Web, Internet Engineering & Signal Processing

`

David C. Wyld,
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

- 12th International Conference on Internet Engineering & Web Services (InWeS 2021), December 11~12, 2021, Chennai, India
- 7th International Conference on VLSI and Applications (VLSIA 2021)
- 8th International Conference on Signal Processing (CSIP 2021)
- 2nd International Conference on Big Data, Machine Learning and IoT (BMLI 2021)

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: 978-1-925953-55-8
DOI: 10.5121/csit.2021.112101 - 10.5121/csit.2021.112107

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

`

Preface

The International Conference on 12th International Conference on Internet Engineering & Web Services (InWeS 2021), December11~12, 2021, , Chennai, India, 7th International Conference on VLSI and Applications (VLSIA 2021), 8th International Conference on Signal Processing (CSIP 2021), 2nd International Conference on Big Data, Machine Learning and IoT (BMLI 2021) was collocated with 7th International Conference on VLSI and Applications (VLSIA 2021). 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 INWES 2021, VLSIA 2021, CSIP 2021 and BMLI 2021 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, INWES 2021, VLSIA 2021, CSIP 2021 and BMLI 2021 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 INWES 2021, VLSIA 2021, CSIP 2021 and BMLI 2021.

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

A. Parthiban,	Lovely Professional University, India
Abd El-Aziz Ahmed,	Anna’s University, Egypt
Abdel-Badeeh M. Salem,	Ain Shams University, Egypt
Abderrahmane Ez-Zahout,	Mohammed V University, Morocco
Abel Gomes,	University of Beira Interior, Portugal
Abhishek Das,	Aliah University, India
Abid Ismail,	University of Kashmir, India
Addisson Salazar,	Polytechnic University of Valencia, Spain
Afaq Ahmad,	Sultan Qaboos University, Oman
Ahmed Farouk AbdelGawad,	Zagazig University, Egypt
Akhil Gupta,	Lovely Professional University, India
Alberto Taboada-Crispi,	UCLV, Cuba
Ali Abid D. Al-Zuky,	Mustansiriyah University, Iraq
Ali Khenchaf,	Lab-STICC, ENSTA Bretagne, France
Amandeep Singh,	NIT Jalandhar, India
Anand Nayyar,	Duy Tan University, Vietnam
Anouar Abtoy,	Abdelmalek Essaâdi University, Morocco
Anurag Sharma,	GNA University, Phagwara
Aridj Mohamed Hassiba Benbouali,	University Chlef, Algeria
Arman Roohi,	University of Central Florida, FL, USA
Asadollah Shahbahrami,	University of Guilan, Iran
Atena Abdi,	Amirkabir University of Technology, Iran
Atul Garg,	Chitkara University, India
Azeddine Wahbi,	Hassan II University, Casablanca, Morocco
Azeddine WAHBI,	University Hassan II, Morocco
B D C N Prasad,	V R Siddhartha Engineering college , India
Barbaros Preveze,	Cankaya University, Turkey
Bouchra Marzak,	Hassan II University, Morocco
Boukari nasim,	skikda university, Algeria
Brahim Lejdel,	University of EL-Oued, Algeria
Carol Anne Hargreaves,	National University of Singapore, Singapore
Cheng Siong Chin,	Newcastle University, Singapore
Christoph Reich,	Cloud Computing and IT Security, Germany
Daniel Hunyadi,	University of Sibiu, Romania
Daniel Imanah,	CEO- Zudan Electric Ltd, Nigeria
Dário Ferreira,	University of Beira Interior, Portugal
Ekbal Rashid,	Cambridge Institute of Technology, India
El Murabet Amina,	Abdelmalek Essaadi University, Morocco
Elżbieta Macioszek,	Silesian University of Technology, Poland
Eng Islam Atef,	alexandria university, Egypt
Fairouz Fakhfakh,	University of Sfax, Tunisia
Farhad Soleimanian Gharehchopogh,	Islamic Azad University, Iran
Farouq Saber Al-Shibli,	Philadelphia University, Jordan

`

Felix J. Garcia Clemente, University of Murcia, Spain
Fernando Bobillo, University of Zaragoza, Spain
Francesco Zirilli, Sapienza Universita Roma, Italy
Gayatri Mehta, University of North Texas, USA
Govindraj Chittapur, Basaveshwar Engineering College, India
Grigorios N. Beligiannis, University of Patras, Greece
Grzegorz Sierpiński, Silesian University of Technology, Poland
Gulden Kokturk, Dokuz Eylul University, Turkey
Hamid Ali Abed AL-Asadi, Iraq University College, Iraq
Hamida Souraya, University of Batna, Algeria
Hedayat Omidvar, National Iranian Gas Company, Iran
Hlaing Htake Khaung Tin, University of Computer Studies, Myanmar
Ihab Zaqout, Azhar University, Palestine
Intisar Al-Mejibli, University of Essex, United Kingdom
Israa Shaker Tawfic, Ministry of Science and Technology, Iraq
Jagadeesh HS, APSCE (VTU), India
Jan Ochodnicky, Armed Forces Academy, Slovakia
Janusz Kacprzyk, Systems Research Institute, Poland
Jawad K. Ali, University of Technology, Iraq
Jorge Bernardino, Polytechnic of Coimbra, Portugal
Julie M David, MES College Marampally Alua, India
K Lal Kishore, JNTUH, India
Kamaraju M, Gudlavalleru Engineering College, India
Karim Mansour, Salah Boubenider University, Algeria
Keivan Navi, Shahid Beheshti University, Iran
Khaldi Amine, Kasdi Merbah University, Algeria
Krishna Maya Manger, University of North Bengal, India
Luisa Maria Arvide Cambra, University of Almeria, Spain
M Padmaja, VR Siddhartha Engg College, India
M V Ramana Murthy, Osmania University, India
M.K.Marichelvam, Mepco Schlenk Engineering College, India
Mahmoud Shafik, University of Derby, UK
Manish Kumar Mishra, University of the People, USA
Manoj Kumar, Vidya College of Engineering, India
Maria Hallo, Escuela Politecnica Nacional, Ecuador
Martin Lukac, Nazarbayev University, Kazakhstan
Maryam Rastgarpour, Islamic Azad University, Iran
Masoud Ziabari, Mehr Aeen University, Iran
Michail Kalogiannakis, University of Crete, Greece
Mohamed Elhoseny, Mansoura University, Egypt
Mohammed Atif, Arab Open University, Kuwait
Mohankumar N, Amrita Vishwa Vidyapeetham, India
Muhammad Elrabaa, KFUPM, Saudi Arabia
Muzhir Shaban Al-Ani, University of Human Development, Iraq
N. Jeyanthi, VIT University, India
Nadia Abd-Alsabour, Cairo university, Egypt
Navaid Z. Rizvi, Gautam Buddha University, India
Nazmus Saquib, University of Manitoba, Nazmus
Nihar Athreyas, University of Massachusetts, USA
Omid Mahdi Ebadati, Kharazmi University, Tehran
Parthiban, Lovely Professional University, India
Paulo Maciel, Federal University of Pernambuco, Brazil

`

Periola Ayodele,	Bells University of Technology, Nigeria
Philippe Merle,	Inria, France
Popa Rustem,	University of Galati, Romania
Pr. Smain Femmam,	UHA University, France
Rahul Pachauri,	Awadesh Pratap Singh University, India
Rajkumar,	N.M.S.S.Vellaichamy Nadar College, India
Rakesh Kumar Mahendran,	Sakunthala Engineering College, India
Ramana Murthy,	Osmania University, India
Ramtin Mohammadi-Zand,	University of Central Florida, USA
Rana Mukherji,	The ICFAI University,India
Rita Li,	Hong Kong Shue Yan University, China
Rodrigo Campos Bortoletto,	São Paulo Federal Institute, Brazil
Ruksar Fatima,	Khaja Bandanawaz University, Kalaburagi
Saad Ali Amin,	University of Dubai, Dubai
Sadeque Reza Khan,	National Institute of Technology, India
Sai Kumar T,	CMR Technical Campus, India
Seyyed Reza Khaze,	Islamic Azad University, Iran
Shahram Babaie,	Islamic Azad University, Iran
Shilpa Joshi,	University of Mumbai, India
Shing-Tai Pan,	National University of Kaohsiung, Taiwan
Shonkh Shuvro,	IIEST, India
Sidnei José Buso,	INEP, Brazil
Smain Femmam,	UHA University France, France
Smaranda Belciug,	University of Craiova, Romania
Subhankar Mukherjee,	Samsung India Software Operations, India
Subhendu Kumar Pani,	Krupajal Computer Academy, India
Sudipta Kr Ghosal,	Nalhati Government Polytechnic, India
Sujatha,	Vellore Institute of Technology, India
Tapalina Bhattasali,	St. Xavier's College (Autonomous), India
Tarig Mohamed Ahmed,	Universty of Khartoum, Sudan
Vahideh Hayyolalam,	Koç University, Turkey
Valli kumari,	Andhra University, India
Vimal,	Ramco Institute of Technology, India
Walid Barhoumi,	LIMTIC Laboratory, Tunisia
Wenwu Wang,	University of Surrey,United Kingdom
Xiao-Zhi Gao,	University of Eastern Finland, Finland
Yew Kee Wong,	HuangHuai University, China
Youye Xie,	Colorado School of Mines, USA
Yuan-Kai Wang,	Fu Jen Catholic University, Taiwan
Yu-Chen Hu,	Providence University, Taiwan
Zhifeng Wang,	Florida State University, USA
Zhihao Wu,	Shanghai Jiao Tong University, China
Zoran Bojkovic,	University of Belgrade, Serbia

`

Technically Sponsored by

Computer Science & Information Technology Community (CSITC)

Artificial Intelligence Community (AIC)

Soft Computing Community (SCC)

Digital Signal & Image Processing Community (DSIPC)

`

7th International Conference on VLSI and Applications (VLSIA 2021)

12th International Conference on Internet Engineering & Web Services (InWeS 2021)

8th International Conference on Signal Processing (CSIP 2021)

2nd International Conference on Big Data, Machine Learning and IoT (BMLI 2021)

SMALL DELAY TRACING DEFECT TESTING

Lakshmaiah Alluri[1] and Hemant Jeevan Magadum[2]

[1]HDG, CDAC, Thiruvananthapuram, India
[2]ITNS, CDAC, Thiruvananthapuram, India

ABSTRACT

This Small Delay Tracing Defect Testing detect small delay defects by creating internal signal races. The races are created by launching transitions along simultaneous two paths, a reference path and a test path. The arrival times of the transitions on a 'convergence' or common gate determine the result of the race. On the output of the convergence gate, a static hazard created by a small delay defect presence on the test path which is directed to the input of a scan-latch. A glitch detector is added to the scan latch which records the presence or absence of the glitch.

KEYWORDS

Delay, Defect, detector, glitch, testing.

1. INTRODUCTION

The traditional delay defect detection techniques having their advantages and disadvantages. In structured test approach, ATPG techniques used to generate input test vector sequences to target classified faults in the circuit. Typically ATPG works in three main phases which are excite the target fault, propagate the fault effect to the observation point through an identified path and justify the values of off-inputs without causing a contradiction. To achieve a desired test coverage, a set of patterns generated by ATPG. The percentage of total number of faults being detected using the generated pattern set out of initial target fault list is called as Test coverage. ATPG targets the faults in two main steps which are generating fault-oriented test patterns and performing fault simulation to determine a list of faults being detected using the generated test patterns. The progressive ATPG tools change these 2 steps into one operation whereas pattern generation. The ATPG generates the test patterns and enables the ATE in order to load pattern data into a chip's scan cells. In order to achieve 100% controllability. Next, these flip-flops are stitched into a single chain or multiple chains based on the number of pads available on the device to perform test. The stuck-at fault model is assumed to be the most common fault model while performing fault simulation. It is being used because of its effectiveness in finding many common defect types. The physical struck at defects can be traced the struck at model capture the struck at zero and struck at one faults by implementing the traditional methods such as Transition delay fault (TDF) , Path-based and Segment delay fault based on ATPG. As late as possible transition fault (ALAPTF) based ATP, N-detect transition fault based ATPG, Timing-aware ATPG.

2. NEED OF SMALL DELAY DEFECT TESTING

This method is for the detection of small delay defects that minimizes the number of delay tests that need to be applied at faster than the rated clock speed. The method requires the longest path to every primary output and scan-latch input (endpoint) to be validated using delay tests or a

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 01-06, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112101

reference path test structure. These longest paths are the reference paths. Once these paths are validated, they are used to refer upper bound delay paths. This process detects small delay defects that occur on the reference paths and on the path segments driving the off-path inputs. The process of testing the off-path segments for little delay defects involves simultaneously propagating signals along both the off-path segment and reference path segment. The gate that serves as the endpoint of both path segments is called the convergence gate. The off-path segment features a smaller delay or a delay adequate to the reference segment delay in normal condition. Otherwise, the reference path is not the longest path, and its transition arrives at the convergence gate before the reference segment signal. If the off-path segment features a small delay defect, the other may occur. The transitions on the inputs to the convergence gate are controlled such that if later a glitch is produced on the output of the convergence gate. A glitch detector, placed at some extent beyond the convergence gate, is employed to capture the glitch, thereby recording the results of the signal race. If a small delay fault occurred, the glitch detector value is scanned out for inspection to determine.

3. ARCHITECTURE OF SMALL DELAY DEFECT TESTING

It makes use of internal signal races along path segments as a way of upper bounding the propagation delay of a test path segment against the delay of reference path segment. So, the upper bounds on the delay of a set of reference paths are determined by using a reference path test structure or by applying standard delay tests. The longest path to every endpoint (scan latch input or primary output) is chosen because the reference path. Figure 2 shows a reference path with multiple endpoints. The longer path to endpoint D is labeled pr for path reference. A transition is launched from a PI or scan-latch A and is captured at endpoint D. If the signal propagates to D within the launch-capture cycle time, then its delay is upper bounded by that point. This process confirms that the reference path does not have a delay defect. The same process is applied to the remaining reference paths. The validated reference paths are then used to bound the delays of other, shorter paths in the circuit. Figure 3 illustrates how this is accomplished. Two path segments are identified as sr (for segment reference) and st (for segment test). Below Figure 1 shows automation flow of methodology.

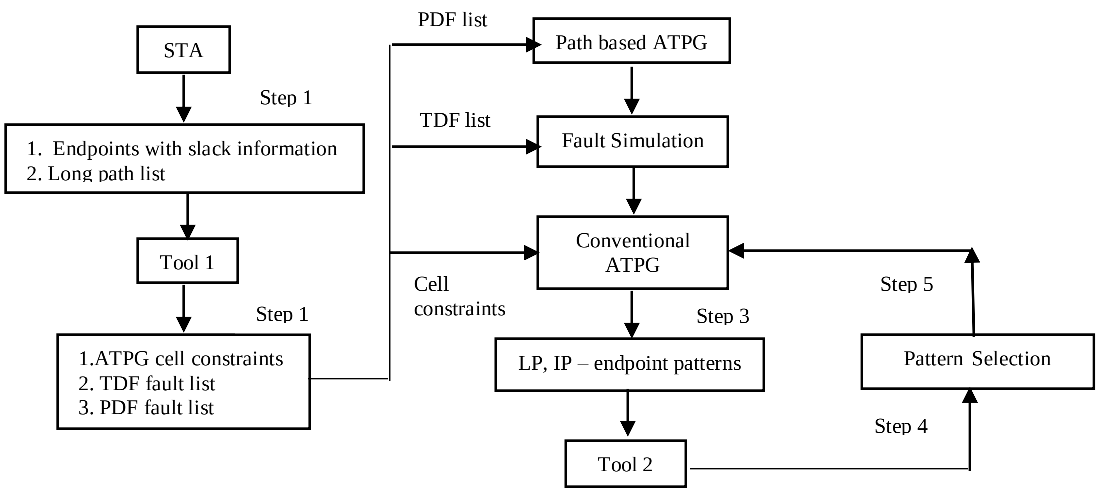


Figure 1. Automation Flow of Methodology

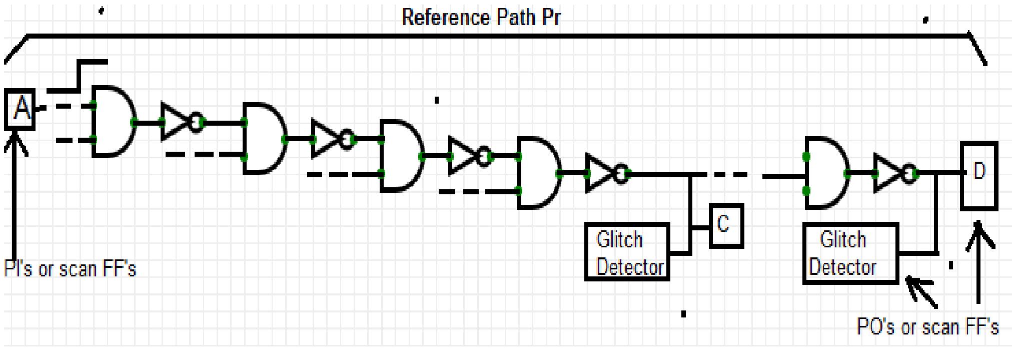


Figure 2. Reference path is longest path to end point

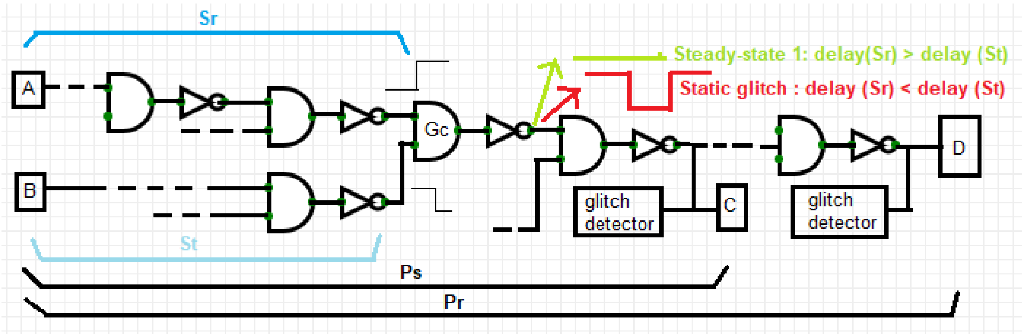


Figure 3. A signal race is created between reference path segment and test path segment driving an off path input to convergence gate.

It launch points labelled A and B, respectively. The endpoints of the segments drive the inputs of a convergence gate, GC. Since the reference path is the longest path to endpoint D, the delay of sr, delay (sr) is greater than or equal to delay(st) by design. The transitions shown at the inputs of GC, i.e., 0 → 1 for sr and 1 → 0 for st cause the output of GC to behave in one of two ways. If delay (sr) >= delay (st) then the 0 along st arrives before the 1 on sr, and the output remains steady-state high. If the opposite is true, i.e., delay(sr) < delay(st), then GC's output switches momentarily with duration proportional to the difference in delays along the two segments.

