies for Covid-19.

2. COVID-19 DATASETS

We categorize the types of COVD-19 related into different types as per the type of research being conducted. Each of this dataset below is supported by a reference (a reference for dataset, a link to the dataset page and the main provider/source of dataset).

2.1. Survey Data

Survey data is defined as the outcome data collected from a sample of survey respondents [4]. This data is extensive information gathered from a specific target audience about a specific topic to conduct research. We present some of the survey data related to COVID-19, in Table 1, to help researchers better understand the influences on participants' daily changes, notably the impact of COVID-19 impacts.

Table 1. Survey data (each underline data source is hyperlink)

Data	Source
- Canadian Perspectives Survey Series (CPSS) [5] - Collection time: January 15, 2020 to March 15, 2021 - The CPSS entails assembling a group of people who agree to take a series of brief online surveys related to COVID-19 over the course of a year. - It consists of following surveys - COVID-19 in Canada: An update on social and economic impacts, Fall 2021 [6] - Impacts of COVID-19 [7] - COVID-19 and working from home, 2020 [8] - Impacts of COVID-19 on job security and personal finances, 2020 [9] - Canadians report lower self-perceived mental health during the COVID-19 pandemic [10] - Mental health of Canadians during the COVID-19 pandemic [11]	Statistic Canada [19]

- Canadian Perspectives Survey Series 2: Monitoring the effects of COVID-19, May 2020 [12] - Food insecurity before and during the COVID-19 pandemic [13] - Food insecurity during the COVID-19 pandemic, May 2020 [13] - Canadian Perspectives Survey Series 3: Resuming economic and social activities during COVID-19 [14] - Precautions that Canadians will take or continue to take as COVID-19 safety measures are relaxed [15] - Canadian Perspectives Survey Series 4: Information sources consulted during the pandemic, July 2020 [16] - Canadians spend more money and time online during pandemic and over two-fifths report a cyber incident [17] - COVID-19 in Canada: A Six-month Update on Social and Economic Impacts [18]	
Canadian COVID-19 Antibody and Health Survey (CCAHS) Collection time: From November 2, 2020, to April 16, 2021 This survey gathers data in two parts. The first section is an automated questionnaire concerning general health and COVID-19 exposure. The second component is an at-home finger-prick blood test, which is sent to a lab to determine the existence of COVID-19 antibodies.	Statistic Canada
Pulse Survey on COVID-19 and its Impacts on Statistics Canada Employees (PSCISCE) [20] Collection time: April 27, 2021 The goal of this poll is to analyze the workforce's health in near real time, give immediate information on the COVID-19 crisis's impact, and shape the development of tools and action plans to better support our employees going ahead.	Statistic Canada
Uniform Crime Reporting Survey – Selected police reported crime statistics – Special COVID-19 report to Statistics Canada [21] Collection time: January 15 to December 31, 2022 The purpose of this survey is to provide timely monthly aggregate data on the types of police-reported activities, including criminal occurrences and other requests for police assistance, that happened during the COVID-19 pandemic's initial months. The data can be used by policymakers, researchers, and the general public.	Statistic Canada

2.2. Epidemiology Data

Epidemiology is the study of how often diseases occur in different groups of people and why they occur [22]. When developing and evaluating disease prevention strategies, epidemiological information is used as a guide for the management of patients who have already developed the disease. Similarly, to the clinical findings and pathology of a disease, the epidemiology of a disease is an integral part of the disease's fundamental description. The subject has its unique techniques for data collection and interpretation, that must be understood.

Table 2. Epidemiology data

Data	Source
COVID-19 daily epidemiology update [23] This data consists of COVID-19 case summaries from across Canada and over time. it also presents a summary of hospitalizations and deaths, testing, variants of concern, and exposures is provided.	Canada.ca [24] , Canada COVID-19 hub [25]
WHO daily data [26] This dashboard presents official daily counts of COVID-19 cases and deaths worldwide.	World Health Organization (WHO) [27]
COVID-19 Canada Open Data Working Group [28], [29] This is an epidemiological data from the COVID-19 Epidemic in Canada.	Open COVID [30]
Covid-19 Cases in Toronto [31] This dashboard gives information about case counts, epidemiological summary of cases, active outbreaks and vaccine data in Toronto.	City of Toronto [32]

2.3. COVID-19 Dataset by Source

We also gathered some data sources that provide COVID-19 statistics by country, province and city. The goal is that one may use the original data sources from the government websites, rather than relying on different other sources. In the rest of this review article, we present the main source of data, more details about each dataset can be found in the respective paper or source.

Table 3. Data sources by country and states.

Data	Source
COVID-19 cases and forecasting hot spots, country wise [33]	Mayo Clinic [34]
COVID Data Tracker [35]	Centers for Disease Control and Prevention [36]
NYTimes Latest Map and Case Count [37]	New York Times [38]
COVID-19 Cases by Country [39]	Johns Hopkins [40]
COVID Risk & Vaccine Tracker [41]	COVID ActNow [41]
Global COVID-19 Tracker [42]	Kaiser Family Foundation [43]
COVID Tracking [44]	COVID Tracking Project[45]
ACLED Data Export Tool [46]	Armed Conflict Location & Event Data Project [47]
COVID-19 cases daily [48]	European Centre for Disease Prevention and Control [48]
60 million anonymized COVID-19 case [49]	Global Health [49]

2.4. COVID-19 Projections Data

The COVID-19 projection models work with a big dataset. When there are larger datasets, there exists the possibility to study statistical changes in the epidemiology, pathogenesis, and spread of the COVID-19 pandemic within a country and across provinces. This type of information is valuable to policymakers in creating predictive models and in order to inform public policy restrictions on wearing masks, maintaining physical distance, and advising on how the economy should function, including rules around the conduct of small businesses and restaurants.

COVID-19 projects data can inform policy changes, help to measure the spread of the pandemic amongst population groups and inform the calculation of mortality rates. Although historical data is rarely a perfect predictor of the future, when used correctly, it can provide some insight into what the coming days and weeks may look like. It is also possible to do predictive modelling

with Tableau and SPSS software that can help predictive logarithmic models to forecast the anticipated increase or decrease in cases over a time period.

Table 4. Data sources by country and states.

Data	Source
COVID-19 projections bi-weekly [50]	Institute for Health Metrics and Evaluation [51]
COVID-19 Staffing Simulator [52]	LEVRUM [53]
State-level social distancing policies [54]	University of Washington [55]
Alcohol Sales During the COVID-19 Pandemic [56]	Alcohol Policy for Information Systems [57]
COVID-19 prison data [58]	Marshall Project [59]

2.5. Lab Data

The number of COVID-19 positive tests tell us only so much about how the pandemic is behaving in each community. This is due, in large part, to testing bias, in which tests are limited to the sickest and most stereotypically presenting patients. To fully comprehend how deeply embedded COVID-19 is in a given community, the total volume of tests must be reported alongside the ever-present positive case count.

Table 5. COVID-19 Lab Datasets

Data	Source
Virology COVID-19 Dashboard [60]	University of Washington Medicine's Department of Laboratory Medicine [55]
Online access to COVID-19 Vaccination Information and COVID-19 lab test results [61]	eHealth Ontario [62]
Vaccine Adverse Event Reporting System [63]	Vaccine Adverse Event Reporting System [64]
Public Health Ontario Data and Analysis [65]	Public Health Ontario [66]
BCCDC COVID-19 [67]	British Columbia Centre for Disease Control [68]

2.6. Mobility Data

Mobility data can also be used to determine the extent to which communities have been shut down. There are six types of areas that Google looks at in its dashboard, which may be broken down by country or region. These include retail and recreation; grocers and pharmacy; parks; transit stations; businesses; and residential sites.

Table 6. COVID-19 Travel and Mobility Data

Data	Source
COVID-19 Mobility data [69]	Google [70]
COVID-19 Reports [71]	Institute for Disease Modeling [72]
Facebook Data for Good [73]	Meta [74]

2.7. Mortality Data

COVID-19 mortality is defined as those people who died because of complications from COVID-19 as well as with other ailments because of their incapacity or reluctance to access a possibly overwhelmed health system, or because they did not have the right treatment in the system at the

time, is also an important data. Plaxovid and retonovir, for example, were unavailable for a long time and resulted in numerous deaths.

Table 7. Mortality Data

Data	Source
US COVID-19 vs other causes of deaths [75]	Flourish [76]
Canada COVID-19 mortality data [77]	Worldometers [78]

2.8. Public and Mental Health Datasets

We also gathered public health and mental health COVID-19 data from miscellaneous sources, which are given below.

Table 8. Miscellaneous sources of Data

Data	Source
Michigan COVID-19 Data [79]	Michigan Government [80]
COVID-19 Public datasets [81]	Google [70]
Unintended consequences of COVID-19: Impact on harms caused by substance use [82]	Canadian Institute for Health Information
Unintended consequences of COVID-19: Impact on self-harm behavior [83]	Canadian Institute for Health Information
Ontario Mental Health Reporting System Metadata [84]	Canadian Institute for Health Information
Mental Health During COVID-19 Outbreak [85]	Centre for Addiction and Mental Health [86]
Impact of COVID-19 on accidental falls [87]	Canadian Institute for Health Information
Flatten: COVID-19 Survey Data on Symptoms, Demographics and Mental Health in Canada [88]	PhysioNet [89]
COVID-19 National Survey [90]	Centre for Addiction and Mental Health [86]
Provincial COVID-19 Vaccine (COVaxON) [91]	Ontario Health Data Platform [92]
COMPASS CIHR mental health [93]	COMPASS SYSTEM [94]

3. TEXT MINING DATASETS AND TOOLS

Text mining, also known as text data mining or text analytics, is a method for extracting valuable information from text [95]. As a result of the COVID-19 pandemic, several datasets containing full text about COVID-19, SARS-CoV-2, and related coronaviruses have been released. These freely available datasets are being made available to the global research community to enable the application of recent advances in natural language processing and other artificial intelligence (AI) techniques in order to generate new insights to aid in the ongoing fight against this infectious disease.

3.1. Text Mining Datasets

The COVID-19 Open Research Dataset [2] is one of the earliest and largest literature corpora created to support COVID-19 text mining. It is a corpus of metadata and full text of COVID-19 publications and preprints released daily by Semantic Scholar at the Allen Institute for AI in collaboration with Microsoft Research, IBM Research, Kaggle, the Chan-Zuckerberg Initiative,

the National. This corpus was first made available on March 16, 2020, at the request of the White House Office of Science and Technology Policy, in order to support community-wide efforts to apply text mining techniques to coronavirus literature.

The corpus includes papers from PubMed Central (PMC), PubMed, the World Health Organization's COVID-19 database and preprint servers bioRxiv, medRxiv, and arXiv. Paper metadata from these sources is synchronised, PDFs are converted to machine-readable JSON using the S2ORC pipeline described in [96]. As of September 15, 2020, the corpus contained over 260 000 paper entries (with 105 000 full text entries).

LitCovid is a curated set of open access COVID-19 papers from PubMed, currently containing more than 240,000 papers and growing. LitCovid is focused on tracking publications specific to COVID-19, while CORD-19 captures the coronavirus literature more broadly, including other coronaviruses (e.g. SARS and MERS) and a wider time period (i.e. before the current outbreak). LitCovid does NOT include pre-prints. LitCovid only includes relevant articles from PubMed.

The other well-known COVID-19 literature repositories are World Health Organization (WHO) COVID-19 database [97] and the Centers for Disease Control and Prevention's COVID-19 research articles database [98]. These databases overlap with other corpora; for example, the WHO database is ingested by CORD-19, and much of the CDC database overlaps with PubMed and PMC, both of which are sources of papers in CORD-19 and LitCovid. The CDC database also includes a collection of white papers and technical reports. Finally, several publishers have compiled and released collections of their COVID-19 literature, such as Elsevier's Novel Coronavirus Information Center [99], Springer Nature’s Coronavirus Research Highlights [100], and other publishers who provide literature under temporary open access licenses. We summarize these text mining datasets related to COVID-19 in Table 9.

Text

Table 9. Text Mining Data

Data	**Source**
CORD-19: COVID-19 Open Research Dataset [2]	Allen Institute for AI [101]
Research COVID-19 with AVOBMAT [102]	AVOBMAT [103]
LitCovid [104]	NLM/NCBI BioNLP Research Group [105]
World Health Organization (WHO) COVID-19 database [97]	WHO [27]
Centers for Disease Control and Prevention's COVID-19 research articles database [98]	CDC [36]
Elsevier's Novel Coronavirus Information Center [99],	Elsevier Publisher
Springer Nature’s Coronavirus Research Highlights [100]	Springer Nature Publisher

3.2. Text Mining Tools

Many text mining systems for COVID-19 literature have been released. We compile a list of important tools in Table 10.

Table 10. Text Mining Tools

Tool	Data source
Covidex [106]	CORD-19, ClinicalTrials.gov
KDCovid [107]	CORD-19
Azure Cognitive Search [108]	CORD-19
CADTH [109]	Multiple sources

4. CONCLUSIONS

Since the COVID-19 emergence, research institutes and governments have made numerous databases publicly available to allow research groups (and independent individuals) to analyse data relating to the COVID-19's spread [110]. These databases are dispersed across a variety of initiatives and sources. The purpose of this article is to compile a list of all of the world's major open databases and data initiatives. Our goal is to provide a road map for establishing links between various scientific fields (health science, epidemiology, artificial intelligence, and mental health). We provide most up-to-date information for a wide range of COVID-19 datasets, which are critical for knowledge synthesis in evidence-based medicine. This review of COVID-19 research can benefit researchers, healthcare professionals, and the general public. Each of these sources is open source and can be downloaded and used in a variety of computational tools and systems.

While this paper includes a list of the most commonly used COVID-19 datasets and tools, it may not include many other COVID-19 datasets. However, because we mainly curated the most important and up-to-date COVID-19 data sources, we believe this list is representative of other COVID-19 datasets as well. We put forward this research direction and give researchers the opportunity to look for more data sources and collaborate to fight the pandemic.

ACKNOWLEDGEMENTS

We would like to acknowledge that this research and manuscript is a part of my CIHR Health Systems Impact Fellowship.

REFERENCES

[1] C. Cheng, J. Barceló, A. S. Hartnett, R. Kubinec, and L. Messerschmidt, "COVID-19 Government Response Event Dataset (CoronaNet v.1.0)," *Nat. Hum. Behav.*, vol. 4, no. 7, pp. 756–768, 2020.
[2] L. Lu Wang *et al.*, "CORD-19: The Covid-19 Open Research Dataset.," 2020.
[3] B. D. Killeen *et al.*, "A County-level Dataset for Informing the United States' Response to COVID-19," 2020.
[4] S. G. Heeringa, B. T. West, and P. A. Berglund, *Applied survey data analysis*. chapman and hall/CRC, 2017.
[5] Statistics Canada, "Canadian Perspectives Survey Series, 2020," 2020. [Online]. Available: https://www.statcan.gc.ca/en/survey/household/5311. [Accessed: 23-Jan-2022].
[6] Statistics Canada, "COVID-19 in Canada: A One-year Update on Social and Economic Impacts," no. March, pp. 1–52, 2021.
[7] Statistics Canada, "Canadian Perspectives Survey Series 1 : Impacts of COVID-19," *Stat. Canada*, pp. 9–14, 2020.
[8] T. Daily, "Working from home during the COVID-19 pandemic Working from home: a new experiment for many Canadian workers and employers," no. April 2020, 2021.
[9] Statistics Canada, "Canadian Perspectives Survey Series 1: personal finances, 2020," *Dly.*, vol. 11-001–X, pp. 17–20, 2020.
[10] H. Gilmour, "COVID-19: Data to Insights for a Better Canada Canadians report lower self-perceived

mental health during the COVID-19 pandemic," *Stat. Canada*, no. 45280001, 2020.
[11] S. Canada, "Statistics:Covid :Mental Health," no. May, p. 2020, 2020.
[12] Statistics Canada, "Canadian Perspectives Survey Series 2: Monitoring the effects of COVID-19," *The Daily*, 2020.
[13] Statistics Canada, "Food insecurity before and during the COVID-19 pandemic , 2017 / 2018 and May 2020," no. May, p. 8300, 2020.
[14] Statistics Canada, "Canadian Perspectives Survey Series 3: Resuming economic and social activities during COVID-19," *Dly.*, 2020.
[15] S. Canada, "PRECAUTIONS THAT CANADIANS WILL TAKE OR CONTINUE TO TAKE AS COVID-19 SAFETY MEASURES ARE RELAXED," p. 35446, 2020.
[16] Statistics Canada, "Canadian Perspectives Survey Series 4: pandemic, July 2020," *Dly.*, vol. 11-001–X, no. July, pp. 17–20, 2020.
[17] T. Daily, "Canadians spend more money and time online during pandemic and over two-fifths report a cyber incident," pp. 40–43, 2020.
[18] A. Arora, "COVID-19 in Canada: A Six-month Update on Social and Economic Impacts," no. September, pp. 1–36, 2020.
[19] Government of Canada, "Statistics Canada: Canada's national statistical agency," 2016. [Online]. Available: http://www.statcan.gc.ca/pub/62f0026m/2016001/chap4-eng.htm. [Accessed: 23-Jan-2022].
[20] Statistics Canada, "Pulse Survey on COVID-19 and its Impacts on Statistics Canada Employees (PSCISCE)." [Online]. Available: https://www.statcan.gc.ca/en/survey/household/5326. [Accessed: 23-Jan-2022].
[21] Statistics Canada, "Uniform Crime Reporting Survey - Selected police reported crime statistics - Special COVID-19 report to Statistics Canada." [Online]. Available: https://www.statcan.gc.ca/en/survey/business/3302. [Accessed: 23-Jan-2022].
[22] The British Medical Journal, "Epidemiology for the uninitiated," *BMJ : British Medical Journal*, 2018. [Online]. Available: https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/6-ecological-studies?hwshib2=authn%3A1535394580%3A20180826%253Ae48fec14-6539-4bab-9bd2-b60a7372a9d1%3A0%3A0%3A0%3ANhpf%2BmFJqPMhlDoGoRukJw%3D%3D. [Accessed: 19-Jan-2022].
[23] Public Health Agency of Canada, "COVID-19 Daily Epidemiology Update," *Coronavirus disease (COVID-19): For health professionals*, 2021. [Online]. Available: https://www.canada.ca/content/dam/phac-aspc/documents/services/diseases/2019-novel-coronavirus-infection/surv-covid19-epi-update-eng.pdf.
[24] Canada.ca, "Home - Canada.ca," 2021. [Online]. Available: https://www.canada.ca/en.html. [Accessed: 23-Jan-2022].
[25] Public Health Agency of Canada, "COVID-19 Daily Epidemiology Update," *Coronavirus disease (COVID-19): For health professionals*, 2021. [Online]. Available: https://www.canada.ca/content/dam/phac-aspc/documents/services/diseases/2019-novel-coronavirus-infection/surv-covid19-epi-update-eng.pdf. [Accessed: 23-Jan-2022].
[26] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 533–534, 2020.
[27] A. EGDAHL, "WHO: World Health Organization.," *The Illinois medical journal*, 1954. [Online]. Available: https://www.who.int/. [Accessed: 23-Jan-2022].
[28] I. Berry, J. P. R. Soucy, A. Tuite, and D. Fisman, "Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canad," *CMAJ*, vol. 192, no. 15, p. E420, Apr. 2020.
[29] I. Berry *et al.*, "A sub-national real-time epidemiological and vaccination database for the COVID-19 pandemic in Canada," *Sci. Data*, vol. 8, no. 1, Dec. 2021.
[30] Opencovid, "COVID-19 Canada Open Data Working Group." [Online]. Available: https://opencovid.ca/. [Accessed: 23-Jan-2022].
[31] T. Publich Health, "COVID-19 Cases in Toronto - City of Toronto Open Data Portal," 2020. [Online]. Available: https://open.toronto.ca/dataset/covid-19-cases-in-toronto/. [Accessed: 23-Jan-2022].
[32] N. Services, "City of Toronto," *Policy Analysis*, 2004. [Online]. Available: https://www.toronto.ca/. [Accessed: 23-Jan-2022].

[33] Mayo Clinic, "U.S. COVID-19 Map: Tracking the Trends." [Online]. Available: https://www.mayoclinic.org/coronavirus-covid-19/map. [Accessed: 27-Jan-2022].
[34] Mayo Clinic, "Mayo Clinic," *Mayo Clinic*, 2019. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/hepatocellular-carcinoma/cdc-20354552. [Accessed: 27-Jan-2022].
[35] CDC, "CDC COVID Data Tracker," *Centers for Disease Control and Prevention*, 2020. [Online]. Available: https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-total-admin-rate-total%0Ahttps://covid.cdc.gov/covid-data-tracker/#vaccinations-pregnant-women%0Ahttps://covid.cdc.gov/covid-data-tracker/#datatracker-home%0Ahttps://covid.cdc.gov/covid-data-trac. [Accessed: 27-Jan-2022].
[36] C. for D. C. and Prevention., "Centers for disease control and prevention," *Indian Journal of Pharmacology*, 2004. [Online]. Available: https://www.cdc.gov/. [Accessed: 27-Jan-2022].
[37] T. N. Y. Times, "Covid in the U.S.: Latest Map and Case Count," *The New York Times*, 2020. [Online]. Available: https://www.nytimes.com/interactive/2021/us/covid-cases.html. [Accessed: 27-Jan-2022].
[38] A. Cowell, "The New York Times - Breaking News, US News, World News and Videos," *The New York Times*, 2003. [Online]. Available: https://www.nytimes.com/. [Accessed: 27-Jan-2022].
[39] Johns Hopkins University, " COVID-19 United States Cases by County Johns Hopkins University," *Https://Coronavirus.Jhu.Edu/Us-Map*, 2021. [Online]. Available: https://coronavirus.jhu.edu/us-map. [Accessed: 27-Jan-2022].
[40] Johns Hopkins University, "Mortality Analyses - Johns Hopkins Coronavirus Resource Center," *Johns Hopkins Coronavirus Resource Center*, 2020. [Online]. Available: https://coronavirus.jhu.edu/data/mortality. [Accessed: 27-Jan-2022].
[41] C. A. Now, "Realtime U.S. COVID Map & Vaccine Tracker - Covid Act Now." [Online]. Available: https://covidactnow.org/?s=22401166. [Accessed: 27-Jan-2022].
[42] G. C.-19 Tracker, "Global COVID-19 Tracker – Updated as of January 27 | KFF." [Online]. Available: https://www.kff.org/coronavirus-covid-19/issue-brief/global-covid-19-tracker/. [Accessed: 27-Jan-2022].
[43] Kaiser Family Foundation, "KFF - Health Policy Analysis, Polling, Journalism and Social Impact Media." [Online]. Available: https://www.kff.org/. [Accessed: 27-Jan-2022].
[44] CovidTracking.com, "The COVID Tracking Project | The COVID Tracking Project," *The Atlantic Monthly Group*, 2020. [Online]. Available: https://covidtracking.com/. [Accessed: 27-Jan-2022].
[45] The Atlantic, "The Covid Tracking Project API," *The Covid Tracking Project*, 2020. [Online]. Available: https://covidtracking.com/api. [Accessed: 27-Jan-2022].
[46] ACLED, "Data Export Tool | ACLED," 2021. [Online]. Available: https://acleddata.com/data-export-tool/.
[47] ACLED, "Home - ACLED," 2021. [Online]. Available: https://acleddata.com/#/dashboard.
[48] R. Evans, "European Centre for Disease Prevention and Control," *Nursing standard (Royal College of Nursing (Great Britain) : 1987)*, 2014. [Online]. Available: https://www.ecdc.europa.eu/en. [Accessed: 27-Jan-2022].
[49] Global.health, "Global.health: A Data Science Initiative - Global.health," 2021.
[50] IHME, "COVID-19." 2022.
[51] M. Arthur, "Institute for Health Metrics and Evaluation," *Nursing Standard*, vol. 28, no. 42. pp. 32–32, 2014.
[52] LEVRUM, "Levrum COVID-19 Simulator," 2021. [Online]. Available: https://covidsim.levrum.com/.
[53] LEVRUM, "Home - Levrum Data Technologies." [Online]. Available: https://levrum.com/.
[54] C. Wu, F. Wu, Y. Chen, S. Wu, Z. Yuan, and Y. Huang, "Neural Metaphor Detecting with CNN-LSTM Model," 2018, pp. 110–114.
[55] University of Washington, "UW Homepage." 2016.
[56] APIS, "About Alcohol Policy | APIS - Alcohol Policy Information System." 2021.
[57] National Institute on Alcohol Abuse and Alcoholism, "Alcohol Policy Information System (APIS)," vol. 2008, no. December 19. 2005.
[58] K. Park and T. Meagher, "A State-By-State Look at 15 Months of Coronavirus in Prisons | The Marshall Project," *The Marshall Project*. 2021.
[59] Marshall Project, "The Marshall Project." 2021.
[60] U. Medicine, "UW Virology COVID-19 Dashboard." 2020.

[61] eHealth Ontario-Ministry of Health, “Online access to COVID-19 lab test results for Health Care Providers | eHealth Ontario | It’s Working For You.” .
[62] eHealth Ontario, “eHealth Ontario | It’s Working For You.” 2017.
[63] “VAERS - Data,” 2021. [Online]. Available: https://vaers.hhs.gov/data.html.
[64] VAERS, “The Vaccine Adverse Event Reporting System (VAERS),” *Vaccine*, vol. 12, no. 10. p. 960, 1994.
[65] Public Health Ontario, “Coronavirus Disease 2019 (COVID-19) – PCR | Public Health Ontario,” *Public Health Ontario*. 2020.
[66] K. Wong and A. Piatkowski, “Public Health Ontario.” pp. 1–30, 2016.
[67] BCCDC, “COVID-19 – BCCDC Foundation for Public Health.” 2022.
[68] Infections, “BC Centre for Disease Control.” 2009.
[69] Google, “COVID-19 Community Mobility Reports.” 2022.
[70] Google, “Google,” 2022. [Online]. Available: https://www.google.com/.
[71] IDMOD COVID-19, “COVID Reports | IDMOD.” 2022.
[72] Institute of Disease Modeling, “Home Page | IDMOD.” 2022.
[73] J. M. Wing, “Data for Good,” 2018. [Online]. Available: https://dataforgood.facebook.com/.
[74] Meta, “Welcome to Meta | Meta,” *Meta*. p. 13, 2021.
[75] Flourish, “Covid vs. US Daily Average Cause of Death | Flourish.” 2022.
[76] Flourish, “Flourish | Data Visualisation \& Storytelling.” 2021.
[77] Worldometer, “Canada COVID - Coronavirus Statistics - Worldometer.” .
[78] Worldometers, “Worldometers: Real Time World Statistics,” *Choice Reviews Online*, 2012. [Online]. Available: https://www.worldometers.info/.
[79] State of Michigan, “Coronavirus: Michigan Data,” *Michigan.gov*. 2020.
[80] “State of Michigan.” 2022.
[81] Google, “Google Cloud Platform,” *Https://Cloud.Google.Com*. pp. 11–12, 2016.
[82] Canadian Institute for Health Information, “Unintended consequences of COVID-19: Impact on harms caused by substance use.” pp. 1–20, 2021.
[83] “Unintended Consequences of COVID-19: Impact on Harms Caused by Substance Use,” 2021.
[84] Canadian Institute for Health Information, “Ontario Mental Health Reporting System,” *Data Quality Documentation*. 2016.
[85] CAMH, “Mental Health and the COVID-19 Pandemic | CAMH.” 2021.
[86] J. C. Negrete, J. Collins, N. E. Turner, and W. Skinner, “The Centre for Addiction and Mental Health,” vol. 49, no. 12. 2004.
[87] CIHI, “Impact of COVID-19 on accidental falls in Canada | CIHI.” 2021.
[88] Flatten, “Flatten: COVID-19 Survey Data on Symptoms, Demographics and Mental Health in Canada v1.0.” 2021.
[89] PhysioNet, “PhysioNet,” 2021. [Online]. Available: https://www.physionet.org/.
[90] The Centre for Addiction and Mental Health, “COVID-19 National Survey Dashboard.” 2020.
[91] P. D. Yadav *et al.*, “Neutralization of Beta and Delta variant with sera of COVID-19 recovered cases and vaccinees of inactivated COVID-19 vaccine BBV152/Covaxin,” *J. Travel Med.*, vol. 28, no. 7, p. taab104, 2021.
[92] OHDP, “OHDP - Researchers - Datasets - OHDP,” 2021. [Online]. Available: https://ohdp.ca/datasets/.
[93] M. C. Buchan, I. Romano, A. Butler, R. E. Laxer, K. A. Patte, and S. T. Leatherdale, “Bi-directional relationships between physical activity and mental health among a large sample of Canadian youth: a sex-stratified analysis of students in the COMPASS study,” *Int. J. Behav. Nutr. Phys. Act.*, vol. 18, no. 1, pp. 1–11, 2021.
[94] COMPASS, “COMPASS CIHR mental health | Compass System | University of Waterloo,” 2021. [Online]. Available: https://uwaterloo.ca/compass-system/compass-system-projects/compass-cihr-mental-health.
[95] A.-H. Tan and others, “Text mining: The state of the art and the challenges,” in *Proceedings of the pakdd 1999 workshop on knowledge disocovery from advanced databases*, 1999, vol. 8, pp. 65–70.
[96] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. Weld, “{S}2{ORC}: The Semantic Scholar Open Research Corpus,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 4969–4983.
[97] WHO, “Global research on coronavirus disease (COVID-19) WHO Databse,” 2021. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-

novel-coronavirus-2019-ncov. [Accessed: 30-Dec-2021].
[98] M. Wilson and P. J. K. Wilson, "Coronavirus Disease 2019 (COVID-19)," *Close Encounters of the Microbial Kind*. pp. 185–196, 2021.
[99] W. E. el. E. M. Salazar, J. Barochiner, "Novel Coronavirus Information Center," *Elsevier*, vol. 20. pp. 2–5, 2020.
[100] Springer Nature, "Coronavirus (COVID-19) Research Highlights | For Researchers | Springer Nature." 2022.
[101] allenai.org, "Allen Institute for AI." 2020.
[102] AVOBMAT, "Research COVID-19 with AVOBMAT – AVOBMAT." 2020.
[103] R. Péter, Z. Szántó, J. Seres, V. Bilicki, and G. Berend, "AVOBMAT: a digital toolkit for analysing and visualizing bibliographic metadata and texts." pp. 43–55, 2020.
[104] Q. Chen, A. Allot, and Z. Lu, "LitCovid: An open database of COVID-19 literature," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D1534–D1540, 2021.
[105] Z. Lu, "Zhiyong Lu - NCBI - NLM," 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/research/bionlp/.
[106] E. Zhang *et al.*, "Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset," *arXiv Prepr. arXiv2007.07846*, 2020.
[107] KDCovid, "KDCovid." 2020.
[108] Azue, "Azure Cognitive Search - Covid-19 Search Demo." 2020.
[109] CADTH, "CADTH COVID-19 Search Strings - CADTH Covid-19 Evidence Portal." 2021.
[110] L. L. Wang and K. Lo, "Text mining approaches for dealing with the rapidly expanding literature on COVID-19," *Brief. Bioinform.*, vol. 22, no. 2, pp. 781–799, 2021.

Learning Structured Information from Small Datasets of Heterogeneous Unstructured Multipage Invoices

David Emmanuel Katz[1], Christophe Guyeux[2],
Ariel Haimovici[1], Bastian Silva[1],
Lionel Chamorro[1], Raul Barriga Rubio [1] and
Mahuna Akplogan[1]
[1] smartlayers.io, [2] Université de Bourgogne Franche-Comté, France

Abstract

We propose an end to end approach using graph construction and semantic representation learning to solve the problem of structured information extraction from heterogeneous, semi-structured, and high noise human readable documents. Our system first converts PDF documents into single connected graphs where we represent each token on the page as a node, with vertices consisting of the inverse euclidean distances between tokens. Token, lines, and individual character nodes are augmented with dense text model vectors. We then proceed to represent each node as a vector using a tailored GraphSAGE algorithm that is then used downstream by a simple feedforward network. Using our approach, we achieve state-of-the-art methods when benchmarked against our dataset of 205 PDF invoices. Along with generally published metrics, we introduce a highly punitive yet application specific informative metric that we use to further measure the performance of our model.

1 Introduction

The problem of structured information extraction from human readable formats to machine readable formats has been an active area of research for both academia and industry due to economies of scale that automated technologies of data extraction provide corporations with. This information extraction is of interest to many disciplines, which see it as a decision support tool and a way to save money. For example, the medical information services of hospitals in most countries in the world must, from the patient record, extract the medical

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 243-258, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121220

procedures performed and encode them according to the ICD10 nomenclature, so that the social security can reimburse the hospital on a fee-for-service basis.

While automatic information retrieval from scanned documents is mature to a certain extent, and there have been a number of success stories, there is substantial room for further application specific innovations. Some documents are more difficult to process, and some applications have specific needs that are difficult to satisfy. For example, the sensitive nature of corporate financial information makes solutions without high levels of accuracy at a target level entity unfeasible.

First of all, each company has its own template, and therefore, in the production phase, we have new templates not contained in the learning set. Moreover, the vast majority of companies interested in automatic information recognition from invoices are small, and can only manually annotate a few dozen to a few hundred invoices. In other words, the learning set is very small, and therefore unsuitable for traditional large scale deep learning techniques. Finally, some fields of these invoices are sensitive and require perfect recognition, such as the invoice amount or the tax percentage. The simple addition or deletion of a digit in these numbers has a large impact in these final results.