The relative timing of the two segments is reflected in the output behaviour of GC. One way to record the output behaviour of GC for subsequent inspection is to monitor the state of the path segment between the convergence gate and an endpoint using a glitch detector. If a transition occurs on its input, the glitch detector is designed with a memory element that flips state. Figures 1 and 2 show two glitch detectors at endpoints C and D. For this test, it is possible to use the glitch detector at endpoint D, selecting an endpoint that is closest to the convergence gate. The glitch detector at endpoint C is better for several reasons. First, differences in pull-up and pull-down strengths of gates along a path can compress the width of the glitch (and even eliminate it), making it more difficult or impossible to detect. Second, hazards produced on off-path inputs between the convergence gate and the endpoint may invalidate the result. So keeping this segment small helps minimize these effects. The example shows the reference segment input to GC changing from the dominate value (0) to the non-dominate value (1), and the test segment input transitioning in the opposite direction. Reversing these transitions allows the relative segment delay of the opposite transition along these segments to be tested. for instance , assume the reference segment transition is 1 -> 0 and therefore the test segment transition is 0 -> 1. If the reference segment transition is slower, i.e., delay (sr) > delay (st) then a static hazard is produced. Therefore the reference path segment delay must be larger than the test path segment delay.

4. GLITCH DETECTOR

As indicated above, the width of the static hazard is proportional to the quantity of additional delay introduced by a little delay defect. Therefore, the planning and layout of the glitch detector must be optimized to detect narrow glitches so as to maximise the sensitivity of the tactic to small delay defects. A second criterion is to stay it small, to attenuate the overhead related to the tactic. Last, the glitch detector must be compatible with launch-off-capture and launch-off-shift delay test methodologies so as for it to be deemed practical. One possible implementation of a glitch detector that meets these criteria is shown in Figure 3. The capture-flop input is shown at the highest of the figure and therefore the capture-flop (with scan) is shown on the proper. The remaining gates constitute elements of the glitch detector. An XOR glitch rectifier as shown in fig consists of two inverters is one of the XOR input. The output of the XOR glitch rectifier fed to the one of the input of the latch circut and the other input that is glitch enable signal is fed to the latch. The output of the latch enables the capture flop through MUX. The capture flop output fed to the next scan flop. Example transitions are shown within the figure. The XOR A output drives the input of a latch, i.e., two NOR gates configured with feedback. The output state of the latch is initialized to 0 before conducting the test by setting glitch_en high. The glitch_en control signal is then set low before application of the delay test patterns. If a static hazard is propagated to the capture-flop input as a results of the test, the first rising edge produced by XOR A flips the state of the latch and generates a 1 on the output of NOR A. The result is stored within the latch than the XOR B is enabled the resultant into scan chain mode. If the output of the latch is 0 then the content of the scan chain remains unchanged.. If the latch output is 1, XOR B flips the state of the bit getting into the capture-flop. The results of this test, also because the results of other tests performed simultaneously on other paths, are scanned out after setting glitch_en to 1. Figure 4 shows Glitch detector design.

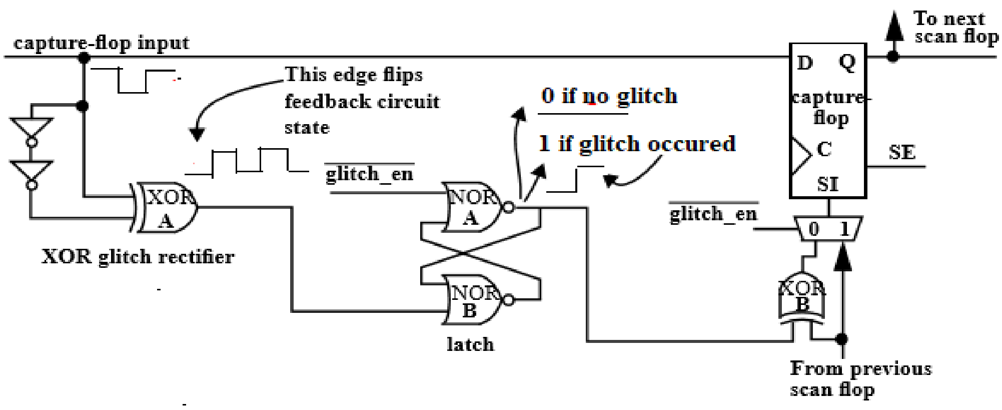


Figure 4. Glitch Detector Design

5. REFERENCE PATH

Reference Path trail sensitization relies on a group of reference paths that are validated to be freed from small delay defects. The reference paths are defined because the longest paths that drive each endpoint (capture latch or primary output). Since all other paths to the endpoints are shorter by definition, it follows that identifying and validation.

An inverter can be inserted in series with this connection as a means of reducing the load to one inverter input, in case of the load capacitance of the inverter and XOR gate is a concern. The coverage of small delay defects in the rest of the circuit can be maximized because of longest path. The validation of the reference paths can be carried out in one of two ways. An at-speed delay test can be applied to check that each reference path is upper-bounded by the clock cycle time. Although this approach works for the reference paths that are also critical paths in the circuit, it cannot be used to confirm a tight upper timing bound for the shorter reference paths. This is true because the shorter reference paths may have significant slack when tested with an at-speed clock. A straightforward solution for testing the shorter reference paths is to use a faster-than-at-speed clock. Unfortunately, testing results in yield loss due to IR drops and other noise effects because of testing at faster-than-at-speed. Our approach avoids the drawbacks associated with the application of a faster-than-at-speed clock by incorporating a special reference path on chip. Figure 5 shows the proposed test structure, subsequently referred to as the reference path test structure (RPTS). It consists of a launch-flop with scan on the left (enabling standard launch-on-shift/launch-on-capture transition testing), a tri-stateable inverter followed by a string of inverters that form a delay chain. The chain is tapped at each successive inverter output using a MUX. The tap point selected determines the delay from the launch-flop to the input of AND followed by NOT A shown on the right in the figure. The MUX delay select inputs control the selection of the tap point, are controlled by the test engineer using a scan chain. The AND followed by NOT A gate serves the role of the convergence gate, GC, described earlier. The bottom input of the AND followed by NOT A is driven by a path-select MUX. The inputs to the MUX are connected to those endpoints that require a faster-than-at-speed validation. The output of AND followed by NOT A drives a glitch detector and is identical to the glitch detector described in reference to Figure 4. While launching a transition from the launch-flop of the RPTS simultaneously, the shorter reference paths can be validated by applying a delay test to them. The expected delay of the reference path is emulated in the RPTS by selecting the appropriate tap point in the inverter chain. The output of the RPTS's glitch detector reflected with result of the race of the transitions along both paths. For example, with the RPTS configured with a delay larger than the reference path under test, the absence of a glitch indicates the reference path under test is shorter and free of small delay defects. The capture flop for inspection off-chip recorded as result. Accurate delay emulation by the RPTS requires knowledge of the actual delay of its delay chain, which can be obtained through calibration. Calibration is performed by configuring the delay chain into a ring oscillator. This is accomplished by setting RO_en to 1 in Figure 4. The delay select inputs to the MUX are configured so that the entire chain of inverters are part of the ring oscillator. A frequency divider (right side of Figure 4) is used to drive an off-chip pin connected to a frequency measuring instrument. The delay of the chain is the inverse of the measured frequency scaled by the value of the frequency divider. It becomes possible to configure specific delays into the RPTS for validating each of the shorter reference paths, when the RPTS is calibrated.

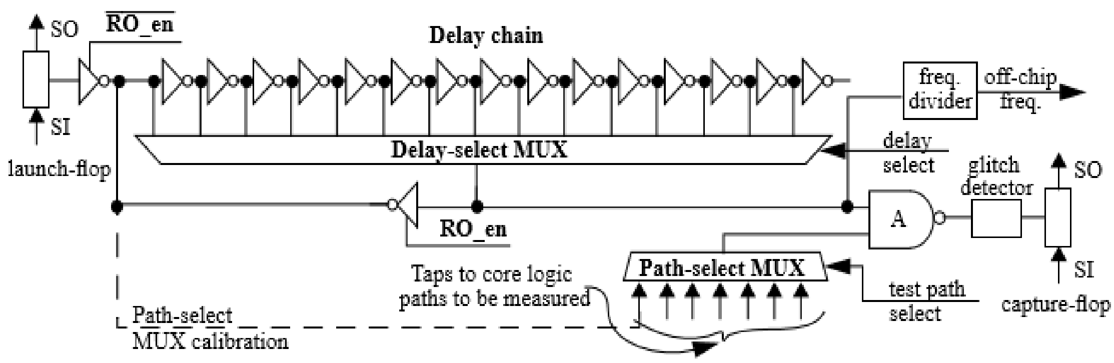


Figure 5. Reference path infrastructure

6. CONCLUSIONS

The advantages of the small delay defect testing include the elimination of a capture clock cycle, which significantly reduces test power issues. Without applying a faster-than-at-speed clock, this method can also detect very small delay defects. It describe a test method that is able to detect very small delay defects without requiring a faster-than-at-speed clock. The strategy also reduces test power by eliminating the capture clock cycle associated with standard delay testing. The technique uses of internal races as a means of bounding the delay of one path segment against another. The result of the test either causes a static hazard to be generated or the transition along a reference path to be halted. Glitch detectors are added to path endpoints as a means of distinguishing these two conditions. A reference path test structure is proposed to validate shorter reference paths against small delay defects. For timing analysis and to measure path delays for validation and debugging of first silicon, this test structure can also be used to aid with correlating models with actual hardware.

ACKNOWLEDGEMENTS

I wish to acknowledge the following colleagues Hardware design group and Intelligent Transportation and networking section who provided their valuable support and guidance to write the paper.

REFERENCES

[1] B. Kruseman, A. K. Majhi, G. Gronthoud and S. Eichenberger, (2004) "On hazard-free patterns for fine-delay fault testing", ITC , 2004, pp. 213-222.
[2] Jubayer Mahmod, Spencer Millican, Ujjwal Guin, and Vishwani Agrawal, (2019) "Delay Fault Testing: Present and Future", IEEE 37th VLSI Test Symposium (VTS).
[3] P. Nigh, (2003) "The Increasing Importance of On-line Testing to Ensure High-Reliability Products", ITC, 2003.
[4] P. Gupta and M. S. Hsiao, (2004) "ALAPTF: A New Transition Fault Model and the ATPG Algorithm", ITC, 2004.
[5] N. Ahmed, M. Tehranipoor and V. Jayaram, (2006) "Timing-Based Delay Test for Screening Small Delay Defects", DAC, 2006.
[6] A. D. Singh and G. Xu, (2006) "Output Hazard-Free Transition Tests for Silicon Calibrated Scan Based Delay Testing", VTS, 2006, pp. 349-355.
[7] B. N. Lee, L. C. Wang and M. S. Abadir, (2004) "Reducing Pattern Delay Variations for Screening Frequency Dependent Defects", VTS, 2005, pp. 153-160.

SEARCH IN A REDECENTRALISED WEB

Thanassis Tiropanis[1], Alexandra Poulovassilis[2],
Adriane Chapman[1] and George Roussos[2]

[1]School of Electronics and Computer Science, University of Southampton, UK
[2]Department of Computer Science and Information Systems,
Birkbeck, University of London, UK

ABSTRACT

Search has been central to the development of the Web, enabling increasing engagement by a growing number of users. Proposals for the redecentalisation of the Web such as SOLID aim to give individuals sovereignty over their data by means of personal online datastores (pods). However, it is not clear whether search utilities that we currently take for granted would work efficiently in a redecentralised Web. In this paper we discuss the challenges of supporting distributed search on a large scale of pods. We present a system architecture which can allow research, development and testing of new algorithms for decentralised search across pods. We undertake an initial validation of this architecture by usage scenarios for decentralised search under user-defined access control and data governance constraints. We conclude with research directions for decentralised search algorithms and deployment.

KEYWORDS

Redecentralisation, search, decentralised systems.

1. INTRODUCTION

The provision that third parties can maintain indexes of Web resources has been a key architectural choice of the Web from the beginning [7], and has played a significant role in its growth by enabling search engines and supporting the discovery of user generated content. However, in recent years, a large part of user activity and generated data has been concentrated on a small number of online platforms that have evolved into data silos, raising concerns and leading to proposals for the redecentralisation of the Web [6]. SOLID[1] [15] is a proposed suite of technologies to support such redecentralisation by envisaging that user data be always maintained in user-controlled personal online datastores (pods) as opposed to online platform-controlled data silos. Online applications need to request and obtain access to user pods in order to function according to this paradigm. This can enable users to share their data with multiple online application providers, fostering data-driven innovation and AI. Nevertheless, this would also require support for large-scale data search across pods.

There has been a large body of previous work on topics closely related to search in such redecentralised environments but currently there is no conceptual model of what search functionalities across SOLID pods would require. We first review relevant literature and then propose an architecture to support search in a Web that has been redecentralised based on the concept of pods as proposed in SOLID. To that end we describe a logical pod structure and identify components that could effectively support search across a large scale of pods. We

[1] https://solidproject.org

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 07-14, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112102

perform an initial high-level scenario-driven validation of these proposals and we propose further research and development roadmaps.

2. RELATED WORK

Search in decentralised data ecosystems is a non-trivial problem since it can involve both keyword and database-type queries distributed over a scale of thousands of datastores to which different search parties can have different access rights and different data governance constraints on data storage or migration may apply. Earlier work has explored distributed queries making use of database schemas and statistics across database endpoints which can have varying types of autonomy [27,20,1,12]. There has also been work on distributed information retrieval using meta-information about databases [8], peer-to-peer (P2P) data management [2, 14, 29, 17], search optimisation in P2P systems [24], social-graph-informed query routing [18], and socio-aware P2P search including work in the Haggle project that used a distributed index for search within groups [30, 23]. There has also been research on schema-based P2P data management with semantic links between data shared by peers [21, 11, 16]. Large-scale distributed search, architectures, query propagation and performance have been explored in Gaian databases [4, 28, 5]. IPFS [3] is a more recent approach to storing and retrieving data on a global, P2P decentralised file system. However, varying access control to available data resources, query endpoints and indexes has not been central to the design of these earlier approaches.

Prior work on securing distributed queries on personal repositories has approached the issue from the angle of addressing privacy threats [13]. Other work has focused on architectures for enforcing access control policies in P2P environments [26] but not in the context of distributed queries. Also proposed is attribute-based search on encrypted data with access control focusing on centrally stored data on the cloud [32]. The complexity of dealing with identities for access control in large-scale fog/edge computing has led to proposals on using distributed hash tables as a substitute for access control lists [31].

There is also a body of work on distributed indexing techniques [9, 10] often focusing on sensor networks and events. Decentralised search engines such as BitClave leverage blockchain to let Web users share their data directly with advertisers, removing intermediaries, but the emphasis is on users protecting their own data rather than on search algorithms across a large scale of users' personal datastores.

3. PROPOSED SYSTEM ARCHITECTURE

3.1. Stakeholders

To meaningfully explore search in a decentralised Web we identify stakeholders and a decentralised search architecture based on SOLID. The first stakeholder to acknowledge is the ***pod user*** who effectively owns and controls their pod, setting the desired access policies over their pod data. There can be more complex ownership models, e.g.: a minor's data held in trust by a guardian; data that is held jointly by two parties, such as marriage-related data between two individuals; or community owned data. Another stakeholder is the ***pod provider*** that provides the digital service and infrastructure to manage and host a pod. There can be many distinct providers in an open marketplace. Search providers can offer an interface for searches to be requested, providing optimisation on queries and metadata shared by pod owners; however, we note that decentralised search does not necessarily require a search provider but can operate on a peer-to-peer basis. We can also identify the ***search issuer*** as the individual or organisation that issues a search query via a search interface. Finally, ***regulating entities*** create and enforce access to information based on local and global laws, such as GDPR[2]. Any system that searches over information in pods must be able to accommodate the needs and requirements of each of these stakeholders.

3.2. Architecture

In the SOLID framework [15] individuals can identify themselves using WebID[3] and maintain their data resources within pods that can be stored in Web-accessible, user-owned or user-rented equipment, e.g. on local hard drives or on the cloud. Users decide who gains access to data in their pods by means of Access Control Lists and the Web Access Control system[4]. Pods can be hosted in pod servers (SOLID enabled Web servers) and their data can be accessed via RESTful interfaces, as in the Linked Data Platform recommendation[5]. Third-party applications can access data in users' pods if their WebID is linked to specific rights recorded for that data in the pod. URLs can be used to identify individuals or groups of individuals and their access rights to resources. SOLID pods may also offer SPARQL support including link-following SPARQL for specific applications, such as the Contracts app [15]. An underlying problem here is that, in allowing access to data in a pod, the user must pre-define who has access to their data; however, search in this case has a multi-faceted, undefined group of users and purposes. Other frameworks for cloud-based personal online datastores include that of the HAT project [22] which enables hosting of user-owned data, especially IoT data, in user-controlled containers similar to pods, and supports micro-services to enable an ecosystem of applications on that data.

Our proposal is for a refined logical structure for pods, compatible with the structure offered by SOLID and HAT, in order to support search algorithms across pods and foster research on optimisation techniques for both keyword and query-type search on a very large scale. There are two main themes required for searching within pods: a) to find appropriate pods that may contain the required data - but not share the data itself, merely identify it so that a contractual access negotiation can take place; and b) access to any publicly available data. In the first case, we need to make accessible to search algorithms possibly available data without impinging upon the privacy of the individual; efforts such as [19] could be drawn upon. We describe the second case in more detail below.

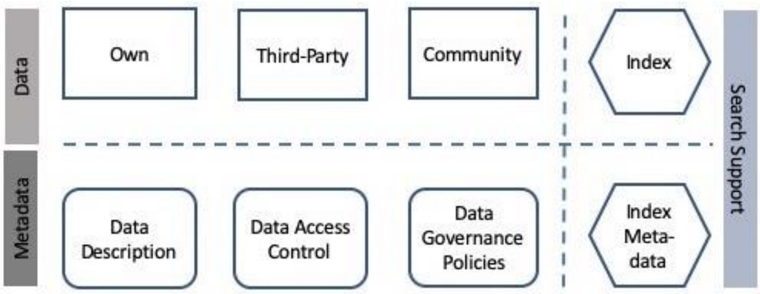


Figure 1. Pod logical structure to support decentralised search.

The logical architecture that we propose (illustrated in Figure 1) distinguishes between three types of pod data: *owner data*, *third-party data* stored in the pod with the consent of the user, and *community data* that are co-owned and potentially co-created by a community of users. SOLID

[2] https://gdpr-info.eu
[3] www.w3.org/2005/Incubator/webid/spec/
[4] www.w3.org/wiki/WebAccessControl
[5] www.w3.org/TR/ldp/

envisages that a user may have several pods [25] but given that the same individual is always in full control of them, we can conceptually work with a single pod. Another distinction in the proposed model is that of *data resources* and *metadata* to describe those resources; that distinction is made so that if a party issues a distributed query across pods and has access to pod metadata, query planning algorithms can decide whether a search query will be executed on a specific pod. Metadata in this sense are similar to terms like *meta-information*, *statistics* or *summaries* in the literature. In addition to such data description metadata, we envisage data *access control metadata* to distinguish metadata that can support optimisation when planning distributed search based on access criteria, for which agreement on schemas can be easier to reach. Finally, we also envisage *data governance policies* as another type of structured and potentiality widely agreed metadata on licensing, copyright, and data storage, migration and retention policies. To emphasise support for distributed keyword-based queries we also identify *indexes* as another type of data resource as well as index metadata to describe information on index access and use by search algorithms.

SOLID enables querying pods for access to locally stored data or for data in other pods using link-following [25]; the issuer of the query is responsible for retrieving links to other pods that can be queried. Our proposed search components architecture shown in Figure 2 aims to be compatible with search as either a third-party or a P2P application. For this reason, we distinguish between index building and index distribution components for optimisation in ways similar to those described in the literature but with additional capabilities to respect user-imposed data governance constraints such as access control, data indexing, and data migration across pods. We also distinguish between distributed keyword search engines and query engines as distinct components that could be used as P2P or third-party applications but could also be combined for more complex hybrid search cases. Finally, we envisage distinct components for query planning, optimisation and API interaction. We differentiate between optimisation and query planning since, especially for hybrid search, optimisation software can determine which query planning algorithms are most suitable each time based on the types of query, pod metadata, network topology etc. Adjacency of components in Figure 2 indicates where possible interfaces are likely to be defined (e.g. optimisation components interfacing with index distribution, distributed query engine and query planning ones) but it does not exclude other possibilities.

4. SUPPORT FOR USAGE SCENARIOS

We consider two scenarios for a high-level validation of our architecture that cover both top-k and exhaustive search under different types of access constraints of pod owners in terms of access control, data export and data processing.

Usage Scenario 1: top-k search to form a community of users who have a common interest in combating the risk of developing Type 2 diabetes by accessing appropriate recipes for meals. The users in this scenario are the *app developer* and a number of *users*; all users have one or more pods, holding non-personal application data for the developer and personal health and nutrition data for the other users. Suppose Eric has enabled access control on the personal data in his pod but has made some metadata public: the fact that he has personal health and nutrition data, in which schemas those data are available, licensing and copyright information on those metadata. He has also authorised an indexer app that maintains a local index for those metadata. The app developer rolls out a Web app which, using search components compatible with pod indexes, can first discover all users (including Eric) with potentially relevant health and nutrition data, and then offer them recipes that present an appropriate nutritional profile. Eric receives and approves a request to provide the Web app with access to his relevant pod data and to allow the app to use existing indexes in his pod or to create new ones for the purposes of the application. Eric can then start using the app; for example, he can request it to search for nutritional information from the *top-k* other users with a similar health profile to him and who have been showing a stable or improving health status.

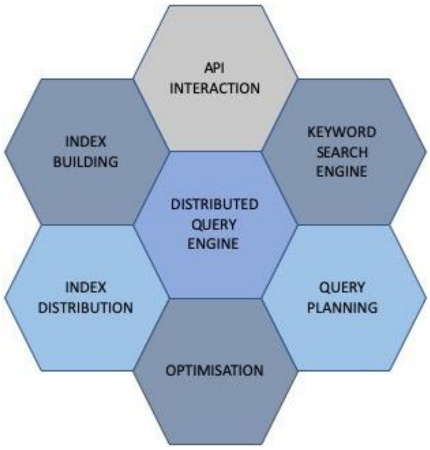


Figure 2. Search service components.

In this scenario, the provision of search components that can be integrated in applications, and indexers that can independently run on user pods, supports the deployment of a Web app that can provide users with relevant information without storing centrally their health and nutrition data. The indexers will use index building components, while the app will use keyword search on indexes to first identify users with relevant data and, once permission is granted, to execute distributed queries to obtain health profile and nutritional data using LDP/HTTP requests and/or SPARQL. Optimisation and query planning components can be integrated into the app to support these queries and link-following features to route queries across pods may be used.

Usage Scenario 2: exhaustive search over thousands of pods to investigate air quality in the city of Birmingham. Helen has a pod with geo-tagged data that she collects on her bike when cycling in Birmingham. She wants to contribute to the improvement of air quality but does not wish to compromise her privacy. She maintains her data in a pod and exposes metadata on the geo-range and time period covered in her pod. A developer provides a Web app that reports on levels of particulates around the city for different times of the day. Helen consents to the app using her pod with the condition that her data or query results cannot be stored by parties beyond her city boundary, and that they will always be aggregated with other query results to reduce the probability of triangulation by a certain factor. The app starts issuing queries to the user network, storing intermediate results in pods respecting their users' settings. Aggregated results are collected in the application pod and reported via the Web interface.

The application can use SPARQL link-following across pods for query propagation inspired by Gaian database approaches [28,5] but with extensions to support access control and restraints on the caching of pod data. Index building and distribution components specific for the app are used for query planning and optimisation. Optimisation algorithms need to consider network structure, user preferences on data aggregation and transfer, and routing distances for exhaustive search. A distributed query engine in combination with keyword search engine components and APIs support the Web app functionality.

5. CONCLUDING REMARKS

We have proposed an architecture to enable research and deployment of decentralised search at scale across SOLID pods, arguing that current distributed search techniques do not fully cover its requirements. We therefore propose that the community revisits distributed search in the context of decentralisation where access control and other data governance constraints are under the control of individuals. There is also scope to validate new decentralised search algorithms and optimisation approaches in existing or emergent ecosystems. Future work also requires supporting the definition of data governance policies in user pods and monitoring their observance.

REFERENCES

[1] Abedjan, Z., Golab, L., Naumann, F.: Profiling relational data: a survey. VLDB Journal 24(4), 557–581 (2015)

[2] Androutsellis-Theotokis, S., Spinellis, D.: A survey of peer-to-peer content distribution technologies. ACM Comp. Surv. 36(4), 335–371 (2004)

[3] Benet, J.: IPFS - Content Addressed, Versioned, P2P File System. Tech. rep., Protocol Labs (2014)

[4] Bent, G., Dantressangle, P., Stone, P., Vyvyan, D., Mowshowitz, A.: Experimental evaluation of the performance and scalability of a dynamic distributed federated database. In: Proc. 3rd Ann. Conf. Int. Tech. Alliance (2009)

[5] Bent, G., Dantressangle, P., Vyvyan, D., Mowshowitz, A., Mitsou, V.: A dynamic distributed federated database. In: Proc. 2nd Ann. Conf. Int. Tech. Alliance (2008)

[6] Berners-Lee, T.: Long Live the Web. Scientific American 303 (2010)

[7] Berners-Lee, T., Cailliau, R., Groff, J.F., Pollermann, B.: World-wide web: the information universe. Internet Research (1992)

[8] Callan, J.: Distributed Information Retrieval, pp. 127–150. Springer (2000)

[9] Danzig, P.B., Ahn, J., Noll, J., Obraczka, K.: Distributed indexing: A scalable mechanism for distributed information retrieval. In: Proc. 14th Ann. Int. Conf. Res. and Dev. in Inf. Retrieval. p. 220–229. SIGIR '91, ACM (1991)

[10] Greenstein, B., Ratnasamy, S., Shenker, S., Govindan, R., Estrin, D.: Difs: a distributed index for features in sensor networks. Ad Hoc Networks 1(2), 333 – 349 (2003)

[11] Halevy, A.Y., Ives, Z.G., Suciu, D., Tatarinov, I.: Schema mediation in peer data management systems. In: 19th Int. Conf. Data Eng. pp. 505–516. IEEE (2003)

[12] Haslhofer, B., Klas, W.: A survey of techniques for achieving metadata interoperability. ACM Comp. Surv. 42(2), 1–37 (2010)

[13] Loudet, J., Sandu-Popa, I., Bouganim, L.: Dispers: Securing highly distributed queries on personal data management systems. Proc. VLDB Endow. 12(12), 1886–1889 (2019)

[14] Lua, E.K., Crowcroft, J., Pias, M., Sharma, R., Lim, S.: A survey and comparison of peer-to-peer overlay network schemes. IEEE Comm. Surv. Tutorials 7(2), 72–93 (2005)

[15] Mansour, E., Sambra, A.V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., Aboulnaga, A., Berners-Lee, T.: A demonstration of the solid platform for social web applications. In: Proc. 25th Int. Conf. Companion on WWW. pp. 223–226 (2016)

[16] McBrien, P., Poulovassilis, A.: P2P Query Reformulation over Both-As-View Data Transformation Rules. In: Databases, Information Systems, and Peer-to-Peer Computing, pp. 310–322. Springer (2006)

[17] Meshkova, E., Riihija ̈rvi, J., Petrova, M., Ma ̈ho ̈nen, P.: A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks. ERCIM News 52(11), 2097–2128 (2008)

[18] Mislove, A., Gummadi, K.P., Druschel, P.: Exploiting social networks for internet search. In: In Proc. 5th Workshop on Hot Topics in Networks (HotNets-V) (2006)

[19] Mork, P., Smith, K., Blaustein, B., Wolf, C., Sarver, K.: Facilitating discovery on the private web using dataset digests. In: Proc. 10th Int. Conf. Information Integration and Web-Based Applications & Services. p. 451–455. ACM (2008)

[20] Naumann, F.: Data profiling revisited. SIGMOD Rec. 42(4), 40–49 (2014)

[21] Nejdl, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmer, M., Risch, T.: EDUTELLA: A P2P Networking Infrastructure Based on RDF. In: Proc. 11th Int. Conf. WWW. pp. 604–615 (2002)

[22] Ng, I., Maull, R., Parry, G., Crowcroft, J., Scharf, K., Rodden, T., Speed, C.: Making Value Creating Context Visible for New Economic and Business Models: Home Hub-of-all-Things (HAT) as Platform for Multisided Market powered by Internet-of-Things. In: HICSS (2013)
[23] Nordström,E.,Rohner,C.,Gunningberg,P.:Haggle:Opportunisticmobilecontent sharing using search. Computer Communications 48, 121 – 132 (2014)
[24] Reynolds, P., Vahdat, A.: Efficient peer-to-peer keyword searching. Springer (2003)
[25] Sambra, A.V., Mansour, E., Hawke, S., Zereba, M., Greco, N.: Solid: a platform for decentralized social applications based on linked data. Tech. rep., MIT CSAIL & Qatar Computing Research Institute (2016)
[26] Sandhu, R., Zhang, X.: Peer-to-peer access control architecture using trusted computing technology. In: Prof 10th ACM Symp. on Access Control Models and Technologies. p. 147–158. SACMAT ’05, ACM (2005)
[27] Sheth, A., Larson, J.: Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Comp. Surv. 22(3), 183–236 (1990)
[28] Stone, P.D., Dantressangle, P., Bent, G., Mowshowitz, A., Toce, A., Szymanski, B.: Query propagation behaviour in gaian database networks. In: Proc. Ann. Conf. Int. Technology Alliance (2012)
[29] Suryanarayana, G., Taylor, R.: A survey of trust management and resource discovery technologies in peer-to-peer applications. Tech. rep., UC Irvine (2004)
[30] Yoneki, E., Hui, P., Chan, S., Crowcroft, J.: A socio-aware overlay for publish/subscribe communication in delay tolerant networks. In: Proc. 10th ACM Symp. on Modeling, Analysis, and Simulation of Wireless and Mobile Systems. pp. 225–234 (2007)
[31] Zaghdoudi, B., Kaffel-Ben Ayed, H., Harizi, W.: Generic access control system for ad hoc mcc and fog computing. In: Foresti, S., Persiano, G. (eds.) Cryptology and Network Security. pp. 400–415. Springer (2016)
[32] Zheng, Z., Xie, S., Dai, H.N., Chen, X., Wang, H.: Blockchain challenges and opportunities: a survey. Int. J. of Web and Grid Services 14(4), 352–375 (2018)

AUTHORS

Thanassis Tiropanis has a research track record on e-Infrastructures, starting with network management, following on with research on Grid infrastructures for content delivery, and more recently infrastructures for Web Observatories. He has worked on distributed queries across federated and distributed data repositories using semantic and peer-to-peer techniques, and on conceptual models of Data Observatories. At the same time, he engaged in interdisciplinary research exploring the impact of the Internet and the Web on society and the economy. More recently he has been contributing to reports and proposals for infrastructures and policies to support data sovereignty. He is a senior member of IEEE, and a chartered IT professional with BCS.

Alexandra Poulovassilis has been Professor of Computer Science in Birkbeck’s Department of Computer Science and Information Systems (CSIS) since 2001, becoming Professor Emeritus in May 2021. Her research is in data integration, querying, visualisation and personalisation. She has published widely in these areas and held numerous grants from the UK Research Councils, EU and industry. She is Founding Director of the Birkbeck Knowledge Lab (Feb. 2016-March 2021), and prior to that Co-Director of the London Knowledge Lab (2003-2015), and has championed and engaged in interdisciplinary research across the sciences, social sciences and arts for many years.

Adriane Chapman is an internationally recognised expert in the field of data provenance. She won ACM SIGMOD's 2016 Test of Time Award for her work on provenance (Buneman et al. 2006), and has over 30 peer-reviewed publications spanning provenance, data integration, and trust. She has designed software systems for agencies across the US government. Relevant to this proposal is her work with the US Food and Drug Administration on searching for related data across independently operated and governed systems. She received a Program Recognition Award for her work on incorporating provenance into the National Geospatial-Intelligence Agency Human-Geography Pilot. She is currently the Director of the ECS Centre for Health Technologies. Data discovery, access and handling are major components of this work.

George Roussos is Professor of Pervasive Computing at CSIS, where he heads the Experimental Data Science Group and the IoT lab. He has over 20 years experience in leading and successfully delivering research projects supported among others by the AHRC, EPSRC, the European Commission and the Michael J Fox Foundation. His work pioneered participatory cyberphysical computing as the predominant methodology for the construction of mobile and pervasive computing systems. His current research interests include applications of the IoT in mobile healthcare, the effects of social activity on mobile system architectures, and exploring mechanisms to support navigation and findability. He has authored four books and over 100 research papers, and since 2004 serves on the ACM US Public Policy Committee.

MULTI-LAYER ENCRYPTION ALGORITHM

Akula Vamsi Krishna Rao[1], V.N. Aditya Datta Chivukula[2],
Sri Keshava Reddy Adupala[2] and Abhiram Reddy Cholleti[3]

[1]Department of Computer Science and Engineering,
CMR Engineering College, Medchal, India
[2]Department of Computer Science and Engineering,
International Institute of Information Technology, Bhubaneswar, India
[3]Department of Electronics and Telecom Engineering, International Institute of Information Technology, Bhubaneswar, India

ABSTRACT

In recent years, security has become a big issue for many applications to defend attacks from intruders. Exchanging credentials in plaintext might expose it to stealers. Many techniques are required to protect the data of the consumers from attackers. Cryptography has come up with a solution to provide security for the users to exchange data securely by the means of the process called as Encryption/ Decryption. In this field, there are basically two techniques of cryptography i.e Symmetric and asymmetric, developed to achieve a secure connection between the sender and receiver. These techniques provide specific goals in maintaining privacy by converting original message to non-readable form and sends it over a communication channel. The unauthorized members try to break the non-readable form but the difficulty depends upon the techniques that were used to encrypt the data. In this paper, we proposed a quadruple encryption algorithm consists of novel phase-shift algorithm, AES (Advanced Encryption Standard), TwoFish and RC4 and making it hard to attack by common methods.

KEYWORDS

Cryptography, AES, Two-fish, RC4, Phase-shift.

1. INTRODUCTION

Cryptography is the process of converting a normal plaintext to unreadable text in the form of using different algorithms [1]. It can be used for authentication, protecting data from criminal which stands against them by locking the particular data using the key. Cryptography involves: Plaintext, Encryption Algorithm, Ciphertext, Decryption algorithm, Encryption key and Decryption key.

Cryptography can provide the following services:

- Confidentiality
- Integrity
- Authentication
- Non-repudiation

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 15-21, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112103

Types of Cryptography:

- Symmetric Key Cryptography: It consists of only one key which is used for both encryption and decryption. Symmetric Key consists of Block and Stream algorithms [2]. Some examples of symmetric encryption algorithms include: Advanced Encryption standard (AES), Data Encryption Standard (DES), Blowfish, RC4(Rivest Cipher 4), RC5(Rivest Cipher 5), RC6(Rivest Cipher 6).
- Asymmetric Key Cryptography [2]: It consists of 2 keys which is public key and private key. if message(m) is encrypted with public key then the encrypted text is decrypted using the private key. Some examples of Asymmetric encryption algorithms include: Rivest, Shamir, Adleman (RSA), Diffie-Hellman, Elliptic curve Cryptography (ECC), El Gamal.
- Hash Function [3] in cryptography: Message Digest 5(MD5), SHA (Secure hashing algorithm) like SHA-0, SHA-1, SHA-2, SHA-3 and CRC32.

2. ALGORITHM

This section discusses about a novel phase-shift algorithm and three existing encryption algorithms i.e AES, Twofish and RC-4. The architecture of the proposed technique is shown in fig.1 with phase-shift algorithm in the first level, AES in second level, Twofish in the third level, followed by RC-4 and at last again with phase-shift algorithm.

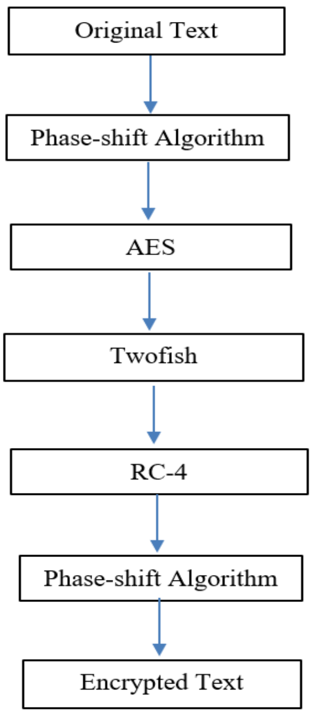


Figure 1. Encryption Process

At each level, the encrypted text is passed to its next level of encryption and the key that is used to encrypt, is stored in a separate array.

For instance, a message is given as an input in the first level. Here, the message gets encrypted and it is passed to the second level. The key that is used to encrypt the message is stored in an array (A1). Now, the second level encrypts the output of the first level i.e encrypted text by first level and stores the key (second level) in the array (A1). Later, the output of the second level is given as an input to the third level. This process is continued till the fifth level.

The previous 5 levels have different set of keys, and the entire set of keys are stored in the form of an array and attached at the end of the data resultant array after four levels of transformations or encryptions. This augmented array is undergone through the final phase shift. Hence, until the angle is known, we cannot perform the inverse phase-shift correctly and thereby, cannot even identify the keys of remaining four levels. This ends the security encryption of the data and algorithm in entirety. Similarly, decryption follows the reverse flow of encryption as shown in fig.1.

2.1. Phase-Shift Algorithm

This algorithm [4] is used to rotate an n-dimensional vector by a given angle 'α' (n>=2) having same magnitude, but, with different arrangement of the values. The algorithm basically adopts a dynamic programming approach to perform the task. Initially, it starts with taking the first two dimensions and compares the angle between the original arrangement (first two dimensions) and every arrangement possible keeping the first dimension fixed and adding second dimension in different ways. This process is repeated keeping in view about the cut-off angle is achieved or not. A detailed example is discussed below,

Assume, the initial vector (init_vector) is [1,2,3] and we have to rotate the vector, such that, the rotated vector combination is 90 degrees to the initial. So,

Step 1: let temp_vector = [1] (first dimension of original vector) and α = 0 degrees

Step 2: take second dimension and calculate all possible angles with the original two-dimensional space of vector.
Two combinations that are possible is [1,2] and [2,1]

Case 1: angle between [1,2] and [1,2] is 0 degrees, hence, α_temp = 0 and check if 90>α_temp> α.