This is why we wanted to place ourselves in the following specific context, which has not been looked at much in the literature: small knowledge base, not representative of the variety of templates, and high accuracy required. Considering this specificity, the main contribution of this article is twofold. On the one hand, a system is proposed that achieves competitive results using a small amount of data compared to the state-of-the-art systems that need to be trained on large datasets, that are costly and impractical to produce in real-world applications. On the other hand, we propose to report the Levenshtein ratio [13] of predicted entities and annotated entities, as a good $f1-$score is not a guarantee of a correct recovered target entity such as a VAT or Invoice Number.

The remainder of this article is organized as follows. State of the art methods for invoice extraction are presented in the next section. Our proposal is explained in details in Section 3. An experimental evaluation is provided in Section 4, and obtained results are discussed too. This research work ends by a conclusion section, in which the contribution is summarized and intended future work is outlined.

2 State of the art

The extraction of information from documents like invoices has long been seen as a task of sequence labeling: after extraction with a OCR tool of the whole text from documents, each obtained token receives a label that follows a rule-based approach for old methods. These rules were defined by humans and based on trigger words, regular expressions, or even linguistic descriptors (e.g., amounts are in the neighborhood of "total" or "VAT"). Such rule-based approaches are used, for instance, in [10] and [6]. Then, as in a lot of fields of research, rule-based methods have been supplanted by machine learning ones [9], like ran-

dom forests [19] or support vector machines [21]. During the most recent years (2020's) deep learning methodologies have surpassed other traditional machine learning techniques, as the experimental evidence of [14] provides.

Among these machine learning based approaches, a collection of tools are available that share the same hypothesis, which is: all possible templates are available in the training set. This is the case for instance in [8], in which a database of field positions is required for each template. This is the case too for [15], where all (field, pattern, parser) triplets must be manually provided for each template, while other approaches of this kind can be found in, e.g., [6] and [4] for a specific invoice framework. Such an hypothesis is however problematic in our context, as in practice, every new customer comes with its own particular form of invoice, which most of the time presents a new template, different from anything already known.

In recent years, deep learning methods have outperformed the more classical ones, both in terms of precision and recall. The most known deep learning based approaches are listed hereafter. Note that these methods frequently need a huge training dataset, which can only be produced by big companies in invoices context.

CloudScan is a commercial software proposed by Tradeshift [17]. It is based on a recurrent neural network that has been trained to recognize 8 fields on a corpus of 300k invoices. A good point is that this platform requires no configuration and does not rely on statistical models. However, half of the 8 fields presented an $f1-score$ lower than 0.87 (e.g., 0.76 for order ID), while to be useful in the invoice information extraction context, a larger score is required. And while invoices are often written in both vertical and horizontal directions, they only choose to apply a left-to-right order. This problem of not considering the spatial information into the key information extraction process is circumvented by CUTIE [23], which stands for Convolutional Universal Text Information Extractor. This is a model based on spatial and contextual information provided to a convolutional neural network coupled with a word embedding layer. Obtained results are quite good, but it presuppose to be able to extract a gridded version of the text, which is problematic in the invoice case. Furthermore, as is a shared theme across invoice extraction models such as [14], its learning stage needs thousands of labeled documents.

Finally, and to the best of our knowledge, the only article that faces the same constraints than us about the dataset size is [11]. They propose two methods which are based on neural networks, and focus on the trade-off between data requirements and performance in the extraction of information. The first method consist to adapt Named Entity Recognition systems for fields extraction of invoices, by fine-tuning BERT [7] to this specific task. The second "class-based" method adapted the features of CloudScan and proposed some extra features to automatically extracts the features, with no preprocessing step nor dictionary lookup. However, the $f1-$scores they obtain are always close to 0.80, which are far from being useful in practice. Furthermore, the smallest set they consider for learning is still huge (20k), which is very far from the concrete use context we aim at. Finally, they only use 1D information as their method is

Natural Language Processing oriented, while invoices structure is 2D.

3 Methodology

Our task consists of taking a set of raw PDF documents with every word on every page being labeled, in order to train a statistical model that can infer the label of all of the words on every page of a new unobserved document. Every document can contain one or more tokens, totaling n tokens in a dataset: the label can be UNDEFINED ℓ_u, or a given target entity ℓ^i, where $\forall i \in [1, n]$, ℓ^i is one of the text listed in Table 1. The layout and structure of every document is obviously not unique.

'Invoice Date',	'Invoice Number',	'Client Name',
'Company Tax Number',	'Total Invoice Money',	'Client Address',
'Logo',	'Total Invoice Money w/o Tax',	'Company Address',
'Company Phone Number',	'Payment Conditions',	'Total Tax',
'Due Date',	'Client Account Number',	'Client Tax Number',
'Tax Percentage',	'Company Name',	'Additional Bussines Information',
'Company Bank',	'IBAN/Account Number',	'Company Email',
'Client Contract Number',	'Company Website',	'Route Number',
'Client Delivery Address',	'Order Number',	'Penalites',
'Additional Financial Information',	'Delivery Details',	'AdditionalTaxInformation',
'AdditionalLegalInformation',	'Date Ordered',	'Client Phone Number',
'Company Order Number',	'Client Email',	'Incoterm',
'LineItemsTable',	'Delivery Date',	'Company Contract Number'

Table 1: List of target features for the information extraction process

Our process can be separated into 5 key steps preceded by a preprocessing, namely: Training of graph embedding, Feature Extraction from words, Feature Merger, Feedforward Network, and Word Token Merger, see Figure 1. These steps are described hereafter.

Initialization We first concatenate pages into a single image file for each document. Then, big connected components are deleted, which removes specific lines, mainly in tables. This helps the OCR tool to extract boxes within table cells. Concretely, this deletion is operated as follows. The image is converted in grayscale, and a thresholding is applied to it (by using THRES_BINARY_INV+THRES_OTSU from OpenCV [2]). We extract the objects with OpenCV's findContours, and we take the rectangle circumscribed to the object. If its length or width is less than 0.1* that of the image, we hide the rectangle. This preprocessing allows to remove various elements such as signatures.

Finally, text boxes and their content are recognized thanks to Tesseract v.5.0 [20], leading to a set of $\{(w^i, x_1^i, y_1^i, x_2^i, y_2^i, \ell^i) | i = 1..n\}$, where for all i, w^i corresponds to the string of characters inside the bounding box i, x_1^i, x_2^i, y_1^i, and y_2^i correspond to the Cartesian coordinates of the corners of this box, and ℓ^i corresponds to the unique label associated manually to the string character in box i. An example of such tuples: ("9,270.00 EUR", 100, 100, 200, 200, "INVOICE AMOUNT").

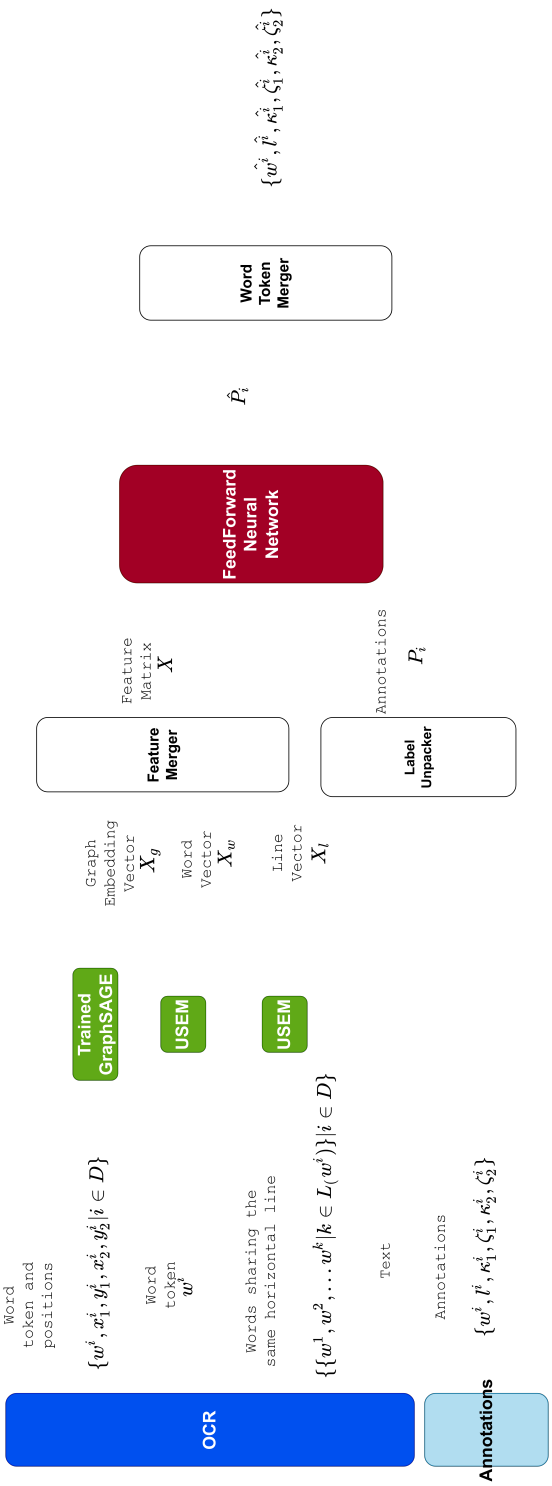


Figure 1: Main pipeline

Training of graph embedding Using the OCR output of Tesseract, we construct a weighted undirected graph for every document in our dataset. The graph nodes are the ocr tokens and the edges are such that two boxes are horizontally connected if they are contiguous in the same line, and they are vertically connected if they have any overlap in the horizontal axis and have no other box vertically between them. The edges are weighted proportional to the inverse of the Euclidian distance between box centers, as depicted in Figure 2. An additional heuristic to decrease memory footprint is that we do not draw edges between tokens that are separated by 5 horizontal lines.

Figure 2: Weighted undirected graph from a given piece of invoice

The individual word cannot be included as an attribute of each vertex in the GraphSAGE [12] step to come, so instead a list of 11 Boolean statistics are produced for each word vertex, cf. Table 2. And words within each node are replaced by these Boolean attributes. The graph construction and graph training process actually does not only take in the 11 Boolean features, but they take in also 3 additional numerical features (to parameterize the nodes in the graph): One multishot vector with the characters of the token (this what we call "Simple Character Encoding"), a one hot vector of the token within the all of the tokens in the corpus (any unseen token is labelled as Undefined), and finally a single TF-IDF score of the token with respect to all the tokens in the document and the dataset.

Finally, an untrained GraphSAGE model is used by training it on the graph structures described above, this model being selected for its ability to encode vertices with properties (here, the Boolean statistics). A fully

is_alpha	is_digit	like_email
like_num	like_url	is_lower
is_punct	is_quote	is_space
is_title	is_upper	

Table 2: The 11 Boolean statistics used in token encoding

trained GraphSAGE model is produced at this stage, fitting an embedding for each node in the graph, the task being the classification of the 39 annotated labels of Table 1, see Figure 3. This step converts each node of the previous graph into a vector X_g of size 1×1024, as illustrated in Figure 4. This 1024 dimension vector systematically represents the bounding box in question, and also captures via the embedding processes samples and aggregations of neighboring nodes.

Figure 3: GraphSAGE training pipeline

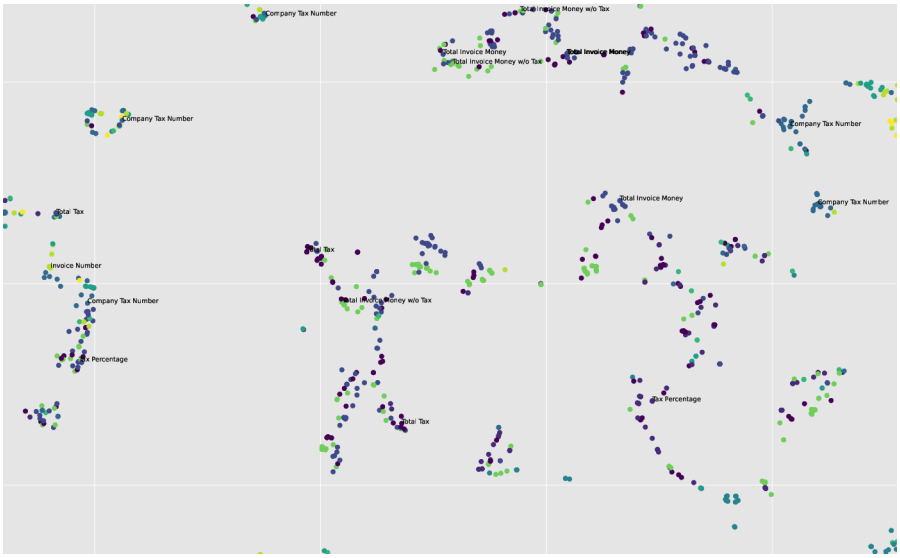


Figure 4: t-SNE [22] reduction of a GraphSAGE embedding, in which inherent structures of the documents are unveiled.

Feature Extraction from words Needed for further downstream processing, we proceed to augment our GraphSAGE nodes by including an embedding of each line and character using USEM [3]. This stage takes the output of the OCR model and aggregates every token with a concatenation of every token that exists on the same horizontal line. An one hot representation at the character level for each token has been applied, while a multihot representation of the tokens in the document has been considered as text vectorizer at the line level.

Both sequences of characters, further referred to as tokens and line tokens (the individual word, and the line) are passed onto a standard Google's Universal Sentence Encoder USEM from TensorFlow [1], to produce two embedding vectors X_w for words, and X_l for lines, both of size 512.

Label unpacking OCR software is not entity aware and will, in an apparently random fashion, split up words or merge them together to create OCR tokens. And due to the potentially poor quality of the documents in question, the OCR model can potentially omit or even add characters that create additional noise. However, these tokens are the only piece of observation we will obtain from unseen data. This is why our annotation process focuses on capturing entire entities that appear on a document. Entities can occupy several words, or even several lines, therefore we need a specific phase that conducts collision detection on the OCR output, to match it with our annotated observations to assign every OCR token i a label ℓ^i.

If a OCR output token is within the area of an annotated token, then the OCR token is assigned the label ℓ^i; if there is no annotated token that surrounds the OCR token, then the OCR token is assigned a label ℓ_u for undefined. After running the label unpacking, we are left with out target vector P with an integer representing each target entity.

Feature Merger With our graph and word embedding vectors created for every word in our document, we proceed to merge our features into a single matrix that can be used for downstream processes: X_g, X_w, and X_l are concatenated to create an individual 1×2048 vector. All tokens are then grouped together to form a matrix X.

Feedforward Network We elected to use a simple feedforward network for the classification task. We train the network using only a subset of 10 target features of particular interest (see Table 3). The network consists of three dense layers of 1024 neurons with activation ReLU plus a dense softmax layer with 10 neurons, while the GCAdam optimizer has been chosen. Due to extreme unbalance of target entities with undefined in our dataset we use sparse categorical cross entropy cost function and $f1-$score for evaluation.

Word Token Merger Due to the nature of our use case, our individual OCR tokens are not suitable for direct consumption by downstream processes.

'Company Name',	'Invoice Number',
'Invoice Date',	'Company Tax Number',
'Due Date',	'Tax Percentage',
'Total Invoice Money',	'Total Invoice Money w/o Tax',
'Payment Conditions',	'Total Tax'

Table 3: List of 10 target features used in the classification task

Our pipeline must take care of assuring end users that the exact sequence of characters (including newlines and spaces) are extracted from a document and labeled with a specific target entity e, hence our training – and most importantly evaluation – pipeline will score target entities at an annotated level, requiring that individual OCR tokens be merged together following a heuristic that depends on a distance metric. Tokens are merged horizontally and/or vertically if their bounding boxes are within ϵ distance from each other and whose labels ℓ match. The word token merger finally produces merged word tokens with next bounding boxes.

4 Experiments

4.1 Experimental protocol

Our invoice dataset consists of individual images for each page of every invoice of our 205 total private invoices. Invoices where assigned a human labeller to draw bounding boxes on every occurrence of a target entity on the page. A second human labeller was used to review the work of the first one. The target entities where one of the labels listed in Table 1. The dataset contained one third of documents in spanish, english and french respectively. It has been split into 70% for training and 30% for testing.

In order to understand the performance gain in introducing graph embedding, we present two baseline methods that are trained directly on OCR tokens and labels, and we compare them to the proposed method. These baseline models consist of a ridge classifier on the one hand, and a neural network on the other hand. They have been trained on the TF-IDF [18] vector of the tokens and the bounding boxes. At each time, the $f1-$score is computed, and it is compared to the results obtained by CloudScan [16] and by Hamdi et al. [11].