If, true, α_temp = α else no change in α.

Case 2: angle between [1,2] and [2,1] is 37 degrees, hence, α_temp = 37 and following previous case, as 90>α_temp > α, α = α_temp.

Step 3: update the temp_vector = [2,1] (vector combination corresponding to greatest α).

Step 4: assign α _temp = 0

take third dimension and calculate all possible angles with the original three-dimensional space of vector.

Three combinations that are possible is [3,2,1], [2,3,1] and [2,1,3].

Case 1: angle between [1,2,3] and [2,1,3] is 21.7 degrees, hence, α_temp = 21.7 and as 90>α_temp> α, therefore, α = 21.7 degrees

Case 2: angle between [1,2,3] and [2,3,1] is 38.2 degrees, hence, α_temp = 38.2 and as 90>α_temp> α.

α = 38.2 degrees.

Case 3: angle between [1,2] and [3,2,1] is 44.41 degrees, hence, α_temp = 44.41 and as 90>α_temp > α, α = 44.41 degrees.

Vector corresponding to highest angle is [3,2,1] and this is the last iteration, hence, temp_vector = [3,2,1]. The highest angle possible with given combination is 44.41 degrees.

Therefore, if the given angle is not possible, the algorithm rotates the vector to near required angle. This entire algorithm can be used for transmission of arrays of data with key as the angle. This is the first level of security transformation applied to data.

2.2. AES (Advanced Encryption System)

The AES algorithm works on a 128-bit block of data and executed N - 1 loop times. The key length of the algorithm is 128, 192 or 256 bits in length [5,6]. The first and last round of the algorithm has different properties where AddRoundkey is added in first rounds whereas MixColumn is removed in the last round.

Here, we use the AES-128 of 128-bit length key for explanation. The AES Encryption algorithm is broken into 4 categories for operation i.e Sub Bytes, Shift rows, Mix Columns and AddRoundkey. We also do key expansion in the round for cipher key. The separate transforms are performed in a number of rounds that are dependent on the cipher key size. Generally, for key size of 128 we perform 10 rounds, for key size of 196 we do 12 rounds and for 256 we do 14 rounds respectively.

The operation in AES are as follows:

- SubBtyes Transformation: In Sub Bytes transformation bytes are transformed using a non-linear S-box which is invertible. Generally, it is represented as a 16*16 array, where rows and columns are indexed by hexadecimal bits. The corresponding value of the row and column are replaced from the values to it in S-box.
- ShiftRows Transformation: In ShiftRows transformation, the function shifts the bytes in each row of matrix, according to their offsets 0,1,2,3. Let there are 4*4 matrix where four rows are shifted cyclically to left. This is a simple permutation and nothing more.
- MixColumns Transformation: The MixColumn operation is basically a substitution. The byte of a column is mapped into a new value that is function of all bytes in the column. Each element of product is the sum of products of elements of one row and one column.
- Add RoundKey Transformation: In this Stage the 128-bit of state are bitwise XORed with 128-bit of round key. This operation is involved as a column-wise operation between the 4 bytes of state column and one word round key. It proceeds at one column at time. The column matrix is added by the round key word.
- Key Expansion: The AES Key expansion takes as input of a four word i.e. 16 byte key and produces a array of type linear in 44 words i.e. 176 bytes. Each round uses four of these words, where each word contains 32 bytes that means each subkey is 128 bits long.

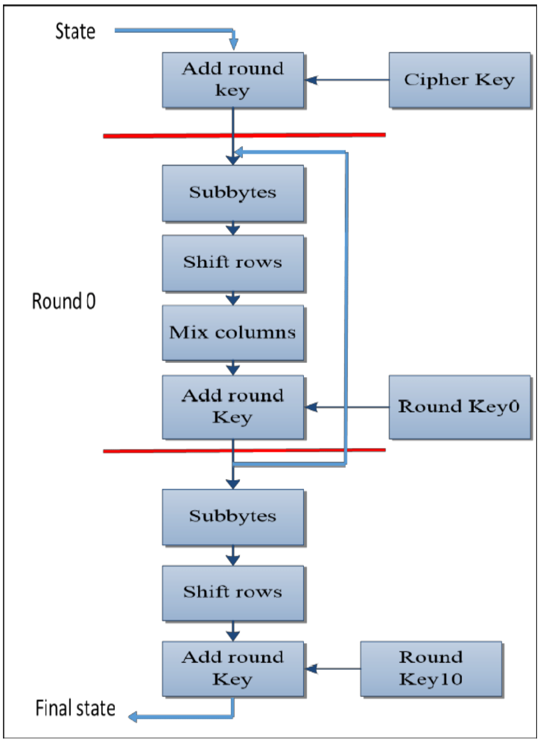


Figure 2. AES Encryption and Decryption

AES Decryption is the reverse process of AES Encryption which inverse round transformations to computes out the original plaintext of an encrypted cipher-text in reverse order. The round transformation of decryption uses the functions Add RoundKey, InverseMix Columns, Inverse Shift-Rows and Inverse Sub-Bytes successively.

- Add RoundKey: Add Roundkey decryption is the operation of inverse transformation of the Encryption.
- Inverse Sub-Bytes transformation: The Inverse Sub-Bytes transformation is done using a once pre-calculated substitution table called InvS-box. That Inverse S-box table contains 256 numbers (from 0 to 255) and their corresponding values.
- InvShiftRows Transformation: Inverse Shift-Rows exactly functions the same as Shift-Rows, only in the opposite direction
- InvMix Columns Transformation: The Inverse Mix Columns transformation is performed using polynomials degree less than 4, which coefficients are the elements in the columns of the state.
- Key Expansion: Key Expansion of AES Decryption is similar to the Key Expansion of AES Encryption.

2.3. Two-fish

Twofish encryption algorithm is a symmetric block cipher which was designed by Schneier in 1998 [7]. Two fish is related to its predecessor block cipher Blowfish. This block cipher considers the block size of 128 bits and key sizes of 128, 192 and 256 bits. This algorithm uses a pre-computed, key-dependent S-box and a relatively complex key schedule [8]. This algorithm was extensively cryptanalyzed and all design elements have a reason. It has 16 rounds Feistel-like structure with additional whitening at the input and output. The basic process of Two fish follows:

- The plaintext is broken up into four 32-bit words and each is XORed with a 32-bit expanded key.
- In each round, two 32-bit words serve as input into the function. Each word is broken up into 4 bytes. Those 4 bytes are sent through four different key-dependent S-boxes.
- The four output bytes are combined using a Maximum Distance Separable (MDS) matrix and combined into a 32-bit word.
- later the two 32-bit words are combined using a Pseudo-Hadamard Transform (PHT). It is added to two round subkeys and then XORed with the right half of the text. There are also two 1-bit rotations going on, one before and one after the XOR.

2.4. RC-4

RC4 is a symmetric cryptographic technique (algorithm) that belongs to the serial cypher (stream cipher) category of symmetric cryptography [9,10]. It's a byte-oriented stream cypher with a variable key length. The RC4 algorithm is distinguished by its simplicity and speed of execution. The key length is also flexible, with a range of 1-256 bytes (8-2048 bits). When the key length equals 128 bits, the violent search key is no longer used, according to current technological support. It's far too likely, thus the RC4 key range should be capable of surviving violent search key attacks for a long period. In fact, there has yet to be discovered an effective attack method for the RC4 encryption algorithm with a 128-bit key length. The key variables in RC4 are

- Key stream: The RC4 algorithm's fundamental feature is that it generates a key stream depending on the plaintext and the key. The length of the key stream and the plaintext are equivalent. That is, the plaintext is 500 bytes long, and the key stream is also 500 bytes long. The encrypted ciphertext is, of course, 500 bytes long, Since the ciphertext i-byte = plaintext i-th byte ^ key stream i-th byte.
- State vector S: The length is 256 characters. S(0), S(1), S(2), S(3), S(4).......S(255), each unit is a byte. Every time the algorithm is run S both contain a set of 8-bit numbers ranging from 0 to 255, with the only difference being the value's position.
- Temporary vector T: Each unit is also a byte, and the length is 256. Assume that the key is 256 bytes long, the key's value is assigned to T directly; otherwise, each byte of the key is assigned to T in turn.
- Key K: The length is 1-256 bytes. It's worth noting that the length of the key keylen isn't always proportional to the plaintext and key stream lengths. The key is usually 16 bytes in length (128 bits).

3. CONCLUSION

The need for secure transmission of data has become the call of the day, and hence, it is very much important to develop new and secure algorithms to increase trust and reliability of the data transmission. The work presented is one such effort towards the need or the cause. We tried to deliver a new way of modelling for secure data transmission and have put forward a new algorithm in combination with the current industry standard algorithms. We are very much looking forward to what the field has got to offer in the future and continue our work in shaping it.

REFERENCES

[1] Delfs, Hans, Helmut Knebl, and Helmut Knebl. Introduction to cryptography. Vol. 2. Heidelberg: Springer, 2002.

[2] Chandra, Sourabh, et al. "A comparative survey of symmetric and asymmetric key cryptography." 2014 international conference on electronics, communication and computational engineering (ICECCE). IEEE, 2014.

[3] Gauravaram, Praveen, and Lars R. Knudsen. "Cryptographic hash functions." Handbook of Information and Communication Security. Springer, Berlin, Heidelberg, 2010. 59-79.

[4] V.N.Aditya Datta Chivukula, Abhiram Reddy Cholleti &Rakesh Chandra Balabantaray, "Lyrics to Music Generator : statistical approach", to appear in Proceedings in International conference on data mining and machine learning (2021).

[5] AES Proposal:Rijandael, Daemen, Joan, and Vincent Rijmen. "AES proposal: Rijndael." (1999).

[6] Supriya, Niranjan, and Mr R. Niranjan. "Realization of AES Encryption and Decryption Based On Null Convention Logic." (2015): 77-81.

[7] Schneier, Bruce, et al. The Twofish encryption algorithm: a 128-bit block cipher. John Wiley & Sons, Inc., 1999.

[8] Bhanot, R. and Hans, R. (2015) A Review and Comparative Analysis of Various Encryption Algorithms. International Journal of Security and Its Applications, 9, 289-306.

[9] T. D. B. Weerasinghe, "Analysis of a Modified RC4 Algorithm," International Journal of Computer Applications, vol. 51, no. 22, 2012.

[10] P. Jindal dan B. Singh, "A Survey on RC4 Stream Cipher," I. J. Computer Network and Information Security, pp. 37-45, 2015.

AUTHORS

Akula Vamsi Krishna Rao is currently an undergraduate student in the Department of Computer Science and Engineering at CMR Engineering College, India. His area of interests is in Cryptography and Offensive Security.

V. N. Aditya Datta Chivukula is currently an undergraduate student in the Department of Computer Science and Engineering at International Institute of Information Technology, India. His area of interest in Machine learning, Deep learning, etc.

Sri Keshava Reddy Adupala is currently an undergraduate student in the department of Computer Science and Engineering at International Institute of Information Technology, India. His area of interests is in Data Analytics, Data Visualization and Machine Learning.

Abhiram Reddy Cholleti is currently an undergraduate student in the Department of Electronics and Telecom Engineering at International Institute of Informational Technology, India. His area of interest is in antenna design and IoT.

A STUDY OF THE CLASSIFICATION OF MOTOR IMAGERY SIGNALS USING MACHINE LEARNING TOOLS

Anam Hashmi, Bilal Alam Khan and Omar Farooq

Department of Electronics Engineering,
Aligarh Muslim University, Aligarh, India

ABSTRACT

In this paper, we propose a system for the purpose of classifying Electroencephalography (EEG) signals associated with imagined movement of right hand and relaxation state using machine learning algorithm namely Random Forest Algorithm. The EEG dataset used in this research was created by the University of Tubingen, Germany. EEG signals associated with the imagined movement of right hand and relaxation state were processed using wavelet transform analysis with Daubechies orthogonal wavelet as the mother wavelet. After the wavelet transform analysis, eight features were extracted. Subsequently, a feature selection method based on Random Forest Algorithm was employed giving us the best features out of the eight proposed features. The feature selection stage was followed by classification stage in which eight different models combining the different features based on their importance were constructed. The optimum classification performance of 85.41% was achieved with the Random Forest classifier. This research shows that this system of classification of motor movements can be used in a Brain Computer Interface system (BCI) to mentally control a robotic device or an exoskeleton.

KEYWORDS

EEG. Machine learning. BCI. Motor Imagery signals. Random Forest.

1. INTRODUCTION

The last decade has witnessed some tremendous advancements particularly in the field of medicine and technology. The inter-dependence of the two said fields is becoming more and more pronounced day by day; virtual surgical theatre, robotic surgery, Brain-controlled wheelchair are the name of the few recent developments. Nowadays, the study of biomedical signals has caught the attention of researchers as it provides the avenue for efficient disease diagnosis, development of assistive technologies, health monitoring of the elderly and aiding humanity in general [1]. This study explores further this very dimension by analyzing different methodologies used in studying Brain-Computer Interface (BCI). Electroencephalogram or EEG is one of the most common non- invasive methodologies of BCI to record brain signals. It measures the electrical activity of the brain using electrodes that are placed over the scalp. EEG is preferred because of its ease of portability and capturing high temporal brain information, however, it fails in capturing high spatial information [2]. BCI uses these EEG signals associated with the user's activity and then apply different signal processing algorithms for translating the recorded signals into control commands for different applications. In an EEG there are five types of oscillatory waves that are commonly used for analysis, which are:

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 23-33, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112104

(a) delta (0.5–4 Hz);
(b) theta (4–7 Hz);
(c) alphaormu (7–13 Hz);
(d) beta (13–25 Hz);
(e) gamma (25–50 Hz).

Motor imagery (MI) is a process in which an individual rehearses or stimulates an action. It is a very popular paradigm in the analysis of an EEG based BCI system. MI activity usually lies in alpha (or mu) and beta bands [3].

In the past few years, significant advances have been made in the BCI systems and they have revolutionized rehabilitation engineering by providing the differently-abled individuals with a new avenue to communicate with the external environment. According to many works of literature, the strength of a BCI system depends upon the methods in which the brain signals are translated into control commands of machines. A novel method namely an arc detection algorithm to find an optimal channel was proposed by Erdem Ekran and Ismail Kurnaz [4]. For feature extraction DWT was used and a number of machine learning algorithms were used for classification purposes, which were SVM, K- nearest neighbor, and Linear Discriminant Analysis. The best accuracy achieved by their methodology was 95% in classifying ECoG signals (BCI competition III, dataset I). Jun Wang and Yan Zhao proposed feature selection based on one dimension real-valued particle swarm optimization, extracted nonlinear features such as Approximate entropy and Wavelet packet decomposition, and achieved the best accuracy of 100% [5]. Aswinseshadri. K et al. used the wavelet packet tree for feature extraction. They used genetic algorithm, applied information gain, and mutual information to find the best feature set and for classification K-NN and Naïve Bayes were employed [6]. Chea-Yau Kee et al. proposed a novel feature known as Renyi entropy that has been employed for feature extraction and BLDA for classification [7]. K. Venkatachalam et al. proposed the use of the Hybrid-KELM (Kernel Extreme Learning Machine) method based on PCA (Principal Component Analysis) and FLD (Ficher's Linear Discriminant) analysis for MI BCI classification of EEG signals. The best accuracy reported was 96.54% [8]. Rajdeep Chatterjee et al. used the AAR (Auto Adaptive Regressive) algorithm for feature extraction, proposed a novel feature selection method based on IoMT (Internet of Medical Things), and classified EEG signals using SVM and ensemble variants of classifiers. The best accuracy reported was 80% [9]. The authors of [10] employed a combination of common spatial patterns (CSP) and local characteristic- scale decomposition (LCD) algorithm for feature extraction, a combination of firefly algorithm and learning automata (LA) to optimize feature selection, and spectral regression discriminant analysis (SRDA) classifier for classifying MI-EEG signals. They have used this method for a real- time brain-computer interface in order to show their method's efficiency.

Most of these studies have worked on the classification of right vs left-hand movement, or hand vs tongue movements, or hands vs legs movements. There are very few works that have studied and classified intricate hand movements such as opening and closing of a hand, or movements of different fingers, or classification of different hand gestures using neural signals, and those who have worked on these subjects either did not achieve high enough accuracy or failed to work in a real-world setup. This paper probed this very aspect of studying intricate human motions and worked on the classification of imagining of opening and relaxing of a hand using MI-EEG signals.

The contributions of this paper are following:

(i) Accurate classification of the motor imagery signals using a very simple algorithm design which is general in nature and hence can be used for other physiological signal classification.

(ii) Identification and ranking of most important features, from which it can be observed that although using a lesser number of features may not seem intuitive but in reality it has improved the classification accuracy.

The organization of this paper is as follows: the first part is the Introduction stage, where a brief introduction was provided and related works were reported, followed by Materials and Methodology stage. In this part, the materials or data that was used in the paper is described and the methodology of this work was elucidated. The third stage involves the results of the study with detailed discussion followed by conclusion.

2. MATERIALS AND METHODOLOGY

2.1. Data Used

The data used in this study was taken from [11]. The data consist of EEG recordings of a single subject. The subject was connected with a high spinal cord lesion and was controlling an exoskeleton (Brain-Neural computer interface) attached to his paralysed limb. The cue-based BNCI paradigm consisted of two different tasks, namely the 'imagination of movement' of the right hand (Class 1) and 'relaxation/no movement' (Class 2).

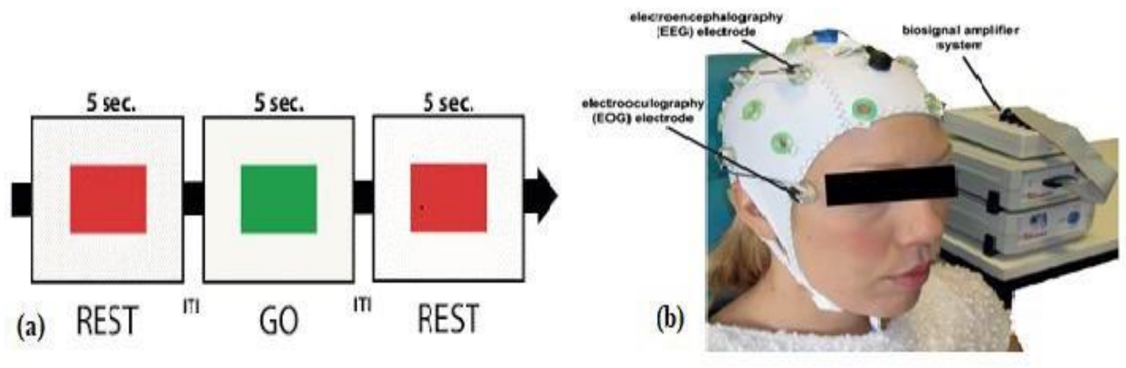


Figure 1. Timing Scheme of the trials used (a) and a subject during the EEG recording of the Dataset (b).

A randomly shown visual cue is used to indicate to the user when to open (for Green square) and when to close (for Red Square). These two indications were given 24 times each in total separated by inter-trial intervals (ITIs) of 4-6 seconds. Each indication was displayed for 5 seconds after which the device was driven back to open position. Re-setting the exoskeleton into op3en position required one second.

EEG was recorded from 5 conventional EEG recording sites F4, T8, C4, Cz, and P4 according to the international 10/20 system using an active electrode EEG system (Acti-cap® and BrainAmp®, BrainProducts GmbH, Gilching, Germany) with a reference electrode placed at FCz and ground electrode at AFz. EEG was recorded at a sampling rate of 200 Hz, bandpass filtered at 0.4-70Hz and pre-processed using a small Laplacian filter.

2.2. Pre-Processing

At this stage, the data was processed or filtered to capture information related to Motor Imagery. For that, the wavelet decomposition was done to obtain four-level details. The wavelet transformation of the EEG record at four levels resulted in four detail coefficients and one approximate coefficient with the frequency ranges listed in Table 1. Many electrophysiological features are associated with the brain's normal motor output channels [12]. Some of these important features are the mu (8-12 Hz) and beta (13-30 Hz) rhythms [13]. We concluded from Table 1 that the details cD2, cD3 and cD4 provide proper representation for the mu and beta rhythms and we decided to extract the vectors of features from these details.

Table 1. Frequency range for the decomposed details and approximation

Signal Component	**Frequency Range(Hz)**
cD1	50-100
cD2	25-50
cD3	13-25
cD4	7-13
cA4	0-7

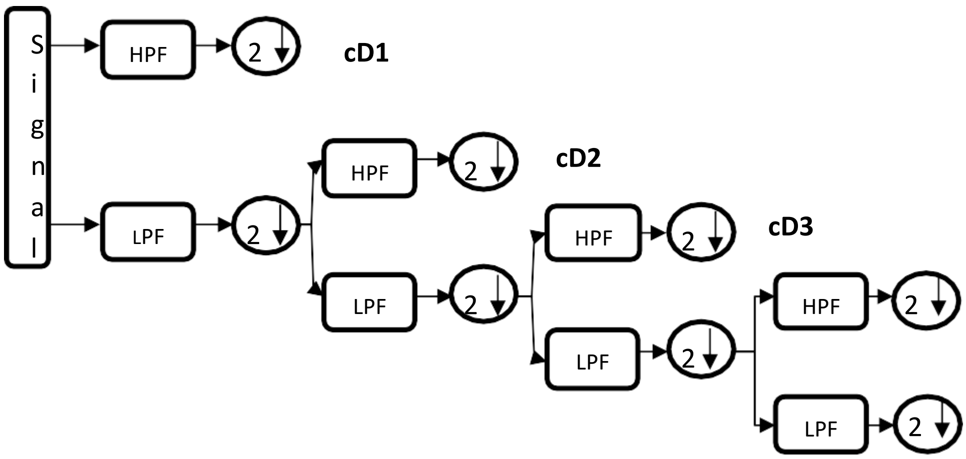


Figure 2. Showing decomposition of EEG signal into different wavelet coefficients corresponding to various frequency bands using wavelet transform.

2.3. Feature Extraction

A feature is a measurable property or characteristic of an observed signal. It should be informative, discriminative and orthogonal to other features. Feature extraction is the method of extracting these features. It can be defined as the process of transforming original data into a dataset with a reduced number of variables but with the most discriminative information.

After the wavelet transforms, the channels F4, T8, C4, Cz and P4 of each EEG record was analyzed using the Daubechies wavelet. Then the features namely Interquartile Range (IQR), Median Absolute Deviation (MAD), Variance, Skewness, Kurtosis, Energy, Mean Absolute Value (MAV) and Standard Deviation were calculated. The choice of these particular features can be understood as these two classes are different. Particularly, the two classes – which corresponds to imagining of opening of hand as 'class 1' and relaxation or no movement as 'class

2' – differ in dispersion. The same can be observed from the histogram of class 1 and class 2 appears, where the class 1 histogram appears to be skewed from the normal distribution. Thereby justifying the choice of IQR, MAD, Variance, Standard Deviation, Skewness and Kurtosis. Energy and MAV were chosen because it has been reported in many works that mu rhythm has a lower amplitude than that of the alpha wave [14].

The following are the mathematical equations of the extracted features:

2.3.1. IQR

It is a measure of statistical dispersion, being equal to the difference between 75th and 25th percentiles, or between upper and lower quartiles. Mathematically, it is defined as:

$$IQR = Q3 - Q1 \qquad \text{(i)}$$

Where, Q3 and Q1 represents the 75^{th} and 25^{th} percentiles of the distribution.

2.3.2. Median Absolute Deviation

It is defined, as the name suggests, median value of the absolute deviations from the data median value.

$$MAD = median(|Xi - median(X)|) \qquad \text{(ii)}$$

Where, Xi is the ith value of the data X.

2.3.3. Variance

It is defined as the expectations of the squared deviation of a random variable from its mean.

$$Var(X) = E[(X - \mu)^2] \qquad \text{(iii)}$$

Where, Var(X) computes of variance of data X, 'μ' represents the average value, 'E' represents the expectation.

2.3.4. Skew

It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

$$\gamma = E\left[\left(\frac{X - \mathrm{u}}{\sigma}\right)^3\right] \qquad \text{(iv)}$$

Where, 'ɣ' represents the skewness of data X.

2.3.5. Kurtosis

It is a measure of 'tailedness' of the probability distribution of a real-valued random variable.

$$Kurt(X) = E\left[\left(\frac{X - \mathrm{u}}{\sigma}\right)^4\right] \qquad \text{(v)}$$

2.3.6. Energy

It is the area under the squared magnitude of the considered signal. Mathematically,

$$Es = \sum_{n=-\infty}^{\infty} |X(n)|^2 \qquad \text{(vi)}$$

2.3.7. Mean Absolute Value

It is defined as the mean value of the absolute values of the data. Mathematically,

$$MAV = \frac{1}{N}\sum_{i=1}^{N} |X_i(n)| \qquad \text{(vii)}$$

2.3.8. Standard Deviation

It is a measure that is used to quantify the amount of variation or dispersion of a set of data values. It can be defined as,

$$SD = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\left(x[n] - \frac{1}{N}\sum_{n=1}^{N} X[n]\right)^2\right)} \qquad \text{(viii)}$$

2.4. Methodology

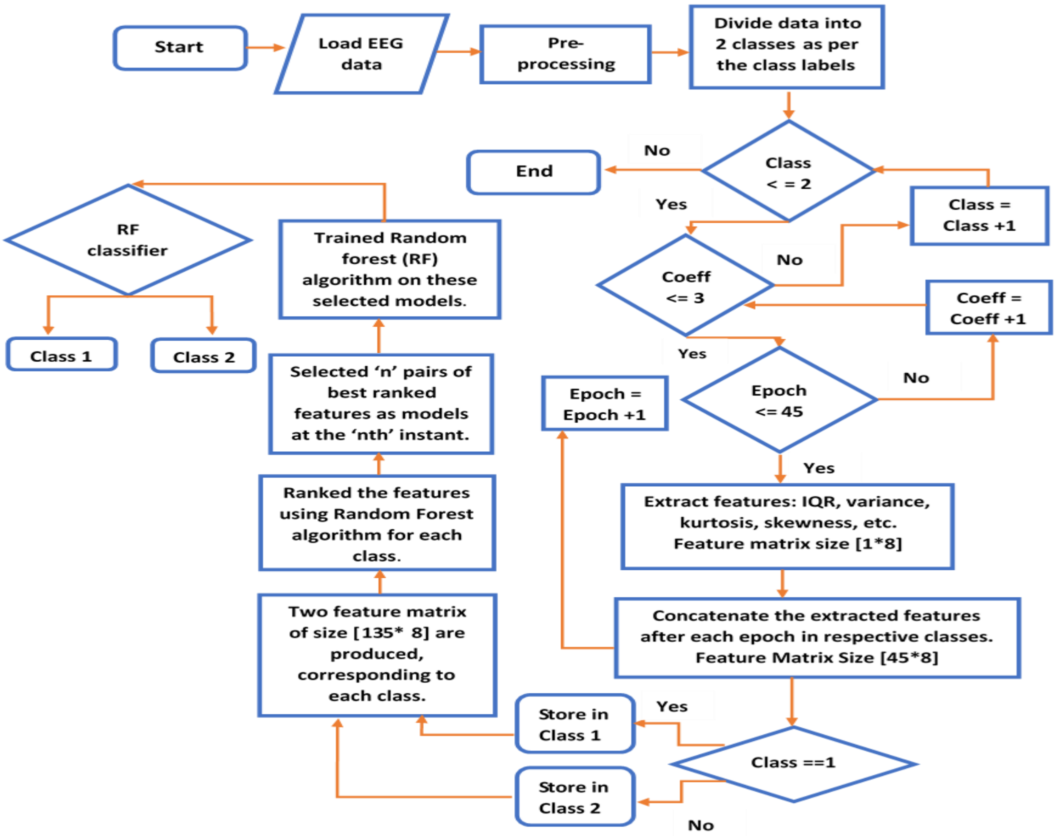


Figure 3. Showing the flowchart of the used algorithm for classification

In this study, there are two classes corresponding to the Motor Imagery (MI) tasks; hand opens and hand relaxes. The total duration of 'Class 1' is 45 seconds and the sample rate is 200 Hz, which produces 9000 data points. The data was pre-processed and decomposed into multiple bands using wavelet transform. Those bands were considered which corresponds to the Mu rhythm (8 – 13 Hz) and Beta rhythm (13 – 40 Hz) as these bands contain motor imagery related brain activity. Therefore, cD2, cD3, cD4 are chosen. A one-second sliding window is considered for the analysis of the signal and for each second 200 samples are considered, for which 8 features were extracted. This process was performed until the end of the recording, thereby producing a feature matrix of the size of [45×8]. This was done for a single coefficient of the wavelet-decomposed signal. Therefore, for all three coefficients, the produced feature matrix size was [135×8]. A similar analysis was performed for the 'Class 2' as well which also produced the feature matrix of size [135×8]. After the feature extraction, these features were made to pass through the feature selection stage where each feature is ranked (or given importance). The Figure 3. shows the flowchart of the used algorithm.

2.5. Feature Selection

This section describes the feature selection stage for the classification of MI-based EEG. Before the classification of signals is done, there are many features that do not provide any extra information than the currently selected features and are known as redundant features. Feature selection ranks the extracted feature based on information content that each feature adds to classify the two classes. As a result, it removes redundant features and improves the computational cost of the system. In this study, the Random forest based feature selection method is used. This selection method consists of a few hundred decision trees and each decision tree is built using a random extraction of the observations from the dataset and a random extraction of the features. The trees are de- correlated and less prone to over-fitting as every tree does not see all the features or all the observations. Based on a single or combination of features each tree represents a sequence of yes- no questions. At every node, the dataset gets divided into 2 groups, each of them consisting of observations that are more similar among themselves and different from the ones in the other group. Therefore, the importance of each feature is derived from how "pure" each of the group is. When a tree is trained, it is possible to measure the decrease in an impurity by each feature and consequently, the more important feature. In random forests, the final importance of the variable can be determined by aggregating (majority vote) the impurity decrease from each feature across several tees. For classification, the measure of impurity is either Gini impurity or entropy (Information gain). The ranked features using random forest are shown in Fig 4. It can be seen that Mean Absolute Value (MAV) has the highest feature importance while Kurtosis has the lowest. This suggests that MAV and Energy should be the best features while Skewness and Kurtosis depict the redundant features.

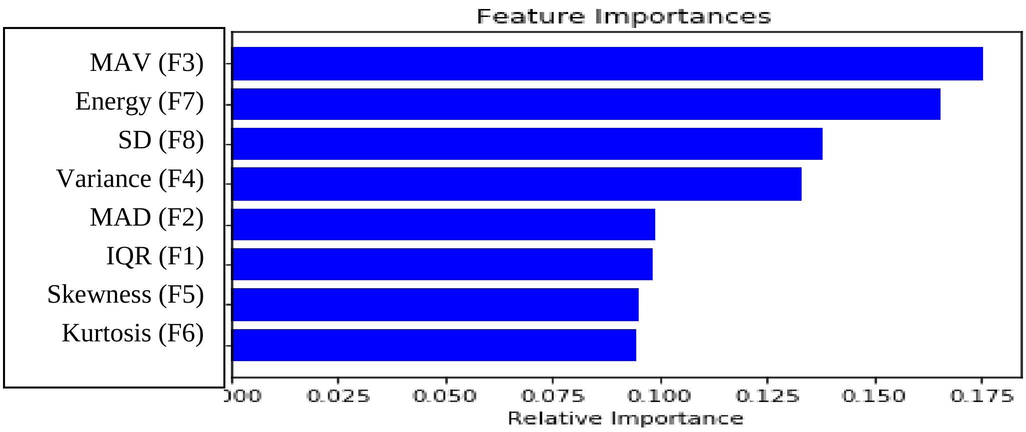


Figure 4. Showing relative feature importance

3. CLASSIFICATION

The classification of EEG signals plays a vital role in biomedical research. According to [15], there are mainly 5 types of classifiers used in BCI research such as linear classifiers, nonlinear classifiers, neural networks, nearest neighbour classifiers and a combination of these.

In this work, the Random Forest algorithm is employed for classifying EEG signals. Random forest is an ensemble learning algorithm. In this algorithm, at each node of the tree, we randomly select some subset of the features $s \subseteq S$, where 'S' is the set of features. The node then splits on the best feature in's' rather than 'S'. To decide on which feature to split is oftentimes the most computationally expensive aspect of decision tree learning. By narrowing the set of features, we drastically speed up the learning of the tree. The majority voting of the classification trees that have been formed obtains the prediction of the classification.

4. RESULTS

The wavelet transform analysis was performed on the dataset and the feature importance of the different extracted features was calculated using the Random Forest Algorithm in Scikit-learn followed by the classification of the selected features into two classes. We have used classification accuracy in order to evaluate the effectiveness of our method.

$$Classification\ accuracy\ (\%) = \frac{TP + TN}{TP + FP + FN + TN} \quad \text{(i)}$$

Where,

TP is True Positive;
TN is True Negative;
FP is False Positive;
FN is False Negative

Table 2 shows the combination of different features according to their relative importance, and with the classification accuracy of those combined features according to the relative importance. Therefore, we constructed eight different combined feature models with a different number of features to obtain the classification. In Table 2, F3 is shown to have the highest feature importance, hence selected first. After which F3 an F4 are combined, then F3, F4 and F8 and so on. From Table 2, it can be seen that though the feature F3 has the highest relative importance among the extracted features, it did not capture the significant distinctive information. However, combining F3 and F4 achieved the highest classification accuracy of this method. This suggests that feature F4 compliments the information captured by F3 and enhanced accuracy. Similarly, when 5 features are selected it achieved a similar accuracy of 85.41 as reported for 2 features.

Table 2. Showing different combination of features with their ranks according to feature selection method and the corresponding accuracies

No. of selected features	Feature name	Classification Accuracy (%)
1	F3	62.5
2	**F3, F4**	**85.41**
3	F3, F4,F8	76.14
4	F3, F4, F8,F2	63.85
5	**F3, F4, F8, F2, F7**	**85.41**
6	F3, F4, F8, F2, F7, F1	76.97
7	F3, F4, F8, F2, F7, F1, F6	78.64
8	F3, F4, F8, F2, F7, F1, F5	72.5

5. DISCUSSION

In this study, statistical features such as Mean Absolute Value, Median Absolute Deviation, Skewness, Kurtosis, Interquartile Range, Standard Deviation, Variance, and Energy were used to extract the underlying information from a dynamic EEG. This study has shown that the proposed features were successful in capturing the relevant distinguishing information. Also, it can be seen that the different combined features #2 and different combined features #5 have the same accuracy while #5 uses 5 features and #2 uses 2 features. This shows that good accuracy can be observed by using a lesser number of features and as a result improving the computational cost of the method. On the other hand, it could very well be true that using two features to represent the signal class may provide good classification accuracy but it could be at the cost of some vital information. Therefore, a future study may look into the choice of selecting lower number of features and the cost of crucial information lose. So, a follow up of this study would be selecting two features gave us good enough accuracy but whether crucial information was lost or not.

The choice of Random forest as a classifier has added robustness to the method. The proposed method can handle low dimensional as well as high dimensional data. However, it was observed that for high dimensional data, the method works slower but gives the same accuracy. Therefore, to reduce computational complexity, features were selected before the classification.

The application of this method in the near future is that it can be used to control an external device i.e. Neuro-prosthetics. The translated commands will be used as input to the external device via a computer (or micro-controller). This will, in turn, provide basic operations of the device. This study could also be used in the supervision of a trained physiotherapist to provide functional restoration to patients with spinal cord injury. In addition to that, this method can also be used in sports Biomechanics.

REFERENCES

[1] S. R. Guntur, R. R. Gorrepati, and V. R. Dirisala, "Robotics in Healthcare : An Internet of Medical Robotic Things (IoMRT)" Perspective. Elsevier Inc., 2016.

[2] M. Radman and A. Chaibakhsh, “Generalized Sequential Forward Selection Method for Channel Selection in EEG Signals for Classification of Left or Right Hand Movement in BCI,” ICCKE, pp. 137–142, 2019.

[3] N. E. Isa, A. Amir, M. Z. Ilyas, and M. S. Razalli, “Motor imagery classification in Brain- computer interface (BCI) based on EEG signal by using machine learning technique”, Bulletin of Electrical Engineering and Informatics, vol. 8, no. 1, pp. 269–275, 2019.

[4] E. Erkan and I. Kurnaz, “A study on the effect of psychophysiological signal features on classification methods,” Measurement, vol. 101, pp. 45–52, 2017.

[5] J. Wang, “EEG signal classification with feature selection based on one-dimension real- valued particle swarm optimization,” International Conference on Mechatronics, Control and Electronic Engineering, pp. 310–314, 2014.

[6] K. Aswinseshadri and V. T. Bai, “Feature Selection In Brain-Computer Interface Using Genetics Method,” 2015 IEEE Int. Conf. Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput., pp. 270–275, 2015.

[7] C. K. S. G. Ponnambalam, “Binary and multi-class motor imagery using Renyi entropy for feature extraction,” Neural Comput. Appl., vol. 28, no. 8, pp. 2051–2062, 2017.

[8] K. Venkatachalam, A. Devipriya, J. Maniraj, M. Sivaram, A. Ambikapathy, and I. S. Amiri, “A novel method of motor imagery classification using EEG signal”, Artificial Intelligence In Medicine, vol. 103, no. December 2019, 2020.

[9] R. Chatterjee, T. Maitra, S. K. Hafizul, and M. Mehedi, “A novel machine learning-based feature selection for motor imagery EEG signal classification on the Internet of medical things environment,” Futur. Gener. Comput. Syst., vol. 98, pp. 419–434, 2019.

[10] Liu, A., Chen, K., Liu, Q., Ai, Q., Xie, Y., Chen, A.: Feature Selection for Motor Imagery EEG Classification Based on Firefly Algorithm and Learning Automata. Sensors. 17, 2576 (2017).