These two tools have been chosen, because CloudScan is a very recent and well-known professional deep learning based software that achieves state-of-the-art results, and because [11] is the study the closest of our particular context. Furthermore, they both provide $f1-$scores, or precision and recall. We do not provide further comparison, because these two tools have the highest scores in the literature. And as it has been evoked before, existing work are not directly applicable to our specific context. Finally, document datasets are always private in this area of research, which makes reproducibility and comparison at least questionable in practice.

Name	precision	recall	f1-score
Total Tax	0.00	0.00	0.00
Invoice Number	1.00	0.05	0.10
Invoice Date	0.00	0.00	0.00
Total Invoice Money	0.00	0.00	0.00
Tax Percentage	1.00	0.05	0.10
Company Name	0.28	0.33	0.30
Total Invoice Money w/o Tax	0.00	0.00	0.00
Payment Conditions	0.90	0.15	0.26
Due Date	0.00	0.00	0.00
Company Tax Number	0.68	0.13	0.22
UNDEFINED	0.79	0.98	0.87

Table 4: Baseline 1: Linear Ridge Classifier

Name	precision	recall	f1-score
Total Tax	0.23	0.16	0.19
Invoice Number	0.88	0.35	0.50
Invoice Date	0.67	0.49	0.57
Total Invoice Money	0.40	0.31	0.35
Tax Percentage	0.71	0.77	0.74
Company Name	0.48	0.49	0.48
Total Invoice Money w/o Tax	0.30	0.14	0.19
Payment Conditions	0.87	0.23	0.36
Due Date	0.43	0.55	0.48
Company Tax Number	0.63	0.25	0.36
UNDEFINED	0.87	0.91	0.89

Table 5: Baseline 2: Neural Model

Note finally that our implementation uses Tesseract V, StellarGraph's GraphSAGE implementation [5], and Tensorflow 2.4.3. And we have trained our model in Google cloud platform using a TESLA V100 GPU.

4.2 Obtained results

In what follows, we present the classification statistics and sequence scoring statistics, and some observations and conclusions from them.

Obtained results with the two baselines are provided in Table 4 for the linear ridge classifier, and in Table 5 for the neural network. Results from CloudScan are reproduced in Table 6, while scores obtained by Hamdi et al. can be found in Table 7. Finally, our own scores are provided in Table 8.

Field	LSTM
Number	0.760
Date	0.774
Currency	0.905
Order ID	0.523
Total	0.896
Line Total	0.880
Tax Total	0.878
Tax Percent	0.869

Table 6: $f1-$scores of CloudScan for unseen templates

Fields	NER-based	class-based
docType	0.87	0.90
docNbr	0.69	0.86
docDate	0.85	0.82
dueDate	0.82	0.84
netAmt	0.51	0.74
taxAmt	0.54	0.77
totAmt	0.51	0.86
currency	0.80	0.89

Table 7: $f1-$scores of the two methods proposed in Hamdi et al. [11]

Name	precision	recall	f1-score
Total Tax	0.93	0.86	0.89
Invoice Number	0.97	1.00	0.98
Invoice Date	0.98	1.00	0.99
Total Invoice Money	0.98	0.97	0.97
Tax Percentage	0.76	1.00	0.86
Company Name	0.98	1.00	0.99
Total Invoice Money w/o Tax	0.97	1.00	0.98
Payment Conditions	0.96	1.00	0.98
Due Date	1.00	0.98	0.99
Company Tax Number	0.87	1.00	0.93
UNDEFINED	1.00	0.72	0.84

Table 8: Our Model: Graph Embeddings + Word/Character Embeddings

From the results it is clear that a linear ridge classifier has the worst statistical results, the two state-of-the-art tools do much better than our two baselines. We can also see that adding the embedding phase to the neural network, which is exactly our proposed method, allows to greatly improve the scores. We can also see that the two state of the art tools produce, in the end, a rather bad

classification, with many $f1-$scores either mediocre, or even very low. Finally, our method obtains much better $f1-$scores, proving that our classification can be used in practice. Let us however remark that it is very hard to compare these types of models across different datasets, and it is not totally fair to compare such $f1-$scores, as the quality of the annotations, data, nature of labels, and many other minor details can have a major impact in predictive power.

To further illustrate this results, we propose the use of a finer metric that exactly computes at which edit distance our results are to the ground truth. Obtained results are summarized in Table 9 for further investigations. The average of document averages Levenshtein scores by field is equal to 0.77, while the full average over the whole corpus is of 0.78. These results start to look good, if we take into consideration that we have customized the Levenshtein ratio as follows: if a token is misclassified, the entire Levenshtein score is set to 0, given that the information is not correctly extracted and has no tangible use to end users. This additional penalization allows us to optimize for industry usability. Concretely, these scores more refined than a simple $f1$ mean that our method is useful in practice, in the specific context of this study, even though further improvements are welcome. Note that due to the very low classification score of the baselines, the final Levenshtein ratio is not reported as it is close to 0. Similarly, we cannot compute this ratio for the two state of the art tools, as we do not have access to their database.

Name	Levenshtein Ratio
Total Tax	0.70
Invoice Number	0.95
Invoice Date	0.97
Total Invoice Money	0.82
Tax Percentage	0.68
Company Name	0.81
Total Invoice Money w/o Tax	0.73
Payment Conditions	0.72
Due Date	0.94
Company Tax Number	0.74

Table 9: Levenshtein scores of the proposed method

4.3 Discussion

The main question at the origin of our experiments can be summarized as follows: can the node embedding we propose add predictive power? Intuitively, we can think that it can, because it is always useful to look at what there is, in the document, in the neighborhood of a given token (this is precisely what humans do). The results obtained confirm this intuition, since without this embedding,

we fall back on our baselines, which contrary to our method are unable to make a classification. And when we add embedding to a simple neural network, we get better results than the state of the art, even though they use the best deep learning tools. We can summarize this by saying that the embedding tool is more important than the classification tool, at least in the particular context of our study: embedding neighbor token significantly leads to greater predictive and empirical results.

Another point to emphasize is that the process of unpacking and regrouping OCR produced tokens into full target entities is not a trivial task. End users expect high levels of quality from their invoice information extraction model, and the only way to empirically guarantee the performance of the model is by scoring it on its ability to produce the entire sequence of characters. Most research today in the field only reports classification metrics of individual OCR tokens. Using Levenshtein scores allows us to look at entity level extraction capability and not just accuracy of classification.

Note finally that our model can scale from small to large scale due to online learning and parallelization, while there is room for improvement creating lighter and more powerful versions of this pipeline. We have thoroughly tested our embedding method before arriving at our proposal, but we have only detailed here the best embedding, due to lack of space, and because it is not useful to show techniques that did not have adequate performance. Among these techniques investigated were object detection approaches, that failed to produce satisfactory results with our limited amount of data.

5 Conclusion

In this article we have studied information retrieval from multi language invoices restricted by the amount of data that is available at training with an emphasis of generating highly accurate results applicable to real world corporate automation projects. To stress its real-world applicability, we proposed to consider finer metrics based on edit-distances rather than f1 scores on token classes only. We have been able to show that the graph node embedding step is a key driver of increased accuracy in 2d document information extraction.

Proceeding our implementation and experimentation phases we are able to achieve competitive state of the art results in addition to industry specific applicability. Our end-to-end approach has been detailed and we have performed various evaluations confirming the practical interest of our approach.

In our future work, we would first like to measure the impact of the final classifier on the quality of the results, and see if an improvement can be obtained either by tuning the architecture or its hyper parameters. Increased efficiency and work can be conducted in the token merging staging, by running studies on different graph construction methods. We would also like to understand how human-corrected predictions can be fed back into our pipeline to increase accuracy, and what is the average marginal effect of a newly corrected prediction our evaluation metrics. Finally, we will look at other types of documents, to see

if what has been done on invoices can easily be extended to other contexts of heterogeneous documents.

Acknowledgement

The authors would like to thank Lila Benhammou founder of Humans4Help and Co-founder of Smart Layers. This study was funded by BPIfrance in the frame of the French Deeptech company.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[3] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. Universal sentence encoder for english. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 169–174, 2018.

[4] Francesca Cesarini, Enrico Francesconi, Marco Gori, and Giovanni Soda. Analysis and understanding of multi-class invoices. *Document Analysis and Recognition*, 6(2):102–114, 2003.

[5] CSIRO's Data61. Stellargraph machine learning library. `https://github.com/stellargraph/stellargraph`, 2018.

[6] Andreas R Dengel and Bertin Klein. smartfix: A requirements-driven system for document analysis and understanding. In *International Workshop on Document Analysis Systems*, pages 433–444. Springer, 2002.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805, 2019.

[8] Daniel Esser, Daniel Schuster, Klemens Muthmann, Michael Berger, and Alexander Schill. Automatic indexing of scanned documents: a layout-based approach. In Christian Viard-Gaudin and Richard Zanibbi, editors, *Document Recognition and Retrieval XIX*, volume 8297, pages 118 – 125. International Society for Optics and Photonics, SPIE, 2012.

[9] Ralph Grishman. Information extraction: Techniques and challenges. In *International summer school on information extraction*, pages 10–27. Springer, 1997.

[10] Ralph Grishman and Beth M Sundheim. Message understanding conference-6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, 1996.

[11] Ahmed Hamdi, Elodie Carel, Aurélie Joseph, Mickael Coustaty, and Antoine Doucet. Information extraction from invoices. In *International Conference on Document Analysis and Recognition*, pages 699–714. Springer, 2021.

[12] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 1025–1035, 2017.

[13] V. I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10:707, February 1966.

[14] Bodhisattwa Majumder, Navneet Potti, Sandeep Tata, James B. Wendt, Qi Zhao, and Marc Najork. Representation learning for information extraction from form-like documents. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 6495–6504, 2020.

[15] Eric Medvet, Alberto Bartoli, and Giorgio Davanzo. A probabilistic approach to printed document understanding. *International Journal on Document Analysis and Recognition (IJDAR)*, 14(4):335–347, 2011.

[16] Rasmus Berg Palm, Ole Winther, and Florian Laws. Cloudscan - a configuration-free invoice analysis system using recurrent neural networks. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 406–413, 2017.

[17] Rasmus Berg Palm, Ole Winther, and Florian Laws. Cloudscan-a configuration-free invoice analysis system using recurrent neural networks. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 406–413. IEEE, 2017.

[18] Anand Rajaraman and Jeffrey David Ullman. *Data Mining*, page 1–17. Cambridge University Press, 2011.

[19] Pappu S Rao and Vasumathi Devara. To improve the web personalization using the boosted random forest for web information extraction. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 13(6):1264–1268, 2020.

[20] Ray Smith. An overview of the tesseract ocr engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*, volume 2, pages 629–633. IEEE, 2007.

[21] Aixin Sun, Myo-Myo Naing, Ee-Peng Lim, and Wai Lam. Using support vector machines for terrorism information extraction. In *International Conference on Intelligence and Security Informatics*, pages 1–12. Springer, 2003.

[22] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[23] Xiaohui Zhao, Endi Niu, Zhuo Wu, and Xiaoguang Wang. Cutie: Learning to understand documents with convolutional universal text information extractor. *arXiv preprint arXiv:1903.12363*, 2019.

CHATFORSENIOR: AN INTELLIGENT CHATBOT COMMUNICATION SYSTEM FOR DEPRESSION RELIEF USING ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING

Hanwen Mai[1] and Yu Sun[2]

[1]Orange Lutheran High School, 2222 N Santiago Blvd, Orange, CA 92867
[2]California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

In recent years, loneliness has appeared in lives for both young and old individuals. As cases of the COVID-19 virus are going up people have dealt more with loneliness and depression especially the seniors [5]. Some have even changed their whole lifestyle because they feel empty and isolated. Others will either try to isolate themselves more or use dangerous ways to quickly get rid of the feeling.To solve this major problem, I have created a digital online communication app which young individuals can have long chats with seniors who are alone and lonely. My application uses real time communication systems which can directly be sent to other users without any issues [6]. Our main goal is to have users have their own way of communicating, using familiar designs of applications we all have used before. By using new features we have created a more user-friendly based user experience which can be experienced throughout our application. Using immersive layouts of applications designs, advanced network connections, visual and data based analytic we are able to solve this major problem.

KEYWORDS

NLP, Mobile Dev, AI.

1. INTRODUCTION

Loneliness is a feeling that most people hate experiencing [4]. Whether you are a young individual or an old senior who has nothing else to live for, loneliness is painful. Feeling emotional pain is something that I, as a young individual, hate seeing [7]. It is also something I've experienced throughout my middle school years of being a transfer student from a different country. When I first arrived in America, I felt alone and confused. I barely knew any english at the time which is why it made it hard for me to communicate with other people. This is one of the major reasons I made this app to solve loneliness and get more people involved with other people. Our application solves this problem by having the benefits of talking to many different people around the globe. With the applications randomizer function, users can have a new individual to talk to once they are ready. A simple consequence users might experience is the lack of monitoring. Spme benefit of using this application is that it can solve this world wide problem and give people a good experience. It can also help young volunteers who want to learn more about teaching and motivation skills [8]. In the end, this problem/topic is important, because of

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 259-270, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121221

how major this problem is. With the increased cases of COVID-19 going up everyday, the loneliness rate also goes up. This problem will continue to grow if no one steps in or does anything to come up with a solution. Therefore, I think it is necessary for more people like myself to pay attention to it.

Some of the existing related methods are any communication applications that allow one user to chat with another. Examples of these applications include WeChat, Instagram, Snapchat, etc. These applications all allow a user to directly communicate to another user using either the messaging feature or the video call feature. However none of these applications have such features that include helpful benefits to life like the applications I have built. Their implementations are also limited in scale which only allows users to chat with only the people they know or have gotten their info from [9]. Many other methods such as Omegle, Paltalk, MeetMe, all have another major problem that cannot be solved. Because all these applications are not secure, they are public to all users who are on the application. Privacy is also the respected thing that none of these mentioned applications provide which could lead to unsafe internet web browsing [10]. A big problem with these apps/websites is that they are not secure nor watched by any admin or moderators who are managing the data and servers. This means that there can be suspicious activities and interruptions of users. Without internet censorship many users might experience unpleasant ads or web internet traffic. When no privacy is presented no one will know whether one is at risk or not.

Our main goal is to help out the problem of loneliness and lessen the elders who are suffering from isolation. We have done precise research on how to work and make these applications so many different kinds of users can all enjoy the friendly based experience that we have to offer. The inspiration that made me want to make this app was the realization of how little I talk to my own grandparents [11]. Realization of this also made me discover a online public communication website called omegle, which is also a big contributor that inspired me to create this project. Some good and useful features of my applications are, first, the easy to use sign in and login screen. With easy access without any need of a third party app you can directly create a new account and start enjoying the apps experience. In comparison to websites like omegle, you will need to provide extra unnecessary authentication steps which can both cause time and safety. Secondly, our application provides a simple messaging feature where one can start messaging to another user right away. Last but definitely not least, we provide a new added feature of the random user finder, which allows users to find different registered users to chat with, similar to omegle. When clicking the button, the user will see a popup of the found user and be provided with a choice of deny or accept. Unlike omegle, we do not have a voice nor a camera feature due to it being quite unsafe for both elderly and young people. Therefore, we think that our application could perhaps help out the seniors and loneliness of many of different ages.

Having tested our applications on two scenarios, which are both using the help of my close friends, we have demonstrated that the techniques that have decreased the isolation of elders. Firstly, I gave one of my friends careful instructions on what to do each week. By using planned evaluation we can get an accurate result. We showed the usefulness of our project by making the test subjects use the application once per week. Using the application to communicate with other participants such as their grandparents. By the end of the week, we have gotten a result that they have gotten closer than ever with their grandparents. They are no longer isolated and alone but rather pleased that their grandchildren talked to them. Secondly, we used the apps provided feature of the random search finder. I have also asked one of my other friends to test this feature out to help out with my experiment. I have asked my friend to talk to a new person every day to see if the engine works or not. By having many registered users in the test we have gathered an accurate result that the random search finder is accurate and will find a different user every time. This feature has proved that by using our application you can find new people to talk to which

can spread the positive and the message that I was a creator is trying to spread out. In conclusion our experiments/test is overall positive which is proof that this application can benefit the problem of loneliness and isolation.

The rest of the paper is organized as follows: Section 2 challenges and problems that occurred while making the project and user's feedback. Section 3: explanation of out solution and carefully planned visuals. Section 4: experimentation and tests on individuals. Section 5: Related work and inspirations. Section 6: conclusion and future work and improvements.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Coming Up Idea and Topic

One of the most difficult challenges of making this project come to life is actually coming up with the idea and problem/topic. The fact that we have to come up with a global wide problem got us stuck for days thinking what is necessary to be solved. Since our goal was to accomplish this problem and lessen this problem our team came up with the communication app. By using this application users can find more people to talk to and get closer to their close ones. The biggest risk of making a communication app is the privacy policy and getting peoples trust and approval. by providing good information we have gotten and ensure that everything is safe and usable for the users. When using our application users can feel a user-friendly experience while chatting.