[11] M. Witkowski, M. Cortese, M. Cempini, J. Mellinger, and N. Vitiello, “Enhancing brain- machine interface (BMI) control of a hand exoskeleton using electrooculography (EOG)”, Journal of NeuroEngineering and Rehabilitation, pp. 1–6, 2014.

[12] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Clin. Neurophysiol., vol. 113, no. 6, pp. 767–791, 2002.

[13] D. J. McFarland, L. a Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements.," Brain Topogr., vol. 12, no. 3, pp. 177–186, 2000.
[14] G. Pfurtscheller, A. Stancák, and C. Neuper, "Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review," Int. J. Psychophysiol., vol. 24, no. 1, pp. 39–46, 1996.
[15] H. Bashashati, R. K. Ward, G. E. Birch, and A. Bashashati, "Comparing Different Classifiers in Sensory Motor Brain-Computer Interfaces", PLOS, pp. 1–17, 2015.

MULTIMODAL DATA EVALUATION FOR CLASSIFICATION PROBLEMS

Daniela Moctezuma[1]and Víctor Muníz[2]and Jorge García[2]

[1]Centro de Investigación en Ciencias de Información Geoespacial, Aguascalientes, Ags., Mexico
[2]Centro de Investigación en Matemáticas, Monterrey, Nuevo León, Mexico

ABSTRACT

Social media data is currently the main input to a wide variety of research works in many knowledge fields. This kind of data is generally multimodal, i.e., it contains different modalities of information such as text, images, video or audio, mainly. To deal with multimodal data to tackle a specific task could be very difficult. One of the main challenges is to find useful representations of the data, capable of capturing the subtle information that the users who generate that information provided, or even the way they use it. In this paper, we analysed the usage of two modalities of data, images, and text, both in a separate way and by combining them to address two classification problems: meme's classification and user profiling. For images, we use a textual semantic representation by using a pre-trained model of image captioning. Later, a text classifier based on optimal lexical representations was used to build a classification model. Interesting findings were found in the usage of these two modalities of data, and the pros and cons of using them to solve the two classification problems are also discussed.

KEYWORDS

Multimodal Data, Deep Learning, Natural Language Processing, Image captioning.

1. INTRODUCTION

Nowadays a large amount of data is generated by users on the Internet, particularly on social media platforms. This shared information usually represents feelings and opinions about social events, people or products; thus, we can also find humorous, sarcastic, offensive, motivational content, among others. The type of data shared is usually unstructured, which means, they do not have a well-defined order or organization and the relationship between their characteristics or variables is not clearly determined. Examples of them are text, images, video or audio. When different types data are involved, one can say we have multimodal data. In fact, our experience as human being is multimodal because we see objects, hear sounds, feel the texture, smell odors, and taste flavors [1]. In this way, the shared information cannot be stored in a traditional relational data structures, since this cannot be treated as usual in relation SQL architectures.

Multimodal data has become very interesting in the machine learning (ML) research community due to many classification problems could be solved by using more than a single modality of the data [2][3]. For instance, when the task is to classify sentiment on twitter users, we would expect that using images and text could be better than using only text. One of the main challenges of using multimodal data relies in the search for the optimal representation of all modalities of

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 35-47, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112105

information combined, where a supervised or non-supervised ML model can be used efficiently and effectively. Sometimes, one could think that the more information the better, but it is not always the truth. The quality and singularity of the data are more important usually.

When social media data is analysed, many insights and applications could be helpful, for instance, determining the attitude over products [4], politics [5], health care [6], events [7], or comments as protests [8]. In many of these problems, not only the text is shared, but also images or videos, and dealing with different modalities of information becomes more difficult but the idea of using the multimodal version of the data is very attractive most of the time.One of the most used platforms of social media adopted by the scientific community is Twitter, not only by computer science researchers, but also from other research areas such as economics, health, environmental studies, and many more.

With the aim of analysing if using two modalities of information for classification problems is better or not, in this paper we address two classification problems where images and text are available. We present a comparison of the solutions obtained by using a support vector machine (SVM) as the classifier and a specific representation of the input, that means of only text, only images, and a combination of both. This comparison provides a measure of how the performance increased or decreased according to the modality of information we used and also with the problem tackled.

Specifically, the contributions of the presented manuscript are the following:

- We obtained a semantic representation of images based on Deep Learning (DL) models, specifically, an Image Captioning model is implemented with visual attention to describe images data. That is, translate the image to text in a semantic way.
- The use of multimodal information in classification tasks is evaluated with experiments using only text, only images, and a combination of both.
- The evaluation was done with two difficult classification problems, tackled usually only with text data: user profiling and memes polarity classification.
- We made an implementation for the Spanish language based on an image captioning corpus available in English using automatic translation tools. In this way, we can use our proposed model for classification problems in Spanish and English languages.

The paper is organized as follows. Related Work Section describes the related work. In Datasets Section the data we used as well as the two classification problems we tackled are described. Methodology Section details all the steps of the proposed methodology, that means, image captioning problem, text representation, and text and image representation for the classification problems. Experiments and results are given in Experiments Section, and finally, the main findings and conclusions are described in Conclusions Section.

2. RELATED WORK

In recent years, there has been an increasing number of research efforts which aims to model and combine the information from each modality of data to tackle specific tasks. State of the art (SOTA) results for many tasks related to multimodal data are based on deep learning (DL) architectures and natural language processing (NLP) models. First research works on that field, combines information from images and text, and prior to the popularization of DL models, most of the research relied on feature engineering on both modalities of information, such as filters (e.g., Gabor or Sobel) for images, and lexical features based on n-grams for text, which were combined based on heuristic rules [9] or simple concatenation of the single-modality features [10]. In both researches, it was shown that the performance of the final classifiers improved significantly when multimodal data was included instead of using just one modality. Popular tasks where image and text has been used are sentiment analysis and author profiling, promoted by the PAN@CLEF [11] [12] and SemEval [13] where SOTA results were achieved using deep encoder-decoder architectures, for both, images and texts, where specialized neural networks are used to extract features automatically, such as pre-trained convolutional neural networks, CNN [14] [15] for images and recurrent neural networks based on long short-term memory, LSTM [16], gated recurrent units, GRU [17] or transformer-based architectures such as BERT [18], for text sequences, where word embeddings are the common vector representation for words [19], [20], [13]. Once obtained the representation for each modality, they are combined generally with early or late fusion techniques [21], [22]. After that, a classifier such as neural network (NN), SVM or random forest (RF) is used, where the input is the whole multimodal representation.

There are other interesting applications where multimodal data has been used. In [23], a classifier for skin lesion is proposed based on multiple imaging modalities (macroscopic and dermatoscopic) and patient metadata such as age, sex, location of the disease, change of lesion, among others. Features for images were obtained with two pretrained CNNs based on ResNet-50 architecture [24], which are concatenated with patient metadata, followed by late fusion with a fully connected NN with two-hidden layers and a five-class softmax output layer in order to combine all features and obtain the corresponding lesion category. The authors show that the multimodal classifier outperforms one based on a single image. Multimodal data has been used extensively in different tasks related to music collections such as genre and mood classification or information retrieval. In this case, it is assumed that there are three main modalities of information [25] which can be associated to musical items: editorial (date of production, genre, composer, country of origin, etc.), cultural, (knowledge produced by the environment or culture, gathered from user profiles, web scraping or collaborative filtering) and acoustic (beat, tempo, rhythm, energy, and music structure). Multimodal information has been used for multi-label genre classification in [26], including cover art images, spectrograms from audio signal and customer reviews, where DL and NLP techniques are used for each modality of data. In [27], a proposal for tag-based music retrieval based on metric learning is presented, where the main idea is to create a shared embedding space based on acoustic and cultural embeddings obtained from Mel spectrograms and a user-song interaction matrix, respectively, in such a form that similarities between music items can be obtained.

3. DATASETS

Several datasets were used in this work. First, the Flickr30K [28] is used for image captioning task, i.e., textual descriptions of images.

For testing the multimodal classification approach, two more datasets related with the two classification problems were considered. These datasets correspond to Memotion Analysis SemEval2020 competition [13], and CLEF author profiling competition [12].

Flickr30k

For the image captioning model, we used the Flickr30k corpus, which consists of 158,915 captions from multiple sources that describe 31,783 images. These captions are in English language and the image's content is focused on people involved in everyday activities and events. These images are obtained from Flickr, a website that allows you to store, order, search, sell and share photographs or videos online, through the Internet.
Figure *1* shows some examples from the Flickr30K dataset.

Gray haired man in black suit and yellow tie working in a financial environment.
A graying man in a suit is perplexed at a business meeting.
A businessman in a yellow tie gives a frustrated look.
A man in a yellow tie is rubbing the back of his neck.
A man with a yellow tie looks concerned.

A butcher cutting an animal to sell.
A green-shirted man with a butcher's apron uses a knife to carve out the hanging carcass of a cow.
A man at work, butchering a cow.
A man in a green t-shirt and long tan apron hacks apart the carcass of a cow while another man hoses away the blood.
Two men work in a butcher shop; one cuts the meat from a butchered cow, while the other hoses the floor.

Figure 1. Images with captions from Flickr30k

In order to obtain captions in the Spanish language, a new version of Flickr30K was generated. This new version is just a translation of the captions to Spanish, which allowed to train a model to generate captions in Spanish.It is important to have a model for the Spanish Language because one of the two classifications task used in this work contemplates a task in Spanish. In Figure *2* we show some captions translated into the Spanish language.

image	...	caption_spanish
1000268201.jpg	...	Un niño con un vestido rosa sube las escalera...
1000268201.jpg	...	Una niña con un vestido rosa en una cabaña de...
1000268201.jpg	...	Una niña subiendo las escaleras hasta su casa...
1000268201.jpg	...	Una niña que sube a una casa de juegos de mad...
1000268201.jpg	...	Una niña entrando en un edificio de madera
1000344755.jpg	...	Alguien con una camisa azul y un sombrero est...
1000344755.jpg	...	Un hombre con una camisa azul está parado en ...
1000344755.jpg	...	Un hombre en una escalera limpia la ventana d...
1000344755.jpg	...	hombre de camisa azul y jeans en escalera lim...
1000344755.jpg	...	un hombre en una escalera limpia una ventana
1000366164.jpg	...	Dos hombres uno con una camisa gris uno con u...
1000366164.jpg	...	Dos chicos cocinando y bromeando con la cámara
1000366164.jpg	...	Dos hombres en una cocina cocinando comida en...
1000366164.jpg	...	Dos hombres están en la estufa preparando com...
1000366164.jpg	...	Dos hombres están cocinando una comida

Figure 2. Examples of translated text to Spanish from the Flickr30k dataset

Memotion Analysis

Memotion analysis [13] is an academic competition organized by SemEval (International Workshop on Semantic Evaluation) in 2020. The purpose of this competition was to provide a dataset including text and image, to tackle the problem of meme's classification, i.e., to detect for example when a meme is offensive or not. Meme's classification is far more difficult than the classification of text (for instance tweets), because a meme contains both, text and images to communicate its content. Nowadays there is not much attention to sentiment analysis in memes. When this task was launched, the objective was to attract the attention of the scientific community towards the automatic processing of memes shared on the internet on platforms such as Facebook, Instagram, and Twitter. Memes are difficult to deal with because they are often derived from our social and cultural experiences, such as television series or popular cartoon characters, and, as is stated in [13], "*these digital constructs are so deeply ingrained in our internet culture that to understand the opinion of a community, we need to understand the type of memes it shares*".

The dataset includes 7,000 images for training and 1,000 images for testing. For each meme, the text contained in the image is also shared. Then, we have the original image of the meme and the correct text extracted from it.
Figure *3* shows some examples of the Memotion Analysis dataset. In these examples, one has the image and the associated text.

Figure 3. Example of memes from the Memotion Analysis dataset

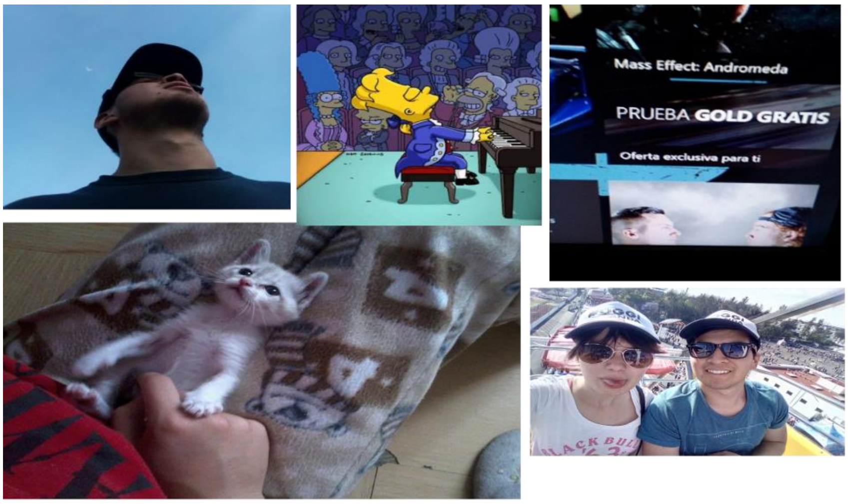
Author profiling

Author profiling is a task proposed by PAN[1], which is a series of scientific events and shared tasks on forensic digital text and stylometry and which in turn belongs to CLEF[2](Conference and Labs of the Evaluation Forum).Author profiling task tries to distinguish between classes of authors studying their sociolect aspect, that is, how language is shared by people. This helps to identify characteristics such as gender, age, native language, or personality type. For this task, the focus is on social networks, since it is of great interest to get insight of how everyday language reflects basic social and personality processes. Specifically, the data is generated by Twitter users.

The data used in this work is from the 2018 competition, which focused on gender identification on Twitter, where text and images are shared as sources of information. The languages addressed were English, Spanish, and Arabic. But, for this work, only the data in the Spanish and English languages were considered. In this dataset, there are 100 tweets and 10 images per user, as well as the gender for each of the 3,000 authors, in each language.

Figure *4* shows some examples of the data shared for this task [12].

[1] https://pan.webis.de/
[2] http://www.clef-initiative.eu/




```
[Bellos recuerdos xD https://t.co/wh07ILJwvJ]]></document>
[@Jokereando ojalá tuviera tu imaginación, yo solo veo una puta esponja :c]]></document>
[@Jokereando XD yo si le atiné cuando me preguntaron, bueno, dije esponja en forma de huevo, pero Masha? Cuerpos sin piel? Xdddd.]]></document>
[Era alguien con quien siempre podía hablar de lo que fuera, así sin miedo a ver su mensaje y no saber qué contestar, siempre había algo.]]></document>
[No hay nada más falso que un "jajaja" contestado tres dias después, ya vio tu msg, ya lo ignoró, y ya se aburrió de que nadie más le habla.]]></document>
[El boxeo de sombra es lo más difícil del mundo, siempre que intento pegarle, la maldita se quita y le doy a la pared :'c #fitness]]></document>
[Mi vida es tan triste que hasta la de soporte técnico de Xbox me decepciona :'v #XboxOne #GoW4 #microsoft https://t.co/k0sl2yepMA]]></document>
[Justo después de comprar gold y canjear el código, hijo de tu puta madre #BillGates  ¿cuántos más? https://t.co/Ylzf13NlV2]]></document>
[@aquilesNfierro @PeaPoblano @borre10000  que sigue siendo mejor que cualquier cosa que el PRI pueda ofrecer...]]></document>
[¿Será que todos somos igual de ojetes y que la diferencia sea el nivel de hipocresía de cada quien?]]></document>
[Tal vez me tratarías con más respeto si supieras que vendí mis dólares para poder salir contigo, sabiendo que iban a subir así de cabrón. 😭]]></document>
["Estamos habituados a que los hombres hagan burla de lo que no entienden, y murmuren a la vista de lo bueno y lo bello". #Goethe]]></document>
["El mundo de los espíritus no está cerrado; tu sentido está obtuso, tu corazón está muerto." #Fausto #Goethe]]></document>
[Y lo peor es que de seguro muchos de estos vándalos pendejos saben que sólo afectan al pueblo, pero igual quieren sus cosas grátis.]]></document>
[No es bueno dejarse arrastrar por los sueños y olvidarse de vivir.]]></document>
[No es que te deseé el mal, pero ojalá te salgan puros Pinsir en tus huevos de 10 km.]]></document>
[@Jokereando cómo se llama esa serie?]]></document>
[@lopezobrador_ Es triste ver cómo la gente comenta tanta mierda, ni siquiera criticas constructivas, tienen el país que merecen y nos llevan]]></document>
[El que vence modificando su estrategia, según la condición de su enemigo, debe ser tomado como un experto en #ElArteDeLaGuerra.]]></document>
[Si no me quisiste cuando manejaba un Platina, no me busques ahora que tengo un Camaro B| ... Es lo que diría si tuviera un Camaro :'v xd]]></document>
[Ese Euler andaba en todo, me pregunto si tenía amiguitos.]]></document>
[@Terafobia si la frase es buena, ¿qué importa? Habrá quien la use porque YOLO y quien la use porque sí la entiende xd.]]></document>
[¿Por qué las mujeres son tan mierda?(Excepto Dalia mi BF) ¿Debería probar la promiscuidad? ¿La homosexualidad? ¿O la promiscuidad bisexual?]]></document>
[¿Han tenido esa sensación de que están en un lugar y/o momento en el que no deberían estar?]]></document>
[No busques perfección, pero sí a alguien que mínimo te respete. Mereces algo mejor. ;)]]></document>
[Ojalá me gustara practicar el deporte tanto como me gusta verlo #Río2016]]></document>
[Tu infancia terminó cuando te diste cuenta que las palabras "heroina" y "erotico" no tienen que ver con superheroes, o que si...]]></document>
[No es lo mismo que te digan: "Voy a estar para lo que necesites", a: "Estaré siempre para ti, para lo que sea".]]></document>
[Dios bendiga al código abierto! :'D]]></document>
[Tan culpable es el que inicia el problema como el que no hace nada por resolverlo.]]></document>
[Antes de quejarte de la actitud de alguien más, mírate a ti.]]></document>
[No siempre se puede fingir una sonrisa para alegrar a alguien, a veces lo mejor que puedes hacer es no mostrarle lo agüitado que estás.]]></document>
[Perder duele, pero cada vez menos]]></document>
[A veces pienso que los sentimientos son nada más que un estorbo.]]></document>
[No importa que nadie lo note, tú sabes que hiciste algo bueno por alguien más :'(]]></document>
[Una simple vicisitud o una completa tribulación, la mayoría de las veces puedes decidir como ver a tus problemas.]]></document>
[No soy adepto a expresar mis emociones ni decir lo que pienso abiertamente... y no me harás empezar hoy. A menos claro, que esto cuente :P]]></document>
[Lección de vida, cortesía de House: "Trata a todo el mundo como si tuviera síndrome de Korsakoff". #TodosMienten]]></document>
[#CocaColaZeroAntártida tocan en la antartida.... no tocan trapped under ice.... me dejaron con las ganas]]></document>
[#CocaColaZeroAntártida es impresionante, me encanta, pero me pregunto porque siempre tienen que tocar las mismas]]></document>
[@Jokereando no se viejo tengo tarea y mi gatita sufre de ataques de nuevo]]></document>
[@Jokereando estoy aqui los fines de semana, holis]]></document>
[@CartoonMX yo viendo la peli de la fusion de goku y vegeta a las 3am y a media pelicula ponen el final de los guerreros de plata!!!]]></document>
[@Jokereando XD pues ven por tus cosas]]></document>
[@Arturo_582 que es eso?? XDD]]></document>
[@arturo_582 se fue mi internet, yo imprimo , sabes el apellido de poulet?]]></document>
```

Figure 4. Examples of images and text shared in the author profiling 2018 dataset

4. PROPOSED METHODOLOGY

The proposed methodology consists of two steps. First, we obtain a textual semantic representation of images by training an image captioning model with the Flickr30K dataset on English and Spanish languages. In the second step, we fuse vector representation (obtained with NLP techniques) of the two modalities of our data by concatenating previously the text information from our data and the one obtained in the first step. On this shared representation, we train two classification models based on SVM to solve the memes classification problem and the user profiling task.

Figure *5* shows a schematic overview of the methodology, and the details are described in the following subsections.

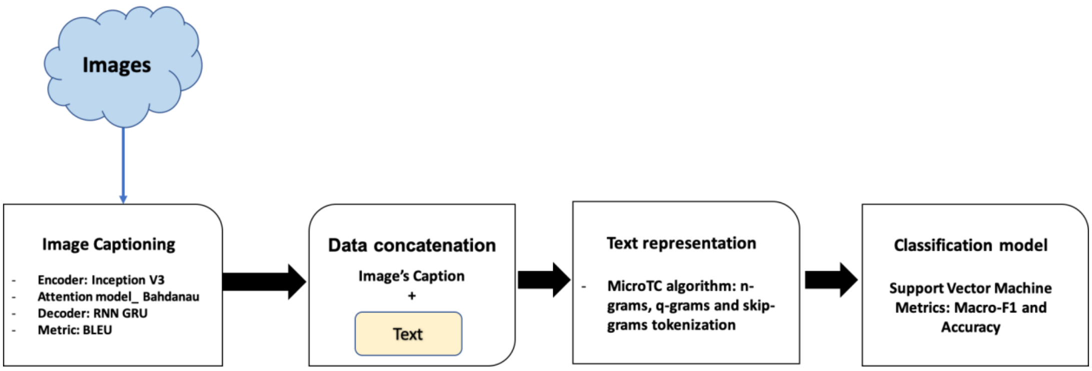


Figure 5. Scheme of our Proposed methodology

Image captioning step

Image captioning is the task which attempts to describe, in a semantic way, an image content. SOTA results for this task are achieved by deep encoder-decoder architectures, as the one we used, which is based on [29]. The encoder consists on a pre-trained CNN based on the Inception-v3 architecture [30] with visual attention [31], in order to relate, or "align" some specific objects of an image with its corresponding text descriptions given in the training data. The output of the encoder is a vector representation of the image, which in turn, is the input of the decoder, which learn to generate an output sequence which is the textual description of the image. In our case, the decoder is based on a recurrent neural network (RNN) with gated recurrent units (GRU). The image captioning architecture is shown in

Figure *6*. As we said before, our training corpus for this task is Flickr30k dataset, described in Flickr30k Section.

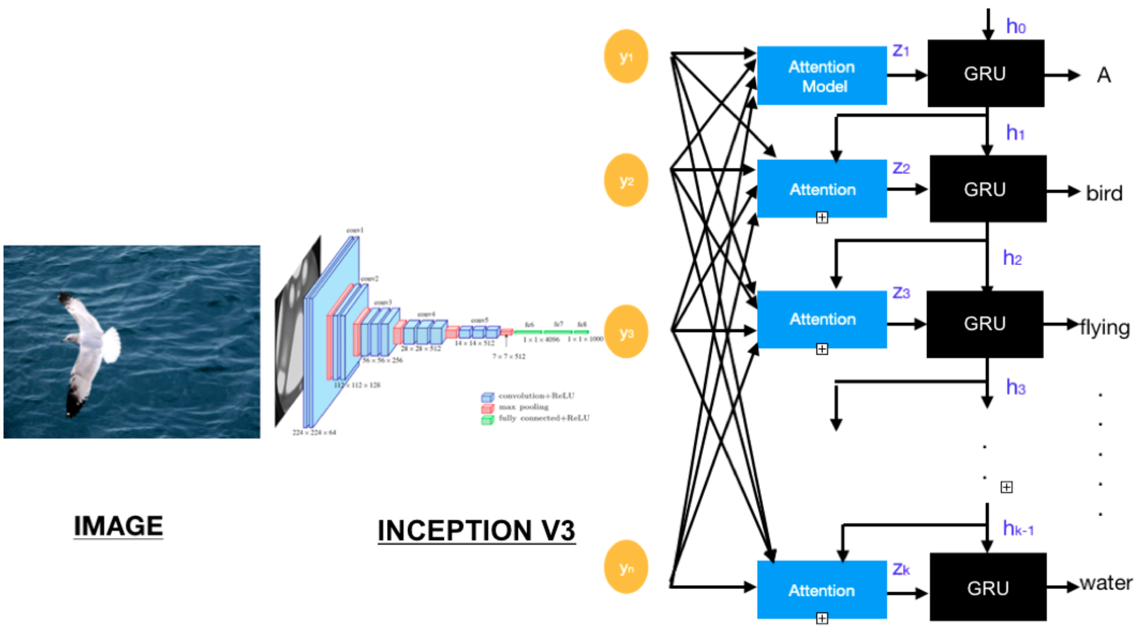


Figure 6. Image Captioning model[3].

[3] Source: https://medium.com/swlh/image-captioning-using-attention-mechanism-f3d7fc96eb0e

Multimodal fusion and classification

Once we have the text which corresponds to the image description and the original text provided in the dataset, multimodal fusion is carried out by learning a vector representation of the concatenated text from both modalities of data. To this end, we used μTC algorithm [32], which is defined as a minimalist text classifier based on SVM, robust to any language and domain. The main idea in [32] μTC is searching for an optimal text representation based on a set of text transformations such as noise deleting, normalization, and tokenization, among others. The optimal representation is the one which has a good performance in a given classification problem. The procedure can be viewed as a combinatorial optimization problem wherein each algorithm iteration, the performance of the parameter's configuration is measured trying to select a better configuration in the next step until the best possible solution is reached. As final step, the vector generated with μTC is used as the input to a SVM classifier. As performance measures, we used macro-F1 and Accuracy metrics.

5. EXPERIMENTS AND RESULTS

In this section, several experiments are presented. First, the performance of the image captioning task is assessed. In this case, we used the bilingual evaluation understudy (BLEU) metric (please see [33] for technical details), which measures the grammatical composition of sentences with n-grams in order to evaluate if the candidate caption obtained, captures the meaning of the reference caption given in the dataset. A key aspect in BLEU is the n-grams considered, which refer to a sequence of words within a window, where n represents the size of the window. For example, for the sentence "yesterday I went to the park to run"the unigram (n=1) represent each word, while for n = 2 we have bigram: "yesterday I", "I went", "went to", "to the", "the park", "park to", "to run". Thus, in BLEU the n-gram of the candidate caption is compared with the n-gram of the reference caption. It is worthwhile to mention that BLEU does not consider the position of the n-grams in the text but the number of matches. In the case of our image captioning model trained with Flickr dataset, the BLEU score obtained based on a 20% testing dataset and unigram were 42 and 38.7 for Spanish and English languages, respectively, which are considered as high quality and good results. In
Figure 7 *and*
Figure *8* we show some representative examples of the image captioning model results. We can see that good results are obtained in both languages, but outstanding results are observed for Spanish.

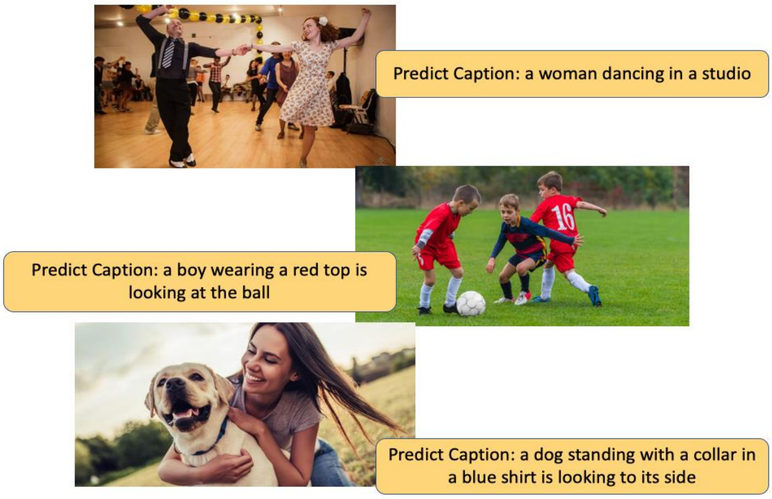


Figure 7. Example of captions in English generated by the image captioning generated model

Figure 8. Example of captions in Spanish generated by the image captioning with our generated dataset (Spanish dataset)

Once we obtained our model for image captioning, we proceed to assess the meme classification and author profiling tasks. For the evaluation, we used macro-F1 and Accuracy metrics.In the evaluation of meme's classification, we tackled two out of three tasks related to the Memotion analysis contest. In task A (polarity classification), the objective is to classify a meme content as positive, negative, or neutral, meanwhile in Task B (humour classification), the objective is to classify a meme as sarcastic, humorous, offensive or motivator, and further, a meme can be classified in more than one category, making this task very difficult.

Two baselines were provided for the Memotion analysis competition [13]. For task A, macro-F1 was 0.2176, and for task B 0.5118. Furthermore, the best results obtained by competitors of the Memotion analysis were 0.3546 for task A, and 0.5183 for task B, both in macro-F1 (see [13] for all results).

Table 1shows our results for meme classification in both tasks. In Task A, we can see very good results in the training stage, but not so good in testing. By using only text information, a macro-F1 of 0.955 is obtained but dropped dramatically to 0.854 in testing. This could be for overfitting in the training stage. When we used Text + Caption, the performance was similar, reaching a 0.976 of accuracy, higher than using only text, but in testing the performance also dropped. In the case of Task B, lower results were reached using Text or Text + Caption. Although we obtained better results in Task-A versus the best result of the competition, in Task-B we achieved a very lower result compared again with the best of the competitors. Our objective is to analyze if our generated caption improves the classification performance, in this case, we can see this has not happened.

Table 1. Memes classification results (macro-F1)

Task	Data	Training	Test	baseline	Best score competition
Task-A	Text	0.955	0.854	0.2176	0.3546
Task-A	Text + Caption	0.976	0.740		
Task-B	Text	0.437	0.380	0.5118	0.5183
Task-B	Text + Caption	0.444	0.369		

For the author profiling task, we used the dataset reported in [12] for English and Spanish languages. Because this competition was held in 2018, we had no access to the baseline nor testing dataset, so, we split the dataset into 80% for training and 20% for testing. For reference, we show in Table *2* the best results for Spanish and English achieved in the competition as was published at that time. The metric used here is accuracy.

Table 2. User profiling contest best results (accuracy)

Language	Data	Training	Test
Spanish	Text	-	0.8200
Spanish	Text + Caption	-	0.8200
English	Text	-	0.8221
English	Text + Caption	-	0.8584

The results reached with our proposed methodology are shown in Table 3. Here, it can be seen the perfect results in training for both languages, but a lower accuracy for test. In both cases, a better result was obtained by considering multimodal data, Text + Caption. Nevertheless, we consider that there is not enough evidence to demonstrate the advantage of including image information by means of its corresponding caption.

Table 3. User profiling results (Accuracy)

Language	Data	Training	Test
Spanish	Text	1.00	0.823
Spanish	Text + Caption	1.00	0.833
English	Text	1.00	0.735
English	Text + Caption	1.00	0.738

As an analysis, we can state that for the meme classification task, a better performance was only obtained in Task A, compared to the baseline reported in the competition. However, it was observed that the classification was better when using only the textual information in both tasks (A and B), and by combining both modalities of data (Text + Caption) it was not possible to achieve good generalization. Then, we could conclude that the descriptions of the images did not improve significantly the results regarding this task. One of the possible reasons for this situation is the composition of the dataset for this task since the sentiment contained in the meme is mainly identified with the textual message of the image, in addition, lower performance may occur in the model by adding the information of the images because the same image is used to create memes with different sentiments, i.e., the same image with different text could be an opposite sentiment perspective.

Respecting to the author profiling task, we obtained better performance when using the textual information combined with the image's description, compared to the performance obtained by classifying with only the user's textual information, in both languages. When we analyse the results reported in this competition, we found that the best result was achieved by using only text than using only images.In our case, although the improvement when using both modalities of information seems to be not so significant, an issue that must be taken into account is that for each user there were around 100 tweets and only 10 shared images and, based on the length of the tweets versus the descriptions generated for each image, the textual information of the added images represents only 9% of the final content for each user, resulting in a significant unbalance in the data modalities.

In both tasks, there are some issues regarding their datasets such as data distribution, topics, categories definition, almost the same data could represent different categories or classes, among others.Maybe, the extension of this analysis with more data could show a better understanding on the relevance of image captioning in this kind of problems.

6. CONCLUSIONS

Our main objective was to demonstrate if using the image captioning approach to use multimodal data could improve the classification results in both tasks, Meme classification, and User profiling.In general, the methodology applied for the classification of multimodal data had a good evaluation in both tasks but not significantly outstanding result to state that using captions in multimodal data improves the performance of the two classification problems tested.

Even so, to conclude on its performance compared to other approaches, we think it is necessary to apply in tasks with balanced data types, i.e., with the same number of images and texts for each sample, in order to analyse if the semantic description of the images provides significant information in the representation of the data. Our methodology did not obtain a good generalization with the description of the images in the classification for new data in the tasks addressed, this is reflected by excellent results in training that decrease in the test dataset, which is a topic of interest as future work, also, it is in our interest to explore another DL architectures, such as those based on transformers.

As future work, more fusion techniques could be applied to both, images and text data. Also, it will be interesting to test our methodology with more adequate datasets to state more solid conclusions and produce more accurate image captioning models to improve the image semantic descriptions.

ACKNOWLEDGEMENTS

This research is supported by the project A1-S-34811 of Basic Science grant by the National Council of Science and Technology (CONACyT) from Mexico.

REFERENCES

[1] Baltrušaitis, T., Ahuja, C., &Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence, 41(2), 423-443.
[2] Kumar, A., & Garg, G. (2019). Sentiment analysis of multimodal twitter data. Multimedia Tools and Applications, 78(17), 24103-24119.
[3] Alam, F., Ofli, F., & Imran, M. (2018, June). CrisisMMD: Multimodal twitter datasets from natural disasters. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 12, No. 1).

[4] Das, T. K., Acharjya, D. P., & Patra, M. R. (2014, January). Opinion mining about a product by analyzing public tweets in Twitter. In 2014 International Conference on Computer Communication and Informatics (pp. 1-4). IEEE.

[5] Jungherr, A. (2014). Twitter in politics: a comprehensive literature review. Available at SSRN 2865150.

[6] Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: a systematic review. American journal of public health, 107(1), e1-e8.

[7] Nichols, J., Mahmud, J., & Drews, C. (2012, February). Summarizing sporting events using twitter. In Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (pp. 189-198).

[8] Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. The Communication Review, 14(3), 197-215.

[9] Kalva, P., Enembreck, F., &Koerich, A. (2007, September). Web image classification based on the fusion of image and text classifiers. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 1, pp. 561-568). IEEE.

[10] Kannan, A., Talukdar, P. P., Rasiwasia, N., &Ke, Q. (2011, December). Improving product classification using images. In 2011 IEEE 11th International Conference on Data Mining (pp. 310-319). IEEE.

[11] Rangel, F., Rosso, P., Potthast, M., Trenkmann, M., Stein, B., Verhoeven, B., &Daelemans, W. (2014). Overview of the 2nd author profiling task at pan 2014. In CEUR Workshop Proceedings (Vol. 1180, pp. 898-927). CEUR Workshop Proceedings.

[12] Pardo, F. M. R., Montes-y-Gómez, M., Potthast, M., Stein, B., Ferro, N., Nie, J. Y., &Soulier, L. (2018). Overview of the 6th Author Profiling Task at PAN 2018: Cross-domain Authorship Attribution and Style Change Detection. In CLEF 2018 Eval Labs Workshop–Work Notes Pap (Vol. 10, p. 14).

[13] Sharma, C., Bhageria, D., Scott, W., PYKL, S., Das, A., Chakraborty, T., ... &Gamback, B. (2020). SemEval-2020 Task 8: Memotion Analysis--The Visuo-Lingual Metaphor!.arXiv preprint arXiv:2008.03781.

[14] LeCun, Y. (1989). Generalization and network design strategies. Connectionism in perspective, 19, 143-155.

[15] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[16] Hochreiter, S., &Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[17] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., &Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[18] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[19] Takahashi, T., Tahara, T., Nagatani, K., Miura, Y., Taniguchi, T., & Ohkuma, T. (2018). Text and image synergy with feature cross technique for gender identification. Working Notes Papers of the CLEF.

[20] Álvarez Carmona, M. Á., VillatoroTello, E., Montes y Gómez, M., &Vilaseñor Pineda, L. (2020). Author profiling in social media with multimodal information. Computación y Sistemas, 24(3), 1289-1304.

[21] Liu, K., Li, Y., Xu, N., & Natarajan, P. (2018). Learn to combine modalities in multimodal deep learning. arXiv preprint arXiv:1805.11730.

[22] Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. IEEE signal processing magazine, 34(6), 96-108.

[23] Yap, J., Yolland, W., &Tschandl, P. (2018). Multimodal skin lesion classification using deep learning. Experimental dermatology, 27(11), 1261-1267.

[24] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[25] Pachet, F. (2005). Knowledge management and musical metadata. Idea Group, 12.

[26] Oramas, S., Barbieri, F., Nieto Caballero, O., & Serra, X. (2018). Multimodal deep learning for music genre classification. Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21.

[27] Won, M., Oramas, S., Nieto, O., Gouyon, F., & Serra, X. (2021, June). Multimodal metric learning for tag-based music retrieval. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 591-595). IEEE.

[28] Young, P., Lai, A., Hodosh, M., & Hockenmaier, J. (2014). From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. Transactions of the Association for Computational Linguistics, 2, 67-78.

[29] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... &Bengio, Y. (2015, June). Show, attend and tell: Neural image caption generation with visual attention. In International conference on machine learning (pp. 2048-2057). PMLR.

[30] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., &Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[31] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[32] Tellez, E. S., Moctezuma, D., Miranda-Jiménez, S., & Graff, M. (2018). An automated text categorization framework based on hyperparameter optimization. Knowledge-Based Systems, 149, 110-123.

[33] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).

AUTHORS

DANIELA MOCTEZUMA is a researcher at the Research Center on Geospatial Information Sciences (CentroGEO) since 2014. She received her Ph.D. in Computer Sciences from Rey Juan Carlos University, Madrid, Spain in 2013. Her research interests include machine learning, computer vision, natural language processing, intelligent video surveillance systems, and remote sensing.