2.2. Connecting to the global database

A second challenge that we have to face while creating this project is connecting to the global database. We will need to have a applications help to connect both the data and the service so we can get the chat to work. To solve this we have used the popular firebase console which utilizes and specializes in things like these [12]. By connecting our servers to the firebase database we can get a visual of how each user is able to communicate and their history of the communication. Seeing this, we are able to get real time messages being sent by both registered users. This also helps with the registered user count which gives us a precise amount of registered users that are currently utilizing our server. With the help of this we are able to get real info sent through users without any interruptions.

2.3. The use of Our in App Feature

The last and final challenge we had to face was the use of our in app feature, the random search finder [13]. This feature allows users to click and find a new stranger to talk to and chat with them. One big problem that has occurred while making this was the process of connecting to the real time registering users. If not connected, users will not find any users while using this feature. The difficulty of doing this is how we can have each new registered user be connected to the server and have pre-registered users link up to the new ones. The solution we found was simple. It was to make a code so that the registered user will directly link to the database right after registering. This allows both users to see the same uptime of the registered users which then allows the search finder to work properly.

3. SOLUTION

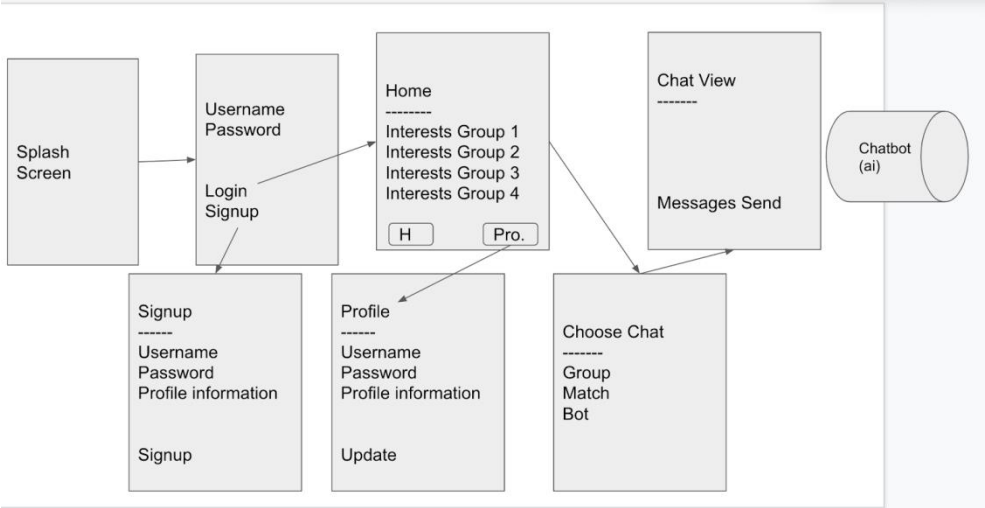


Figure 1. Basic layout and planning format draft

Chat4Seniors is a digital mobile communication application, where users can chat online through our database-connected messaging feature. Young volunteers can log on to the app and quickly make an account to start chatting with seniors that have already registered into our database. With our random search finder, seniors can find and chat with any random volunteers around the globe. Chat4Seniors also provides a user-friendly experience, which we will respect your privacy to the fullest and make sure that all application features work the best to your standards. This application's components work like any other application users have used before. The splash screen is digitally connected to the login and sign-up screen, which upon clicking on the app, it will directly force the user to the login screen. After logging in or creating an account, the system will automatically transfer the user to the next screen which is the application's main home screen. Here is where the user can use all our application features, such as the random search finder, settings, and messages. The settings icon is where the users can find our log-out button which will bring the user back to the login and sign-up screen [15]. The main technical challenge of our system is managing the database and recording each user's registered accounts. We have solved this problem by having our system be connected to the firebase console database; which can track real-time user interactions and account registrations. In order to achieve our desired goals, our application consists of 5 main components:

- A data layout of all different screens and functions and features
- A system connection to the database which stores personal data and registration info.
- A messaging screen which sends real time messages and users advance signals for connections.

For your further information we have provided detaily explained diagrams and pictures displaying all of the components and methods used in this project.

Figure 2. Icon

```
flutter_icons:
  android: "launcher_icons"
  ios: true
  image_path: "assets/icon.png"
```

Figure 3. Code of icon

3.1. Splash Screen

The splash screen is a screen which will only appear for a certain amount of time. It also uses a simple code (code above) which directly uses our selected picture (icons.png) from our selected folder (assets). After the time is up, the splash screen will automatically transfer the user to the next screen which is the log-in screen. On this page the user can see both the log-in and the sign-up button.

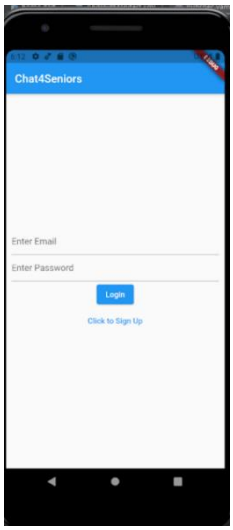


Figure 4. Log-in/Sign-up screen

3.2. Log-in/Sign-up Screen

The log-in and sign-up screen follow a similar concept with the splash screen. However, it does not have a time limit, it will only move to the next screen when either the user clicks on the login or the signup button. As for the login and signup, the user can either choose to log back in to their existing account or they can click the “sign up” button and create a new account. Once inside the sign up screen (Figure down below) the user will be able to create an account. The username and password however will require at least 3 letters and 6 numbers. If the user inputs something that doesn't meet the requirements, it will reject their request for creating a new account. Both screens use a code which utilizes the “Elevated button” code. The code simply adds a “click” command to your chosen buttons, and once a user has clicked the code it will transfer the user to the next screen.

```
return Scaffold(
  appBar: AppBar(
    title: Text('Chat4Seniors'),
  ), // AppBar
  body: Container(
    padding: EdgeInsets.symmetric(horizontal: 10),
    child: Form(
      key: widget._formKey,
      child: bodies[index]
    ) // Form
  ), // Container
); // Scaffold
}
void toggleIndex({required int targetIndex}) {
  setState(() {
    index = targetIndex;
  });
}
String? emailValidator(String? email) {
  if (email != null && !email.contains('@')) {
    return 'Please enter a valid email';
  }
  return null;
}
String? usernameValidator(String? username) {
  if (username != null && username.length < 3) {
    return 'Please enter a valid username with at least 3 characters';
  }
  return null;
}
String? passwordValidator(String? password) {
  if (password != null && password.length < 6) {
    return 'Password must be at least 6 characters long';
  }
  return null;
}
```

Figure 5. Code of sign up screen

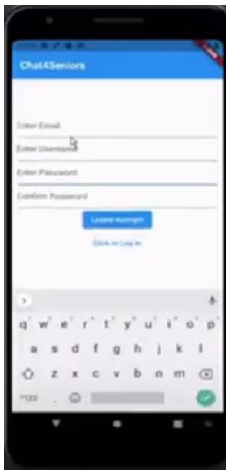


Figure 6. Home page

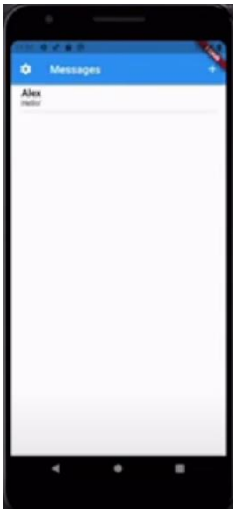


Figure 7. Message page

3.3. Home Screen

After the users have created an account or logged in to their existing account they are able to access the home screen which will provide them with many different features to experiment with. The home screen is the screen that many users will spend their most time on because it has the messaging feature which is used to communicate between users. In the figure above it shows the main screen with a fake user named Alex. On the upper left hand corner, there is a gear icon. This button will navigate the user to the sign out screen where the user can log out on demand whenever they want.

Figure 8. Setting page

Figure 9. User randomizer

3.4. Plus Icon/User Randomizer

In the upper right hand corner of the home page there is a plus icon which upon clicking will show a pop up showing a random user that has been found nearby. The current user can either accept the request or deny it. Upon accepting it, the user will see that a new tab of message will appear. The tab will show the most recently accepted user. In order to keep everything organized the randomizer will only search for people who have registered into our database and are nearby to you. As for the database and account checking we have linked our system to the firebase console which will track each and every user and their registration.

Figure 10. Messaging screen

3.5. Messaging Screen

The messaging screen is the screen that the user uses to communicate with other users. The blue bubble of a message (Seen Above) is the sender's message color. The sender will send a blue bubble message while the receiver gets it in green. The receiver's message icon is green. The messaging screen is only for one purpose which is to communicate and send messages to the other communicator.

4. EXPERIMENT

4.1. Experiment 1

Our solution solves the problem by making users communicate to more people and get less lonely elderlys. The problem is solved because we have helped out and built our own feature of the random search finder which a user can use to find any other registered user to talk to. With the help of this the solution is now fixed and there will be less lonely people due to the design of this advance feature. Throughout out time experimenting to see if our hypothesis was true or not I have asked many people including my own family members to try our application. We have provided them with instructions and guides of how to navigate and use the app effectively, with times of when to use the app, how to use the app, and how the app is helpful. As a result most of the feedbacks and comments came out positively with most of the people saying that this app was helpful and only a small amount saying that it was non helpful or just semi helpful.

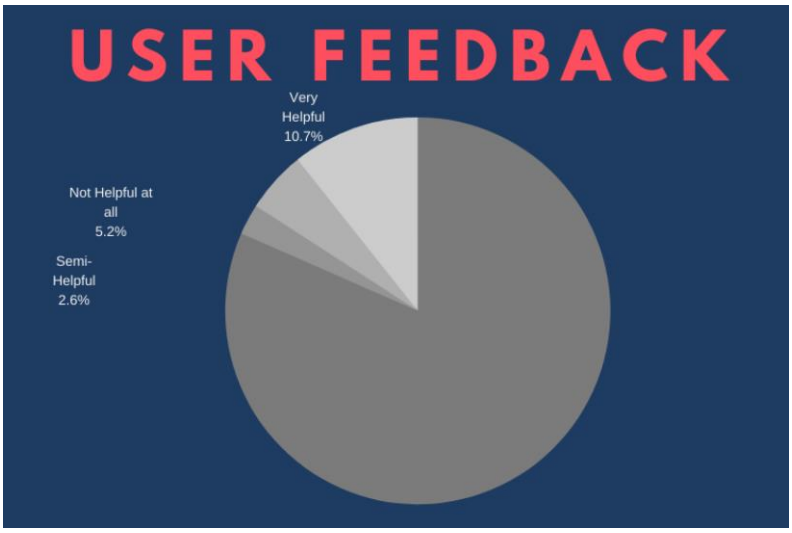


Figure 11. User feedback 1

The first experiment was a major success as almost 80% of the people said that our application was helpful. Looking at our graph we can see that only a small amount of people (10%) said that this application was very helpful. With the knowledge of this I have figured that many of my close friends are indeed experiencing or know someone who is experiencing isolation and loneliness. The result of this graph shows that most of the people find it helpful which means that our app is capable of solving this problem. Therefore I believe that the app is helpful to some and will help many who struggle with the problem while experimenting with it.

4.2. Experiment 2

With our second experiment we have solved the problem of getting the problem out to many other people including strangers. The solution was to get random people who wanted to help someone they know that is currently struggling with loneliness. Based on our experiment I have asked several strangers around school and community areas to try out our app. We gave them a demo account which was authorized by us and let them chat with any of their own friends. As a result of our second experiment many came back and rated that the app was too simple. Many said that there are many apps like this which are more helpful and have more features. However, around 30 percent of the people I've given this app to, said that the app actually helped them get to know more people and talk about their own struggles with strangers online who are also using the app.

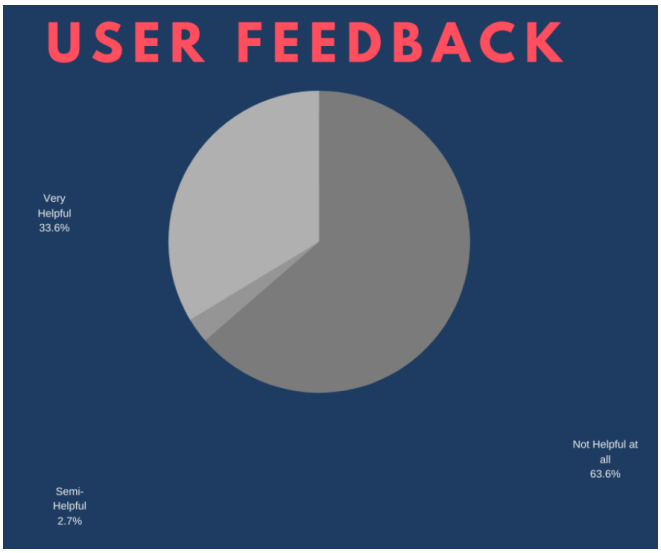


Figure 12. User feedback 2

Based on our graph of the second experimentation we can see that 63% of the total 20 strangers I've tested the app on said that it was not helpful. However 33% of the people said that it is helpful. This shows that strangers are more reliant on other application with the same features and functions, which makes my application seems simple and plain. With the information I've gathered many said that the app was just way too simple or the app had some small issues or problems. The remaining 2% of the people said that the app was semi helpful just thought that the app was creative and uses features of other apps. Due to the feedbacks I've gathered I have concluded that there will be more improvements to be made to the application in order to achieve further success in solving our problem.

In conclusion, the two experiments that we have conducted were overall a success with feedback that said that it helps them in some way. Many people in the first experiment said that the app was indeed helpful which helped them with loneliness and isolation. The problem that I've gotten so far from the first experiment was that It was too easy to use or had too few features. However we have overcome the challenge of trying to solve loneliness. Looking at our second experimentation we have gotten more negative feedback than positive. Since we tested this on strangers most of them didn't enjoy the app as much as my friends and family did. This is mainly because of the wide variety of other similar apps that have the same features as mine. Many have said it was just too underwhelming and not getting the same experience as other apps were giving them. By the end of the day, I think this application did quite meet my expectations with the results of both experiments. Though I think I would have expected more of a positive result from the strangers I tested my app on since we are mainly trying to focus on the worldwide problem of loneliness. Therefore, I think my app is a success in a way because It could help some people out with their problems and struggles.

5. RELATED WORK

Margaret Lubas, Jeniffer Mitchell, Gianluca De Leo have presented us with a digital alternative device that helps children with severe autism disorders. [1] They have created a Augmentative and Alternative Communication App with ads for kids with autism, using many different methods including a input and output design with screens and visuals to help the learning experience. [1] Comparing the two works of mine and theirs we have a similar purpose of trying to solve a worldwide problem. Although we have different app functions and features our

application can still solve problems that need to be looked at. The strength and differences of my app to his is that I definitely have more advanced features such as our random search finder. But he had used Graphic-based methods that allow for the creation of messages through combining symbols and images, and can often then be translated into speech.

Walker Z, McMahon DD, Rosenblatt K, Arner T have created an application which utilizes the similar format of the popular game Pokemon GO and uses its Augmented reality to create a learning environment for people of all ages. [2] Their work is fairly complicated with uses of advanced technology such as AR and UDL technology to help with this Augmented reality. To compare my work we also have a similar purpose of helping people. However what makes us different is that we are helping a specific group of people, the seniors and they are just helping the overall group of the whole population.

Guarino, G. Aceto, D. Ciuonzo, A. Montieri, V. Persico and A. Pescapè have made an app that analyzes the effect and pandemic speed and increase rates of the recent COVID-19 virus. [3] with rates going up they have created an app that helps see the cases and improve upon it to have an accurate value. The comparison between my app and theirs is that we have more of a messaging app not a analytical data based graph app. The main difference between ours is that we trying to spread awareness to the other people of the world. With both of our apps wanting to solve a problem that exists in the world.

6. CONCLUSIONS

To conclude we have made an online messaging app which helps out the major problem of loneliness and isolation in seniors [14]. With the use of our online database and real time messaging between users we can find that communication can be easier. Based on our experiments we can tell that most of the tests were positive and came back making the app better. Our method of the random search finder really helped many who tried the app and liked it, it also helped many to meet new people who are also trying to find other people.

The current state of the app is poor and is in definite need of improvements. Before starting the project I have thought of adding in maybe more features like a profile picture setting or a bio for the users. I have also thought about adding a video chat function where users can share their thoughts as a group. As of right now, the limitation of the app is that you can't have a full-on experience with the other user. In the future, I would like to add more advanced features that can help with a better user experience and improve it so that many users can have that experience with the other user.

To further evaluate our problem we have decided to keep on improving the application which can help out the problem much better. By improving features and breaking limitations we can make the app more useful to the everyday people who need it and want to use it for good.

REFERENCES

[1] Lubas M, Mitchell J, De Leo G. User-Centered Design and Augmentative and Alternative Communication Apps for Children With Autism Spectrum Disorders. SAGE Open. January 2014. doi:10.1177/2158244014537501

[2] Walker Z, McMahon DD, Rosenblatt K, Arner T. Beyond Pokémon: Augmented Reality Is a Universal Design for Learning Tool. SAGE Open. October 2017. doi:10.1177/2158244017737815

[3] Guarino, G. Aceto, D. Ciuonzo, A. Montieri, V. Persico and A. Pescapè, "Characterizing and Modeling Traffic of Communication and Collaboration Apps Bloomed With COVID-19 Outbreak,"

2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI), 2021, pp. 400-405, doi: 10.1109/RTSI50628.2021.9597263.
[4] Hawkley, Louise C., and John T. Cacioppo. "Loneliness matters: A theoretical and empirical review of consequences and mechanisms." Annals of behavioral medicine 40.2 (2010): 218-227.
[5] Altar, C. Anthony. "Neurotrophins and depression." Trends in pharmacological sciences 20.2 (1999): 59-62.
[6] Bondy, Andrew S., and Lori A. Frost. "The picture exchange communication system." Focus on autistic behavior 9.3 (1994): 1-19.
[7] Borsook, David, and Lino Becerra. "Emotional pain without sensory pain—dream on?." Neuron 61.2 (2009): 153-155.
[8] Geen, Russell G. "Social motivation." Annual review of psychology 42.1 (1991): 377-399.
[9] Obermann, Stuart F., and Michael J. Flynn. "Division algorithms and implementations." IEEE Transactions on computers 46.8 (1997): 833-854.
[10] Montgomery, Alan L., and Christos Faloutsos. "Identifying web browsing trends and patterns." Computer 34.7 (2001): 94-95.
[11] Thrash, Todd M., et al. "The psychology of inspiration." Social and Personality Psychology Compass 8.9 (2014): 495-510.
[12] Moroney, Laurence. "The firebase realtime database." The Definitive Guide to Firebase. Apress, Berkeley, CA, 2017. 51-71.
[13] Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." Journal of machine learning research 13.2 (2012).
[14] Wang, Ian J., and Gideon S. Bradburd. "Isolation by environment." Molecular ecology 23.23 (2014): 5649-5662.
[15] Burks, Arthur W. "Icon, index, and symbol." Philosophy and phenomenological research 9.4 (1949): 673-689.

EMAILTRACKER: AN INTELLIGENT ANALYTICAL SYSTEM TO ASSIST EMAIL EVENT TRACKING USING ARTIFICIAL INTELLIGENCE AND BIG DATA

Joyce Zheng[1] and Yu Sun[2]

[1]Central High School, 1700 W Olney Ave, Philadelphia, PA 19141
[2]California State Polytechnic University, Pomona,
CA, 91768, Irvine, CA 92620

ABSTRACT

Recent studies have shown an increasing demand for software that assists in social media applications, such as Gmail [1] [2]. This paper develops a software that utilizes a pixel image in order to assist Gmail users with the status of their email [3]. This intelligent analytical system can be used to tell whether an email has been opened or not. After conducting a qualitative evaluation of this approach, the results provided evidence of the system's usability and the reliability of the system to give accurate results and data.

KEYWORDS

Chrome Extension, AI, Machine Learning.

1. INTRODUCTION

The technology of instantaneous communication over the internet began with Ray Tomlinson's creation of the electronic mail, which we refer to as Gmail [4]. At the time, messages could only be left on the same computer, however while working for ARPANET, Tomlinson came up with the idea to create a medium on the ARPANET system that allowed users to send electronic messages across multiple computers [5]. However, as user activity increased, problems relating to the status of emails began to appear. This application utilizes an invisible pixel image to assist users in tracking and understanding the status of their sent emails. Each pixel image contains its individual tracking number, which triggers the sender's system, thus allowing them to precisely see whether the email has been opened or not. Not only does this application save time for both the recipient and the sender, it also reduces the rate of follow-up emails [6].

Most of the existing systems that have been proposed are paid applications that notify users on the status of their email and allow users to view the data and analytic of multiple emails at once. However, these proposals don't take into consideration the rate of consumers that are willing to pay for the application, which is rarely the case in most practices. According to a study of 100 million users, the results have shown that only about 5% of current app users spend money on in-app purchases. Their implementations are also limited in scale, with samples given for … Other techniques, such as … They … Because …, the method/algorithm used cannot be too sophisticated and often results in … A second practical problem is that some users find it hard to understand…

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 271-276, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121222

The rest of the paper is organized as follows: Section 2 gives insights on the challenges that we met during the designing of the email tracking application; Section 3 introduces solutions to the challenges mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, followed by a comparison of other similar works in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. How to ensure accurate results despite Gmail's security system

One challenge while creating the application was being able to get accurate results in spite of Gmail's security system [7]. Gmail's security system scans through emails, including attachments and links, for malicious content. The system might give a false positive because Gmail's security system had scanned and loaded the pixel image. This interferes with the application's ability to give accurate insight into the status of the email.

2.2. Difficulties with email tracking for group emails and forwarded emails

Because this email tracker uses an individual pixel image to track an email, it is meant for one person. Consequently, the extension is unable to create a tracking link for an email sent to multiple recipients. Because the image tracking link is sent with an email, it only tracks the message that includes the pixel image. This means that forwarded emails are unable to be tracked. However, chrome extensions like MailTrack have used other codes in their system to allow them to track group emails [8].

2.3. Recipients might have tracking blockers and other extensions that might render the accuracy of the email tracking

Email tracking blockers like Ghostery and Smart Pixel can prevent the pixel image in the email from loading, making tracking emails sent to recipients who have these blockers impossible [9] [10]. Additionally, if a recipient previews an email, the pixel image might still load, which gives the system a false result.

3. SOLUTION

Email Tracker is a chrome extension based on a system of pixel images that notifies users on the status of their email and allows users to view the data and analytic of multiple emails at once. The system uses the recipient's email address as well as the subject in regards to the email in order to create an individual tracking link, which is extremely helpful in identifying and tracking certain emails. This tracking link generates a pixel image, which allows senders to precisely see whether the email has been opened by the intended recipient or not. When the receiving end opens the email containing the pixel image and the pixel image is loaded, the image sends a message back to the server indicating that the email has been opened by the recipient. As shown in the graph above, the sender would compose an email, by inserting the recipient's email, email subject, and the message. Then, the sender would click on the email tracker extension in the upper right corner, which opens a window that generates a unique tracking link. Before sending the email, the user would insert the link into the photo section of the email, which generates an invisible pixel image. When the recipient clicks on the email, the pixel image is loaded, which sends the message back to the main system, allowing the sender to see that the email has been opened.

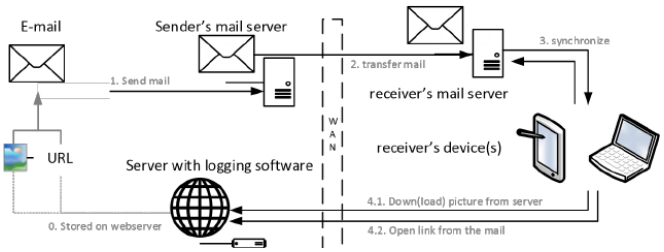


Figure 1. Overview of the system

The system uses the code to the right to create a window that creates the url for the pixel image file. The window utilizes the inputted subject and recipient to create the individual tracking link.

```
@app.route('/test')
def testImage():
    email = request.args.get("email")
    subj = request.args.get("subject")

    ip_list = request.access_route
    print(ip_list[0])
    temp = ip_list[0].split('.')
    my_ip = temp[0] + '.' + temp[1]

    if my_ip == '130.211':
        print("I opened the image!")
    else:
        currEmail.append([email, subj])

    return send_file("smallpixel.png", mimetype = "image/gif")
```

Figure 2. Screenshot of code 1

The system uses the code to the right in the background to request the data from the content's send message (such as the recipient name, recipient's email, and the subject) and send it back to the content.js to create the tracking link

The sender would then copy the tracking link and insert it into the photo section of the email and send it. Once the recipient opens the email, the pixel image will load and send a message back to the system to let them know that their email has been opened.

```
chrome.runtime.onMessage.addListener(
    function(request, sender, response){
    console.log(request);
    const arr = request.split("&")
    var d = JSON.stringify({"name": arr[0], "subj": arr[1]});
```

Figure 3. Screenshot of code 2

4. EXPERIMENT

4.1. Experiment 1

This experiment uses the results of 10 sent emails to show the accuracy of the system's ability of showing the status of an email (before the email is opened and after the email is opened). This experiment shows that despite the challenges faced, such as overcoming Google's security system, the email tracker still gives reliable results.

Table 1. Result of experiment 1

Email Trials	status before the email is opened	status after the email is opened
#1	unopened	seen
#2	unopened	seen
#3	unopened	seen
#4	unopened	seen
#5	unopened	seen
#6	unopened	unopened
#7	unopened	seen
#8	unopened	seen
#9	unopened	seen
#10	unopened	seen

Based on the results of the experiment, out of the 10 sent emails, almost all of them give accurate results. For the 6th trial run, after the email is opened, the email initially gives a status of unopened, but after reloading the page, it gives “seen”. Consequently, this proves that our system is reliable and capable of giving correct results.

4.2. Experiment 2

This experiment uses Mailtrack and our system to compare and contrast the results and abilities of different email tracking systems by sending 10 emails with both systems (before the email is opened and after the email is opened).

Table 2. Result of experiment 2

Email Trials	Our System: status before email is opened	Our System: status after email is opened	Mailtrack: status before email is opened	Mailtrack: status after email is opened
#1	unopened	seen	not opened	opened
#2	unopened	seen	not opened	opened
#3	unopened	seen	not opened	opened
#4	unopened	seen	not opened	opened
#5	unopened	seen	not opened	opened
#6	unopened	unopened	not opened	opened
#7	unopened	seen	not opened	opened
#8	unopened	seen	not opened	opened
#9	unopened	seen	not opened	opened
#10	unopened	seen	not opened	opened

The results of the experiment shows that although both systems give accurate results, the abilities of the two email trackers vary. For example, our email tracker uses an invisible pixel image that tracks the email without the recipient knowing. However, Mailtrack includes a watermark in their emails, which is visible to both the sender and recipient. Additionally, Mailtrack requires the user to pay for upgraded services, however, our system does not.

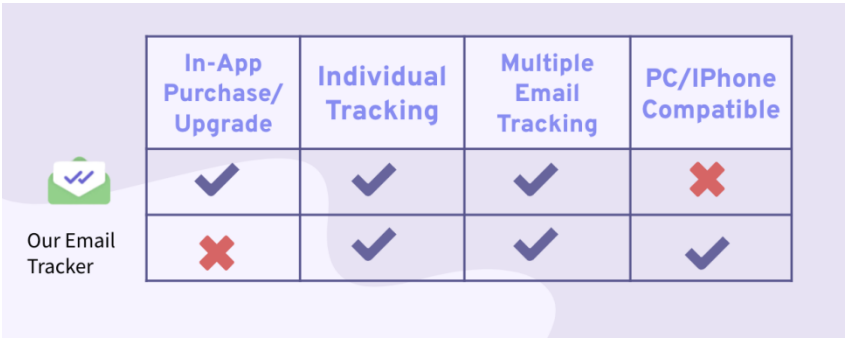


Figure 4. Vary results of Email Tracker

The experiment of the system ' s accuracy proves that despite the challenges faced, such as overcoming Google ' s security system, the email tracker still gives reliable results. The experiments also show that despite many other competitors in the email tracking field, our system has abilities and tools, such as not having to pay for upgraded versions, individual tracking, and anonymous tracking, that many other email tracking extensions don't have.

5. RELATED WORK

MailTrack is an email tracking chrome extension that utilizes pixel images to notify users on the status of their email and allow users to view the data and analytics of multiple emails at once [11]. While our system requires the user to copy and paste the given tracking link, MailTrack automatically tracks an email by pressing "compose". You can choose to not track the email by deleting the MailTrack watermark included in the email. However, this also means that the recipient can see that the email is tracked.

Mixmax is an email tracking software for Gmail that gives users email tracking alerts [12]. It also allows users to see which individual opened the email in cases with group emails. This feature is something not available on our software. Like our email tracking system, mixmax offers individual tracking. They also offer statistics and insights into how many times the email has been opened. However, the only disadvantage is that mixmax uses a watermark visible to both the recipient and sender.

Snovio is another email tracking extension that tracks emails using a pixel image system [13]. Like Mailtrack and Mixmax, it offers desktop push notification, something that our email tracker extension doesn't offer. It also allows users to see the full history and insights of an email. This includes the date and time they viewed the emails and the number of times the email is opened. However, based on some of the reviews, many users experience false results. For example, by opening the email you sent, it gives the system a false "seen".

6. CONCLUSIONS

Email Tracker is a free chrome extension that utilizes an invisible pixel image to help users track and understand the status of their email. The pixel images are devices of individual tracking links, which will trigger the sender's system when the images are loaded. The experiments shown above have verified the accuracy of the email tracking (in giving accurate results) and shown the email tracker in comparison to other email tracking devices. The experiment of the system's accuracy shows that out of 10 sent emails, 9 emails give accurate initial results. The results also show that in spite of Gmail's security system, which scans emails (and unintentionally loads the

pixel images) for malicious content, our system still gives accurate results. Additionally, in comparison to MailTrack, it has more pros, such as having individual and multiple tracking, IPhone and Android compatibility, and allows users to use the extension (and all of its tools) without an app purchase [14].

The experiment on the system's accuracy shows that out of 10 sent emails, 9 of them give initial accurate results. For the last email, after the email is opened, the email initially gave a status of unopened, but after reloading the page, it gave "seen". This shows that the user might have to reload the page a few times to get accurate results. Compared to other email tracking applications, our software requires a few extra steps to retrieve the tracking link, such as entering the email's subject and recipient to generate a tracking link.

In the future, I plan to fix the code in order to get the system to give accurate results immediately, without reloading the page [15]. Additionally, I plan to change the layout of the tracking software, by allowing users to press track on an email without having to physically generate a link.

REFERENCES

[1] Carr, Caleb T., and Rebecca A. Hayes. "Social media: Defining, developing, and divining." Atlantic journal of communication 23.1 (2015): 46-65.
[2] Chen, Mia Xu, et al. "Gmail smart compose: Real-time assisted writing." Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019.
[3] Fisher, Peter. "The pixel: a snare and a delusion." International Journal of Remote Sensing 18.3 (1997): 679-685.
[4] Sajithra, Karin, and Rajindra Patil. "Social media–history and components." Journal of Business and Management 7.1 (2013): 69-74.
[5] Hauben, Michael. "History of ARPANET." Site de l'Instituto Superior de Engenharia do Porto 17 (2007).
[6] MacFarlane, Andrew, et al. "Sender vs. recipient-orientated information systems revisited." Journal of Documentation (2021).
[7] Dunning, Ted E. "Accurate methods for the statistics of surprise and coincidence." Computational linguistics 19.1 (1993): 61-74.
[8] Liu, Lei, et al. "Chrome Extensions: Threat Analysis and Countermeasures." NDSS. 2012.
[9] MacBeth, Sam. "Tracking the trackers: analysing the global tracking landscape with GhostRank." Retrieved October 3 (2017): 2018.
[10] Hinton, H. Scott. "Progress in the smart pixel technologies." IEEE journal of selected topics in quantum electronics 2.1 (1996): 14-23.
[11] Xu, Haitao, et al. "Privacy risk assessment on email tracking." IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018.
[12] Savvidy, Konstantin G. "The MIXMAX random number generator." Computer Physics Communications 196 (2015): 161-165.
[13] Cselle, Gabor, Keno Albrecht, and Rogert Wattenhofer. "BuzzTrack: topic detection and tracking in email." Proceedings of the 12th international conference on Intelligent user interfaces. 2007.
[14] Remneland-Wikhamn, Björn, et al. "Open innovation, generativity and the supplier as peer: The case of iphone and android." International Journal of Innovation Management 15.01 (2011): 205-230.
[15] Haupt, Johannes, et al. "Robust identification of email tracking: A machine learning approach." European Journal of Operational Research 271.1 (2018): 341-356.

DETECTING DEPRESSION IN SOCIAL MEDIA USING MACHINE LEARNING

Ruoxi Ding[1] and Yu Sun[2]

[1]Woodbridge High School, 2 Meadowbrook, Irvine, CA 92604
[2]California State Polytechnic University, Pomona, CA, 91768

ABSTRACT

Social Media Depression Detection is an Intelligent System to automate the detection of Youth Depression with social media (Instagram) using AI and Deep Learning. The student is the targeted group because most students with depression express themselves on social media rather than seeking help from doctors. This app gathers captions and images from the user's personal Instagram profile through web scraping using Instagram private API to check whether or not the posts are depressive. The google cloud dataset supports the captions and pictures analysis performed by the app [6]. Caption sentiment analysis depends on sentiment analysis, and the pictures analysis depends on classifying images by custom labels. The app reports the image and the caption analysis results back to the user. Python is used for the back-end functionality, while Dart and Flutter are used for the front-end. It was tested by 2 experiments, the first experiments returned the feedback of 15 students demonstrates that the program has the capability of detecting depression through the captions with relatively high accuracy. The second experiment of testing the app functionality on the same account demonstrates that the program is stable and consistent. The purpose of the app is to detect depression at an early stage to prevent the condition from worsening.

KEYWORDS

NLP, Web Scraping, Machine Learning.

1. INTRODUCTION

Depression is one of the most frequent and devastating mental illnesses that has a significant social impact. It is common knowledge that mental health is an important aspect of public health. Depression is one of the most common causes of disability in the world. Millions of individuals suffer from depression around the world, according to the results of World Health Surveys (Moussavi et al., 2007) [1]. Recent research has attempted to use social media to detect depression since the patterns of ideas and thoughts represented in posted text and photos can, to some extent, mirror users' mental states [2]. According to WHO estimates from 2014, approximately 20% of children and adolescents have suffered mental illness, with half of these diseases beginning before the age of 14. Furthermore, mental and substance use disorders were responsible for almost 23% of all deaths worldwide. Because of the close interaction that exists between social media platforms and its users, these platforms have come to mirror the users' personal lives on a number of levels [3]. Apart from the severity of mental problems and their impact on one's mental and physical health, societal stigma (e.g., "mental disorders cannot be treated") and discrimination have caused individuals to be ignored by the community and to avoid receiving necessary treatment [3]. Mental illnesses, such as depression, should be

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 277-291, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121223

recognized early on to avoid the condition from worsening and to maximize the chances of patients receiving appropriate treatment.

Some techniques and systems have been proposed to detect depression in social media by asking patients to fill out interactive questionnaires that contain specific questions designed to detect depression, allowing the user to report whether they have depressive symptoms such as sadness or depressed mood, loss of enjoyment, major weight change, insomnia or excessive sleep, and others [4]. These proposals, however, presume that patients with depression are willing to reply to the survey, which is rarely the case in practice because most patients do not want to open their minds to interact with the survey, and some patients do not acknowledge having depression. Another method is to check for physical indicators of depression, such as back discomfort, headaches, limb pain, joint pain, gastrointestinal issues, sleep issues, and so on [4]. About half of those who suffer from depression are never recognized or treated, and not receiving treatment can be fatal [4]. Because doctors are required to physically contact patients in order to assist them find out the root of their symptoms, and because the majority of depression sufferers do not seek care, this method could not be too complex, and it frequently leads to patients' late depression. Patients can also write down their concerns about any specific depressive symptoms or unusual behaviors they are experiencing, as well as their detailed family history from relatives, but they will need medical assistance. This makes it difficult for patients who primarily use social media and refuse to seek medical assistance. The majority of research into how to direct mental illness in social media platforms has centered on feature engineering [3]. Researchers had to identify each mental disease by extracting traits that overlapped. This procedure isn't always accurate, and it's difficult to forecast which mental disease a patient will be diagnosed with.

In this paper, I follow the same line of research by finding the depressive symptoms and implementing it into the project. My goal is to detect depression in Instagram by analyzing the users' posts including texts, videos, and images [7]. My method is inspired by feature engineering, which is the process of transforming raw data into features that better represent the underlying problem to the predictive models [8]. The mobile app I created is an Intelligent System to automate the detection of Youth Depression with social media (Instagram) using AI and Deep Learning. I choose the student to be the targeted group because most students with depression are not likely to seek help and also teenagers use social media a lot. This app gathers captions, images, and videos from the user's personal Instagram profile to check whether or not the posts are depressive. The program utilized sentiment analysis in the text analysis part, google.cloud vision image annotator, image label filter, safe search, and face detection for the image analysis. The results splits into slightly depressive, depressive, and positive. The app also informs the user of the results in order to encourage them to get proper treatments and prevent possibilities of mental health-related problems.

I did two experiments to prove my program. The first experiment is conducted on 15 students who frequently make instagram posts. The feedback received demonstrates that 71% of slightly depressive students, 75% of depressive students, and 100% of positive students were correctly-identified for the caption analysis. For the image analysis, 70% of depressive students and 80% of positive students were correctly-identified. Considering the difficulty and the complexity of finding possible depressive symptoms only through the social media posts, my program that combines the captions and the image analysis proves that it can accurately detect depression in most of the cases. The second experiment was conducted based on the performance of the program 50 times for each of the 2 Instagram accounts: one with completely positive posts, and the other with completely depressive posts. The results received demonstrate that the caption and the image analysis were both 100% consistent. The average time taken for getting the final analysis results of the positive account is slightly greater than the depressive accounts since the positive account doubles the amount of posts of the depressive account. The 100% consistency of

the 50 trials for each account proves that the analysis results generated by the program are trustworthy.

The rest of the paper is organized as follows: Section 2 gives the details on the challenges that we met during the experiment and designing the sample; Section 3 focuses on the details of our solutions corresponding to the challenges that we mentioned in Section 2; Section 4 presents the relevant details about the experiment we did, following by presenting the related work in Section 5. Finally, Section 6 gives the conclusion remarks, as well as pointing out the future work of this project.

2. CHALLENGES

In order to build the tracking system, a few challenges have been identified as follows.

2.1. Collecting Data and Getting the Updated Instagram Posts

Collecting the data required for building the mobile program and getting the updated user posts on instagram are challenging [9]. It's easy to get lots of users' posts from a large dataset, but it doesn't work for the purpose of gathering captions and videos from the updated version of the user's personal Instagram program. It's complicated to scrape the data stored in another social media and implement my program to analyze these data. People want to collect information that ranges anywhere from collecting usernames, finding followers, collecting comments, and analyzing conversations that include the keywords. Also, sensitive data that can be used to manipulate others should not be gathered and used for any purposes. In addition, the data extracted from the Instagram social media platform is challenging due to the unstructured nature of the text posted by users [11]. The posts are introduced with misspelled words, new terms, character limitations, and syntactic errors when composing a message.

2.2. Image Analysis

Since Instagram posts are all published with images, images can better express feelings than words sometimes through the color, darkness, character, and feeling it creates [12]. In order to build the depression detection app, image analysis is required to determine whether the user is positive, slightly depressive, or depressive. Conducting the image analysis is complicated because there's no standard. For example, dark color doesn't mean that the picture is sad and depressing, and bright color also doesn't mean that the picture is positive and joyful. It might be easy for humans to look over the pictures and decide which category that certain picture should be in, but it's complicated for machines to analyze. In addition, pictures also include different elements or symbols that represent a certain thing, which makes it complex for computers. Irony has also been increasingly used in images, which often trick computers to consider the image by completely off or opposite meaning.

2.3. Dart Front-end Development

Since the Instagram Depression Detection is an app that I'm building, I need to develop the front-end, or the client side to make the app user-friendly. I also want to publish the app on both IOS and Android systems so I choose Dart, which is a language that I don't have any experience with [13]. I encountered a lot of trouble during the process of switching pages and building the necessary elements, such as enabling the user to enter username and password, linking the analysis result with the result page, demonstrate the users' pictures and captions along with the

analysis, and changing the settings including fonts, backgrounds, size, page size. I also need to take into consideration different mobile phones' sizes.

3. SOLUTION

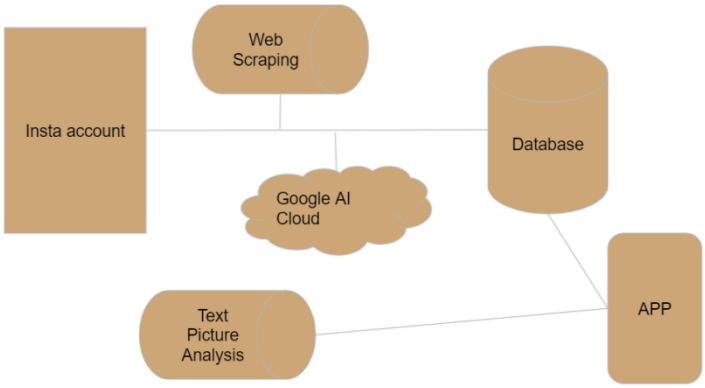


Figure 1. Flow of The Process

DepressionEyes is an Intelligent System to automate the detection of Youth Depression with social media (Instagram) using AI and Deep Learning. As shown in figure 1, DepressionEyes gathers captions and images from the user's personal Instagram profile through web scraping with the help of client in the Instagram private API [6]. Google cloud generates a large database of text sentiment analysis and pictures labels [10]. The captions and pictures analysis performed by the app are supported by google cloud dataset. Caption analysis is dependent on google cloud language_v1 package text sentiment analysis. All text input will receive a score from -1 to 1, -1 being the extreme negative, and 1 being the extreme positive. The pictures analysis is dependent on google cloud vision image label filter, label detection, safe search annotation, and human face annotations. Due to the large dataset contained in google cloud vision, image input will be detected either as depressive or not depressive. The app divided the total number of the images received from the user's Instagram posts by the total number of depressive images detected to find and report whether half of the user's pictures are depressive depending on the percentage of depressive images among all. For the caption analysis, the app finds the average of the sentiment analysis performed on all captions, and splits the results into slightly depressive, depressive, and positive captions. In the front-end app building, I wrote the code in dart and flutter, which I included five pages: about page, info page, login page, main page, and pictures page. The app links the back-end python code with the front-end in the info page, so the results will be demonstrated after the user puts in their instagram username and password.

This part explains how I implemented the Back-end of DepressionEyes written in Python. As shown in figure 2, the instagram private api client class enables the user to pass in their username and password and the program then has the ability to access the user's instagram. The program used the self_feed() function to get authenticated user's own feed [6]. Since the program only needs captions and pictures, it collects information inside 'item' among the user's own feed. The pictures are presented as links, which are stored in the picsLinks list. The captions are stored in the captions list.

```
def getInstaPosts(user_name, password):

  picsLinks = []
  captions = []

  api = Client(user_name, password)
  results = api.self_feed()
  items = results.get('items')
```

Figure 2. Getting User’s Posts

In order to determine whether the input captions are depressive, the program included a textAnalysis function defined with a parameter text, which received the passed-in captions. The text sentiment analysis is dependent on the google cloud language_v1 package [7]. The program first instantiates a client through the LanguageServiceClient() function, then passes the caption input to the Document() function. The program detects the sentiment of the captions using the analyze_sentiment() function and the document_sentiment. document_sentiment contains the overall sentiment of the document, which consists of score and magnitude. The sentiment score runs from -1.0 (negative) to 1.0 (positive) and refers to the text's total emotional learning. The document's magnitude reveals how much emotional content it contains. Magnitude is significant when analyzing documents with a neutral score since it allows one to determine how much emotional content is present. The textAnalysis(text) function returns the sentiment score multiplied by the magnitude as a result.

```
def textAnalysis(text):
  # Instantiates a client
  client = language_v1.LanguageServiceClient()

  # The text to analyze
  document = language_v1.Document(content=text,
                                  type_=language_v1.Document.Type.PLAIN_TEXT)

  # Detects the sentiment of the text
  sentiment = client.analyze_sentiment(request={
      'document': document
  }).document_sentiment

  return (sentiment.score * sentiment.magnitude)
```

Figure 3. Text Analysis Function

The program finds the scores of the captions stored in the captions list. As shown in figure 4, the captions average is either identified as slightly depressive, depressive, and positive. Average score above 0.0 is considered positive, average below -0.25 is considered negative, and average between -0.25 and 0.0 is considered slightly depressive.

```
  for text in captions:
    textAverage += textAnalysis(text)

  textAverage /= len(captions)

  if textAverage < 0 and textAverage > -.25:
    captionsResults = "Your posts are slightly depressive"
  elif textAverage <= -.25:
    captionsResults = "Your posts are depressive"
  else:
    captionsResults = "Your posts are positive!"

  toJson(imagesResult, captionsResults, picsLinks, captions)
```

Figure 4. Caption Results Supported by The Text Analysis Function

DepressionEyes utilized google cloud vision AI to conduct the images analysis [8]. With AutoML Vision, Vision AI employs machine learning to analyze images with industry-leading prediction accuracy, train machine learning models that classify images by custom labels, and detect objects and faces, read handwriting, and construct important image metadata [8]. After instantiating a client and loading the image into memory, the program performs label detection on the image file. It stored the labels, safe search results, and the face features results of the images in different variables as shown in figure 5.

```
def imageAnalysis(picture):

  imageLabelFilter = ["Darkness", "Fearful"]

  # Instantiates a client
  client = vision.ImageAnnotatorClient()

  # The name of the image file to annotate
  file_name = os.path.abspath(picture)

  # Loads the image into memory
  with io.open(file_name, 'rb') as image_file:
    content = image_file.read()

  image = vision.Image(content=content)

  # Performs label detection on the image file
  response1 = client.label_detection(image=image)
  labels = response1.label_annotations

  response2 = client.safe_search_detection(image=image)
  safe = response2.safe_search_annotation

  response3 = client.face_detection(image=image)
  faces = response3.face_annotations

  # Names of likelihood from google.cloud.vision.enums
  likelihood_name = ('UNKNOWN', 'VERY_UNLIKELY', 'UNLIKELY', 'POSSIBLE',
                     'LIKELY', 'VERY_LIKELY')
```

Figure 5. Label Detection in Image Analysis

If the labels in the image include “darkness” and “fearful” defined in figure 5, the picture is considered as depressive. If the safe violence presented in the picture, it’s considered depressive. Considering the face features in case some posts included faces, faces with anger and sorrow are considered depressive, and faces with joy are considered positive as shown in figure 6.

```
for label in labels:
  if label.description in imageLabelFilter:
    return True

if likelihood_name[safe.violence] == 'POSSIBLE' or likelihood_name[
    safe.violence] == 'LIKELY' or likelihood_name[
        safe.violence] == 'VERY_LIKELY':
  return True

else:
  return False

if faces:
  if likelihood_name[
      faces.joy_likelihood] == 'POSSIBLE' or likelihood_name[
          faces.joy_likelihood] == 'LIKELY' or likelihood_name[
              faces.joy_likelihood] == 'VERY_LIKELY':
    return False

  elif likelihood_name[
      faces.anger_likelihood] == 'POSSIBLE' or likelihood_name[
          faces.anger_likelihood] == 'LIKELY' or likelihood_name[
              faces.anger_likelihood] == 'VERY_LIKELY':
    return True

  elif likelihood_name[
      faces.sorrow_likelihood] == 'POSSIBLE' or likelihood_name[
          faces.sorrow_likelihood] == 'LIKELY' or likelihood_name[
              faces.sorrow_likelihood] == 'VERY_LIKELY':
    return True

  else:
    return False
```

Figure 6. Categorizing Different Images

In figure 7, the program finds the depression percentages by dividing the total images with the depressive images. It reports whether or not half of the images are depressive.

```
os.chdir(origDir)
for image in os.listdir("images/"):
  result = imageAnalysis("images/" + image)
  if result == True:
    numTrue += 1
  #os.remove("frames/" + image)
if numTrue >= 1:
  percentDepression = imgCount / numTrue
else:
  percentDepression = 3

if percentDepression <= 2:
  imagesResult = "At least half of your images are depressive"
else:
  imagesResult = "Less then half of your images are depressive"
```

Figure 7. Images Results Supported by The Images Analysis Function

I chose flutter, an open-source UI software development kit created by Google, for the front-end development of my program because flutter can be used to develop cross platform applications for iOS, Android, Linux, Mac, Windox, etc. There are five pages in the programs, the main page that leads the users to the program, the login page that enables the users to put in their Instagram username and password, about page that give instructions on how the program works to the users, info page that demonstrates the caption and image analysis of the users' instagram results, and pictures page that demonstrates all pictures in users' instagram posts. Considering switching between screens, I used a gesture detector to navigate to another page through the sensitivity of the hit test behavior as shown in figure 8.

```
child: GestureDetector(
  behavior: HitTestBehavior.translucent,
  onHorizontalDragUpdate: (details) {
  // Note: Sensitivity is integer used when you don't want to mess up vertical drag
  if (details.delta.dx > sensitivity) {
    Navigator.push(
      context,
      MaterialPageRoute(builder: (context) => LoginPage()),
    );
  }
  else if(details.delta.dx < -sensitivity){
    Navigator.push(
      context,
      MaterialPageRoute(builder: (context) => LoginPage()),
    );
  }
```

Figure 8. Swipe Picture from Main Page to Login Page

On the login page, I used the input decoration, the outline input border, and the ChakraPetch fonts inside the input box as shown in figure 9. The first box on the page is asking for the username, and the box below is asking for the password. A button to the about page is also on the login page.

```
Padding(
  padding: EdgeInsets.fromLTRB(20, 15, 20, 20),
  child: TextField(
    style: TextStyle(color: Colors.white),
    onChanged: (v)=>setState((){username=v;}),
    decoration: InputDecoration(
      focusedBorder: OutlineInputBorder(
        borderRadius: BorderRadius.all(Radius.circular(4)),
        borderSide: BorderSide(color: Colors.white),
      ),  // OutlineInputBorder
      enabledBorder: OutlineInputBorder(
        borderRadius: BorderRadius.all(Radius.circular(4)),
        borderSide: BorderSide(color: Colors.black),
      ),  // OutlineInputBorder
      labelText: 'Username',
      labelStyle: TextStyle(fontFamily: 'ChakraPetch', color: Colors.deepPurple),
    ),  // InputDecoration
  ),  // TextField
),  // Padding
```

Figure 9. Login Page Input Box

Figure 10 demonstrates the first three screens of the application. The first one from the left is the main page, the second one from the left is the login page, and the third one from the left is the about page. The theme is purple, which remained consistent throughout the application.

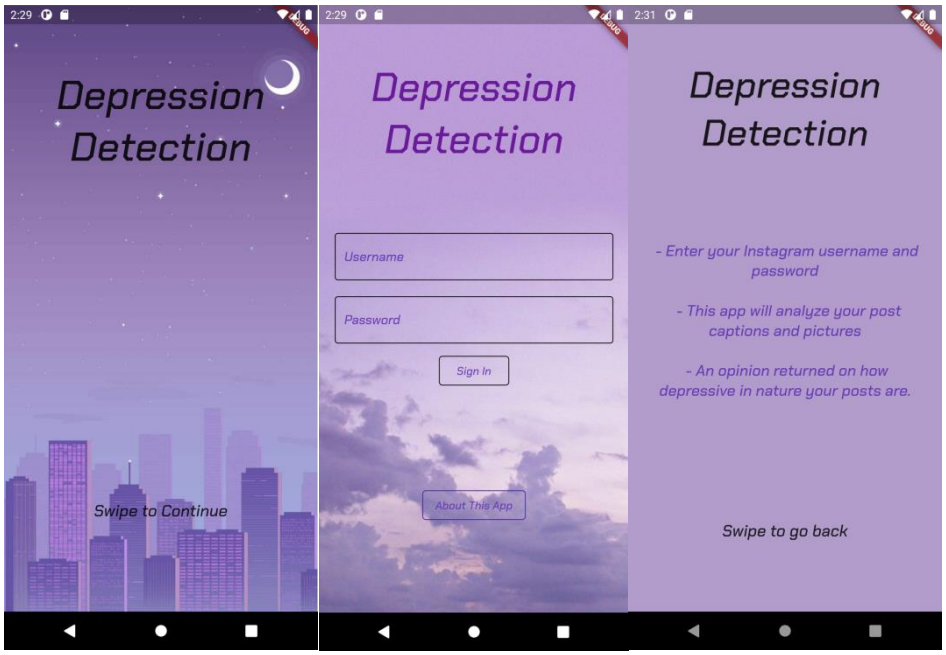


Figure 10. Demonstration of Main Page, Login Page, and About Page

As shown in figure 11, I linked my back-end Python program with flutter in the Info Page by making a get request to my program on repl.it with the username and the password passed.

```
_makeGetRequest(username, password) async {
  var url = 'https://depression2.aliviading.repl.co/results/$username/$password';
  var response = await http.get(url);
  print(response.body);
  return json.decode(response.body);
}
```

Figure 11. Info Page Linking to Back-end

The info page also serves the purpose of demonstrating the caption and the picture analysis results. As shown in figure 12, I received the caption results from the “Captions Sentiment” element of the analysis, and demonstrated the results in ChakraPetch 20 fonts. The pictures results demonstrate in similar ways.

```
FutureBuilder(
  future: _makeGetRequest(widget.username, widget.password),
  builder: (context, snapshot) {
    if(snapshot.hasData) {
      return Container(
        child: Column(
          children: <Widget>[
          Text('Captions Result: ',
            style: TextStyle(fontSize: 20, fontFamily: 'ChakraPetch', fontWeight: FontWeight.bold),
            textAlign: TextAlign.center,
          ),  // Text
          Text('${snapshot.data["Captions Sentiment"]}',
            style: TextStyle(fontSize: 20, fontFamily: 'ChakraPetch'),
            textAlign: TextAlign.center,
          ),  // Text
```

Figure 12. Info Page Elements

Figure 13 is a demonstration of the info pages. The one on the left is tested by a depressive instagram account, so the caption and picture results all show that the Instagram account is depressive. The one on the right is tested by a positive instagram account, so the caption and picture results all show that this instagram account is positive.

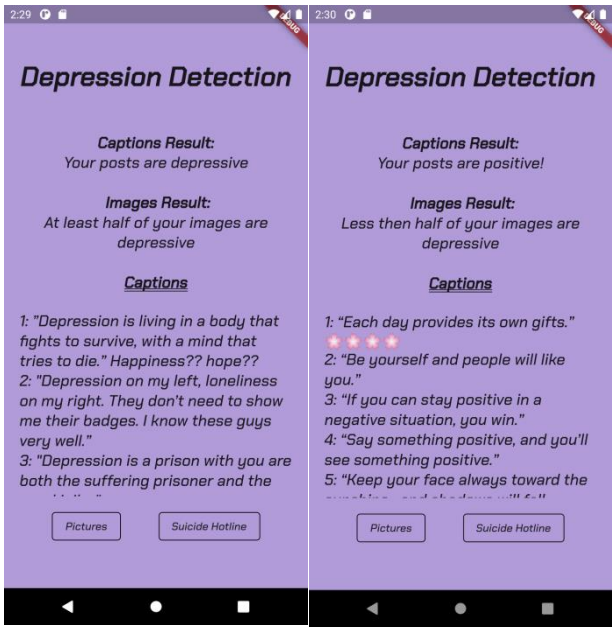


Figure 13. Depressive and Positive Info Page Demonstrations

As shown in figure 13, there's a picture button linking to the picture page that displays all pictures from the user posts. The program converted the pictures to links, and I stored and displayed the link in the pictures array as shown in figure 14. The program also have a "Suicide Hotline" button on the info page, which linked to https://suicidepreventionlifeline.org/ in case the users need help.

```
picturesList(info) {

  List<String> pictures = [];

  for (int i=0; i < info["Pics Links"].length; i++) {
    pictures.add(info["Pics Links"][i]);
  }

  return pictures;
}
```

Figure 14. Getting The Pictures

Figure 15 demonstrates the pictures page of the same positive and depressive instagram accounts the program used for the Info page demonstration. The one on the left has depressive Instagram posts and the one on the left has positive instagram posts. The picture pages listed all pictures in the posts.

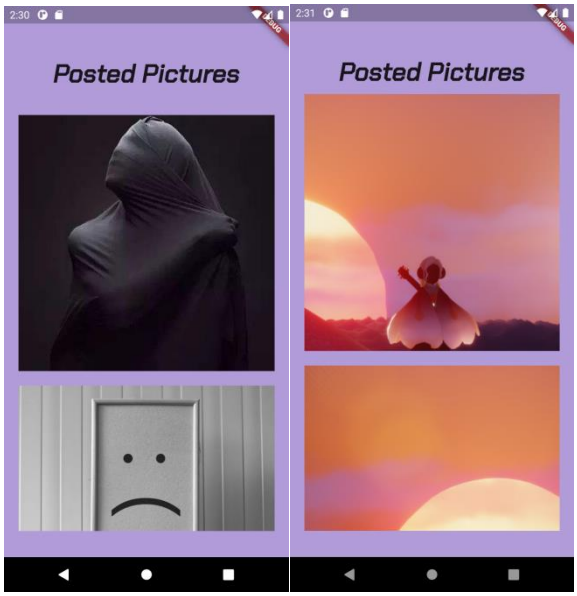


Figure 15. Depressive and Positive Picture Pages Demonstrations

4. EXPERIMENT

Experiment 1

The program served the purpose of detecting depression among the youth generation through their social media posts. In order to test the efficiency and the accuracy of the program, I conducted an experiment on 15 students that frequently used Instagram. They put in their username and password to the program and check if the captions and the image analysis presented match what they think. The purpose of the experiments is to find out the percentage of students that's detected currently.

Table 1 demonstrates the accuracy of the text sentiment analysis. Rows represented the number of students who considered themselves as either depressive, slightly depressive, or positive. Columns representing the number of students are being correctly-identified, wrongly identified, and calculated the match percentage. For each row, the correctly-identified and the wrongly-identified students add up to be the total number of students that considered themselves as the corresponding category.

Table1. Text Sentiment Analysis Results

	Correctly-identified	Wrongly-identified	Percentage match
Depressive	3	1	75%
Slightly Depressive	5	2	71%
Positive	4	0	100%

Table 2 demonstrates the accuracy of the image sentiment analysis. Rows represented the number of students who considered themselves as either depressive or positive. Columns representing the number of students that are being correctly-identified, wrongly identified, and the percentage match shows. The numbers of correctly and wrongly identified added up the total students of each corresponding category.

Table 2. Image Analysis Results

	Correctly-identified	Wrongly-identified	Percentage match
Depressive	7	3	70%
Positive	4	1	80%

Experiment 2

In order to test the consistency and also the accuracy of the program, the second experiment tested the program performance on two Instagram accounts, each for 50 trials. One of the accounts is intended to be depressive, with depressive captions (i.e., depressive quotes) and depressive images, and a total of five posts. The other account is designed to be positive that includes positive contents with a total of 10 posts. This experiment measured and recorded the caption and image analysis result along with the time(sec) takes for the analysis to process. The purpose of the experiments is to find out how accurate and consistent the program is.

The caption analysis and picture analysis results for all 50 trails for the positive account and the depressive account are the same, as shown in table 3. Therefore, the percentage match for both is 100 percent. The average time spent for getting a result is 16.5 seconds for the positive account, and 14.25 seconds for the depressive account. The depressive account has less posts than the positive accounts, and the time taken is more consistent than the positive account with more posts.

Table 3. Consistency Tests on Two Accounts with either Positive or Negative Posts

	consistency of caption analysis	consistency of image analysis	percentage match (accuracy for caption analysis)	percentage match (accuracy for image analysis)	Average time spent for result (caption analysis + image analysis)
positive	100% same	100% same	100%	100%	16.5 sec
depressive	100% same	100% same	100%	100%	14.25 sec

The first experiment demonstrates that 71% of slightly depressive students, 75% of depressive students, and 100% of positive students were correctly-identified for the caption analysis. The slightly depressive students have the lowest percentage because their posts are more neutral than students who are strongly depressive or positive. This proves that the program has the capability of detecting depression through the captions. It has the ability to improve through training the model by captions that are not extremely negative or positive. For the image analysis, 70% of depressive students and 80% of positive students were correctly-identified. This shows that the image analysis has lower accuracy than the caption analysis, but it can still detect most of the depressive pictures. Combining the captions analysis and the image analysis, the program proves that it can accurately detect depression in most of the cases.

The second experiment demonstrates that the caption and the image analysis were both 100% consistent. The average time spent for the analysis results including the caption and image analysis for the positive account is 16.5 seconds, and the average time for the depressive account is 14.25 seconds. Considering that the positive account doubles the amount of posts of the depressive account, longer time is reasonable. Also, the 100% consistency proves that this program is trustworthy.

5. RELATED WORK

Many well regarded scales and criteria have been developed based on the user research or questionnaire survey [5]. Beck's Depression Inventory [Beck et al., 1961], for example, consists of 21 questions about the mental and physiological status of the user. Another example is the CES-D Scale [Radloff, 1977], which includes 20 items about mental health issues such as users' guilt sentiments and sleep patterns. My program directly analyzes the posts in social media instead of asking the users' survey questions. One of the strengths of my work is that detection can be made even when the patient refuses to answer the questions.

With the passage of time, we, including depressed users, are nearly unable to live without social media. As a result, researchers began to study the online behaviors of depressed individuals. Park et al. [2012] used real-time moods acquired from Twitter users to investigate the usage of language in portraying depressive moods as a preliminary study. Park et al. [2013] performed face-to-face interviews with 14 active Twitter users to investigate their depressive behaviors in their follow-up study. My program also used social media as a platform to detect depression. My program performed a text analysis on the user's caption. In addition to that, my program has image analysis which enables posts with only images to be tested. The benefit of their program is that they conducted interviews with active Twitter users to explore their depressive behaviors, so they know how people with depression act in social media.

Another study (Wang et al., 2013) presented a method for detecting depression in microblogs based on subject-dependent sentiment analysis. They think that a lack of good emotions and a lack of negative emotions are essential markers of depression. They used ten psychologically-based features of depressed users: the number of emoticons, interaction features (how users connect with one another), and behavior features (frequencies of the posting, active period). My program did similar things to detect depression, such as the sentiment analysis of the posts. One difference is that I only considered text/pictures depressive when it contains negative contents, Wang's approach classifies microblog as depressive when there's lack of positive emotions. This may help prohibit the omission of anyone with depression, but it might not be specific enough.

6. CONCLUSIONS

Social Media Depression Detection is an Intelligent System that uses AI and Deep Learning to automate the detection of Youth Depression on social media (Instagram). Because most students with depression express themselves on social media rather than seeking medical treatment, they are the target group. This software uses web scraping and the Instagram private API to collect captions and images from a user's personal Instagram profile in order to determine whether or not the postings are depressing. The descriptions and image analysis performed by the app are supported by the Google cloud dataset. Sentiment analysis is required for caption sentiment analysis, while image classification by custom labels is required for image analysis. The image and caption analysis findings are returned to the user via the app. The backend is written in Python, while the front-end is written in Dart and Flutter [14]. It was put to the test in two ways. The first trial yielded input from 15 students, demonstrating that the algorithm is capable of identifying depression through captions with a high degree of accuracy. The app's operation was tested again on the same account, demonstrating that the software is stable and consistent.

For the image analysis, 70% of depressive students and 80% of positive students were correctly-identified. This shows that the image analysis has lower accuracy than the caption analysis. The pictures that the user posts in Instagram is important to the result, so the program needs to improve the image analysis accuracy. Also, this program can only be conducted on the user's own Instagram, because it requires username and password. It would be a lot better if it can be used to test other public accounts. An extension that contains these functionality and can be used to test all public accounts is more ideal. The look of the pages can also be more attractive.

In order to improve the accuracy of the image analysis, I will try more libraries or use the library to scratch my own algorithm [15]. For example, I will try the openCV package with python, tensorflow with vuforia, and enhance the way it classifies something as depressive. I will make the mobile app function the same on the computer and potentially make it an extension.

REFERENCES

[1] Stankevich, Maxim, et al. "Feature Engineering for Depression Detection in Social Media." ICPRAM. 2018.
[2] Lin, Chenhao, et al. "Sensemood: Depression detection on social media." Proceedings of the 2020 International Conference on Multimedia Retrieval. 2020.
[3] Orabi, Ahmed Husseini, et al. "Deep learning for depression detection of twitter users." Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018.
[4] Beck, Aaron T. "The diagnosis and management of depression." (1973).
[5] Shen, Guangyao, et al. "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution." IJCAI. 2017.
[6] Garraghan, Peter, Paul Townend, and Jie Xu. "An analysis of the server characteristics and resource utilization in google cloud." 2013 IEEE International Conference on Cloud Engineering (IC2E). IEEE, 2013.
[7] Instagram, Instagram. "Instagram." Facebook, https://www. instagram. com (2016).
[8] Turner, C. Reid, et al. "A conceptual basis for feature engineering." Journal of Systems and Software 49.1 (1999): 3-15.
[9] Coombs, Clyde H. "A theory of data." (1964).
[10] Data, Modeling Historical, and Guide User. Database Design. Perancangan Basis Data) merupakan salah satu, 2015.
[11] Bogdanov, Evgeny, et al. "A social media platform in higher education." Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON). IEEE, 2012.
[12] Koenderink, Jan J. "The structure of images." Biological cybernetics 50.5 (1984): 363-370.

[13] Al-Qershi, Fattoh, et al. "Android vs. iOS: The security battle." 2014 World Congress on Computer Applications and Information Systems (WCCAIS). IEEE, 2014.
[14] De Smedt, Tom, and Walter Daelemans. "Pattern for python." The Journal of Machine Learning Research 13.1 (2012): 2063-2067.
[15] Freese, Frank. "Testing accuracy." Forest Science 6.2 (1960): 139-45.

AUTHOR INDEX