VÍCTOR MUÑIZ received a PhD in computer science at Research Center in Mathematics (CIMAT) in Guanajuato, Mexico. He is working at CIMAT Monterrey, Mexico, in liaison and research projects. His research interests are Machine Learning, Natural Language Processing and spatio-temporal models.

JORGE GARCIA Jorge Sánchez García received a bachelor degree in mathematics from Universidad Autónoma de Nuevo León (UANL) and a master degree in statistical computing from Research Center in Mathematics (CIMAT). His research interest is computer vision.

THERMAL COMFORT OF THE ENVIRONMENT WITH INTERNET OF THINGS, BIG DATA AND MACHINE LEARNING

Matheus G. do Nascimento and Paulo B. Lopes

Graduate Program in Electrical Engineering and Computing,
Mackenzie Presbyterian University, São Paulo, Brazil

ABSTRACT

This research proposes to evaluate the level of thermal comfort of the environment in real time using Internet of Things (IoT), Big Data and Machine Learning (ML) techniques for collecting, storage, processing and analysis of the concerned information. The search for thermal comfort provides the best living and health conditions for human beings. The environment, as one of its functions, must present the climatic conditions necessary for human thermal comfort. In the research, wireless sensors are used to monitor the Heat Index, the Thermal Discomfort Index and the Temperature and Humidity Index of remote indoor environments to intelligently monitor the level of comfort and alert possible hazards to the people present. Machine learning algorithms are also used to analyse the history of stored data and formulate models capable of making predictions of the parameters of the environment to determine preventive actions or optimize the environment control for reducing energy consumption.

KEYWORDS

Big data, internet of things, machine learning, thermal comfort.

1. INTRODUCTION

People spend more than 90 % of their day in indoor environments, such as offices, schools, colleges, commercial buildings, industrial buildings and homes. The concern with analysing the internal habitats seeks to increase people's quality of life and health as, under normal conditions, the environment should not cause feelings of discomfort and stress for human beings [1], [2].

Thermal comfort is a state of well-being that the organism experiences in a given environment. It is the result of the combination of climatological variables in the environment and, in cases of discomfort, it can have a direct impact on people's health and quality of life. Each human being can make a different judgment of the comfort state of the environment due to differences in the levels of heat exchange and production [3]. These differences are related to vascular responses, subcutaneous fat thickness, individual clothing habits, eating habits, the activity that the organism was or is performing and other conditions [4]. For World Health Organization (WHO), people are sensitive to changes of up to 0.5°C/h, that is, the greater the variation, the greater the proportion of discomfort [5].

Each location has its specific climatic factors and particularities that directly influence temperature and humidity. Therefore, it is possible to evaluate the suitability of the environment to human beings with basis on these two parameters [6].

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 49-66, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112106

Monitoring thermal comfort is very important in a work environment as it directly influences the development of daily tasks, impacting people's productivity during the workday [4], [5], [7]. The judgment of comfort is performed through the response of each organism to the environment. Nonetheless, in huge commercial and residential buildings, typical of modern cities, the volume of data to be collected, stored and analysed in both space and time dimensions prevents the use of traditional tools.

With the use of IoT and Big Data tools, it is possible to implement a project to classify, in real time, the level of discomfort in the environment using relationships between the air temperature (℃) and the relative humidity (%) collected in the studied surroundings. Through these parameters it is possible to calculate the thermal comfort indexes to assess the environmental conditions and alert places that may generate risks to the health of exposed people [8], [9]. The use of comfort indexes helps to assess the impact of the environment on the thermal state of humans to determine personal comfort.

The devices and tools implemented in this research use the smart home concept, providing an IoT module in each chamber of the habitat. Each module has a temperature and humidity sensor on an IoT mote connected to the Wi-Fi network that will perform an HTTP POST request to a PHP script to insert the sensor readings into a LAMP server installed on the Raspberry Pi 3B+. By collecting data from IoT devices, it will be possible to monitor and generate analyses that can contribute to thermal comfort. In these circumstances, it will be necessary to use Big Data techniques to store, process and visualize the large volume of information collected in both time and space, ensuring security and privacy for the user [10], [11].

The environmental parameter collection history is used to build and train machine learning models, using the MindsDB tool, which will be directly connected to the application database, to perform the environment parameter predictions through an SQL query. MindsDB automates and simplifies the steps for model development in order to speed up the analysis of your results, reducing the need for specialist development. During the elaboration of the algorithms, it was considered that the collected data may present imperfections, inconsistencies and be of different types in order to guarantee the security, trust and privacy of the information [12], [13], [14]

This paper leverages IoT, Big Data and Automated Machine Learning (AutoML) to enable predictive environmental analysis of the internal environment location to be used in corporate environments to assess the level of comfort and productivity of its employees. It can also be used in domestic households with a focus on elderly people who need special care, being possible to identify the conditions of the environment and alert possible dangers through devices connected to the internet, physically distant from the monitored environment. By leveraging the results of predictive analysis, one can anticipate events, take actions to avoid closed environment related issues and even optimize the use of the energy in Heating, Ventilating and Air Conditioning (HVAC) systems.

2. THERMAL COMFORT

According to American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), thermal comfort is a mental condition that expresses the body's satisfaction with the environment. For the organism to remain in thermal comfort it is necessary that the heat fluxes between the environment and the organism are null [3], [15].
Metabolism is responsible for maintaining an almost constant temperature in the body in a process that, through internal biochemical reactions combining organic fuels, produces energy for the body. The human body is always exchanging heat with the environment through the skin, that

is, the skin is the thermoregulatory organ of mankind. Skin temperature depends on the blood flow that is flowing through it, and the greater the flow, the higher the temperature [4].

The American Ventilation Commission conducted a study in 1916 to analyse the influence of thermo-hygrometric conditions on labour income. The research motivation was to analyse the effects on production caused by the Industrial Revolution and extreme situations of war. With the study, it was possible to determine that an increase of the ambient temperature in the factories from 20°C to 24 °C and from 20°C to 30 °C caused worker´s productivity to decrease by 15% and 28% respectively. Likewise, it was identified that with the increase in temperature in mines from 19°C to 27 °C, the miners' income decreases by 41%. Seasons impact industry production and thermally uncomfortable work environments have high incident rates [16].

The population's health and well-being are directly related to climatological variables, directly influencing people's behaviour. To analyse how much the population can suffer from thermal discomfort, the Heat Index (HI), Thermal Discomfort Index (TDI) and the Temperature and Humidity Index (THI) have been universally used [9], [17], [18] in industry, heath and civil engineering. The idea of calculating comfort indexes aims to relate several variables into a single parameter.

The Heat Index is a measure to define the intensity of heat a person feels, relating the air temperature (°C) with the relative humidity (%). The Thermal Discomfort Index and the Temperature and Humidity Index are also calculated using the air temperature (°C) with the relative humidity (%). They are important parameters to determine whether the environment provides thermal comfort for people, relating them to alert levels and physiological symptoms.

To calculate the Heat Index, Equation 1 was used, in which the HI is the Heat Index in °C, the T is the temperature in °C and the UR is the relative humidity of air in % [17, p. 517].

$$\begin{aligned} HI = -8.78469475556 + 1.61139411 \cdot T + \ 2.33854883889 \cdot UR \\ + (-0.14611605 \cdot T \cdot UR) + (-0.012308094 \cdot T^2) \\ + (-0.0164248277778 \cdot UR^2) + 0.002211732 \cdot T^2 \cdot UR \\ + 0.00072546 \cdot T \cdot UR^2 + (-0.000003582 \cdot T^2 \cdot UR^2) \end{aligned}$$

Equation 1. Heat Index Equation

In Table 1, the HI is related to danger level and physiological symptoms [17].

Table 1. Relation of HI with possible physiological symptoms

HI[°C]	**Danger level**	**Heat syndrome**
Less than 27°C	Absence of alert	Absence
From 27°C to 32°C	Warning	Possible fatigue in cases of prolonged exposure
From 32°C to 41°C	Very careful	Possibility of cramps, exhaustion and heat stroke for prolonged exposure and physical activity
From 41°C to 54°C	Danger	Cramps, heat stroke and burnout likely. Possibility of brain damage (CVA) for prolonged exposure with physical activity
Greater than 54°C	Extreme danger	Heat stroke and impending cerebrovascular accident (CVA)

To calculate the Thermal Discomfort Index, Equation 2 was used, in which the TDI is the Thermal Discomfort Index in °C, the T is the temperature in °C and the UR is the relative humidity in % [18, p. 37]. The relationship between TDI and danger level is shown in Table 2.

$$TDI = T - (0{,}55 - 0{,}0055 \cdot UR) \cdot (T - 14{,}5)$$

Equation 2. Thermal Discomfort Index Equation

Table 2. Thermal discomfort level as a function of the TDI

TDI[°C]	**Danger level**
Less than 21°C	No discomfort
From 21°C to 25°C	Less than 50% of the population feels discomfort
From 25°C to 28°C	More than 50% of the population feels uncomfortable
From 28°C to 30°C	Most of the population feels discomfort
From 30°C to 32°C	Everyone feels discomfort
Greater than 32°C	State of medical emergency

Equation 3 was used to calculate the Temperature and Humidity Index, in which the ITU is the Temperature and Humidity Index in °C, the T is the temperature in °C and the UR is the relative humidity in % [9, p. 154].

$$ITU = 0{,}8 \cdot T + \frac{UR \cdot T}{500}$$

Equation 3. Temperature and Humidity Index Equation

The relationship between ITU and danger levels is depicted in Table 3. For ITU values less than 21 °C the environment can be considered partially uncomfortable.

Table 3. Thermal comfort level as a function of the THI

THI[°C]	**Danger level**
From 21°C to 24°C	Comfortable
From 24°C to 26°C	Slightly uncomfortable
Greater than 26°C	Extremely uncomfortable

3. INTERNET OF THINGS, BIG DATA AND MACHINE LEARNING

The Internet of Things made it possible to interconnect the physical and digital world using smart devices, allowing the transmission and reception of collected information. The term "things" is used to define everyday equipment that uses hardware, software, sensors and actuators to collect parameters from the physical world so that they can be stored and processed on a local server or in the cloud. The IoT devices are limited in storage and processing due to the limited computing capacity of each device [19], [20].

The basic architecture of the IoT is composed of the physical, gateway and data processing layers. In the physical layer, all IoT devices are identified on the network and the information is collected and transmitted through the gateway, which will read and transmit the data obtained to the information management system, which is connected to the internet, making it possible to transmit the information to the cloud. The data processing layer is responsible for performing data analysis and mining, using Big Data Analytics concepts [11], [21].

Big Data is an abstract concept that emerged in 2010 and is based in the literature with the 5Vs: volume, velocity, variety, veracity and value. The first challenge is to work with the large volume of data, which is growing more and more, and it is necessary to use ways to treat this data in order to transform them into useful information [11], [16].

Data is stored for consultation and further analysis on storage servers, typically specific to IoT services. For storage, Big Data technologies are used, in addition to supporting the large volume of data, resources are offered for processing and viewing information. The servers also provide the time stamp for each stored data and visualization models using dashboards [11], [22].

The most common servers adopt relational databases that use Structured Query Language (SQL) to perform queries and manipulations on the data. The relational model presents a database composed of relations stored in tables with the respective values of each parameter. All information received is stored in a row of a table composed of specific columns for each variable. Not only SQL (NoSQL) database models are used, which offer greater performance and scalability for projects that use a large volume of information [19].

Many IoT applications are unable to handle data in real time due to computational limitations and latencies in the collection process. For real-time processing, agile software is needed, which can obtain information by means of data simultaneously. For the analysis of historical data, the information is already stored and available to use statistical techniques to obtain conclusions and relationships between the data and the context studied [22].

After performing the data analysis, it is possible to identify patterns and make some predictions of possible future results, using algorithms and statistical analysis. Cognitive analysis is present in projects that have learning mechanisms that can make decisions autonomously.

Machine learning is a data analysis method that automates the construction of analytical models, a branch of artificial intelligence based on the idea that systems can learn from data, identifying patterns with minimal human intervention. Algorithms are trained to abstract information and identify patterns in a large amount of data, to enable predictions and decisions to be made as the data is processed [23].

With a greater amount of data provided by big data, machine learning algorithms become more reliable and accessible, offering more efficient and intelligent machine learning. The learning model uses information from the dataset to formulate a model capable of performing prediction of system parameters [24], [25].

Before implementing the machine learning model, it is necessary to select and prepare the dataset to solve the proposed challenge. In a table, for example, each row represents a data point and its columns represent the characteristics that describe the data. Columns can be classified as features and can also contain labels. Features are used by the model to predict the labels, output attributes. After training, the user will provide a set of features for the model to return the expected label [12], [23], [24].

For data analysis it is necessary that the information is unique, randomized and also checked if there are trends or imbalances that could impact the training [23]. After this treatment, the data are divided into two sets, the training subset, used to train the model, and the evaluation subset, responsible for testing and refining the model. The set of algorithms to be applied varies according to the dataset attributes, the amount of data and what should be predicted [12], [24].

AutoML combines automation with machine learning, proposing to automate steps required for model development, enabling you to quickly build machine learning solutions using techniques that are easier to apply and reducing the need for specialists for development. This model proposes to streamline procedures so that it is possible to analyse their results as quickly as possible [12]. Diverse algorithms are applied to the problem and the best solution is selected, enabling the obtention of a fast result.

In recent years AutoML has become a much researched topic for addressing, in complement to traditional machine learning algorithms, applications in computer vision, data mining and natural language processing [12]. Machine learning automation only became possible due to the development of complex big data techniques on very fast computers and also due to the great demand for machine learning tools [24].

In the traditional machine learning model, the selection of the model and algorithms is done through human resources that abstract information from the available data using machine learning tools, based on professional knowledge. In addition, it is necessary to carry out resource engineering, consisting of planning and building resources from the data to abstract the maximum amount of information from the data set. Feature selection also removes unnecessary information, so models are more efficient, simpler, and interpretable. In the AutoML model these steps are automated by computer programs and are designed to be performed within a limited computational budget. Basic AutoML tools use a controller composed of an optimizer and an evaluator [12], [24], [25].

4. DEVELOPMENT

Based on the theoretical reference, it was possible to assess the level of comfort in the environment by measuring the indices of thermal comfort used. The assessment consists of judging the level of comfort in the environment, considering the level of danger and possible physiological symptoms for the individuals present.

For data collection regarding air temperature (°C) and relative air humidity (%), IoT motes were used. Each one, illustrated in Figure 1, is composed of a DOIT ESP32 – ESP-WROOM-32 board connected to a Wi-Fi network, a DHT11 temperature and humidity sensor.

The temperature and humidity collection modules were installed in the rooms of a household located in the metropolitan region of São Paulo to analyse the comfort levels of the place. The house is inhabited by three adults who work remotely and share the rooms with each other. The modules were deployed in the living room, kitchen, office, bedroom and bathroom.

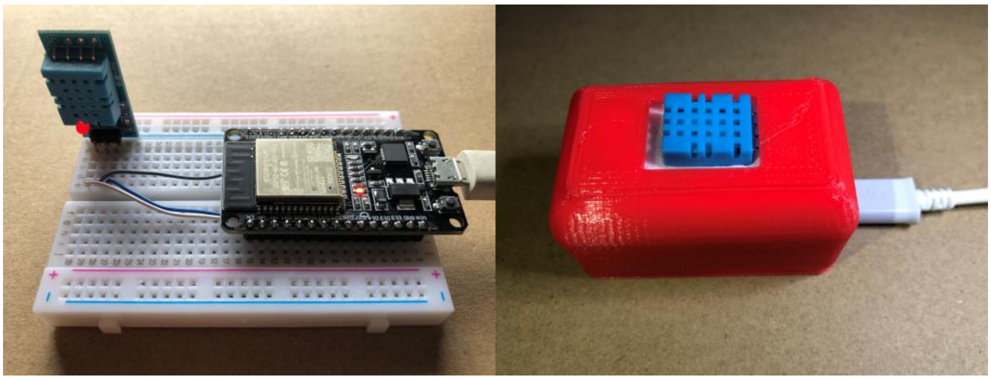


Figure 1. Prototype for collecting air temperature (°C) and relative air humidity(%)

The development board DOIT ESP32 – ESP-WROOM-32 manufactured by the company DOIT (Doctors of Intelligence & Technology) uses as base the microcontroller ESP32. This component is a high-performance, low-power IoT device developed by Espressif Systems [26].

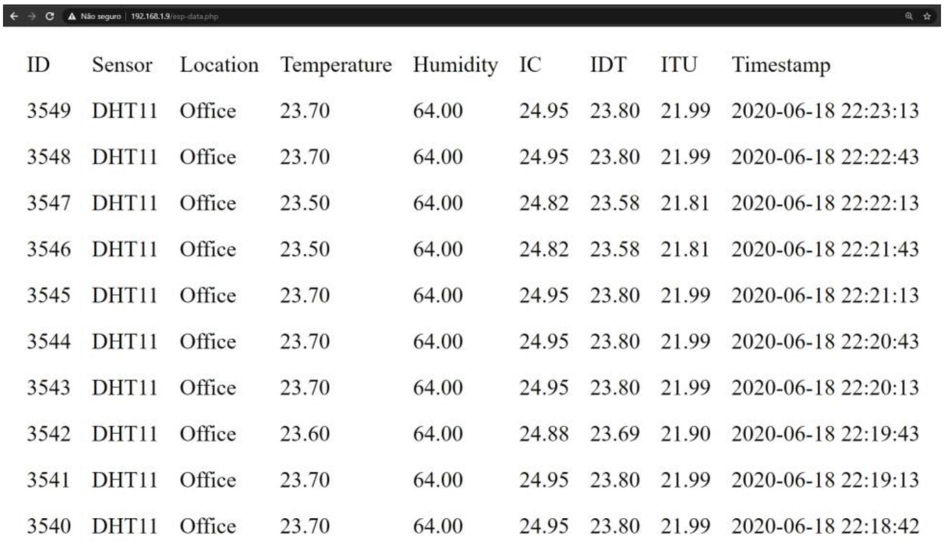
The DHT11 sensor was developed by ASAIR (Aosong Electronics Co., LTD.), includes a resistive-type HR202 humidity meter, an NTC thermistor for measuring temperature, and a high-performance 8-bit microcontroller. The temperature is measured with an accuracy of 2°C, in the range from 0°C to 50°C. Humidity is calculated with an accuracy of 5%, between 20% and 90%. DHT11 sensors do not need to be calibrated to be used because during the production process the calibration coefficients of each sensor are stored in the internal memory and are used for temperature and humidity calculations [27].

The prototype is also connected to the environment's Wi-Fi network and to the LAMP server (Linux, Apache, MariaDB and PHP) installed on the Raspberry Pi 3B+. After collecting the temperature and humidity, the DOIT ESP32 – ESP-WROOM-32 board of each environment will perform the calculation of the heat index, the thermal discomfort index and the temperature and humidity index. The apiKeyValue parameter is defined in the prototype code, which is a string used to authenticate each prototype before publishing the information in the database. The script /post-esp-data.php was implemented on the server to identify the requests sent by the prototypes and insert them into the MariaDB database of the server. Every thirty seconds an HTTP POST request is made which will send the collected information to the SensorData table in the esp_data database on the server. To display the content stored in the database, the esp-data.php script was developed, which will display all the content stored in the database through a web page when an HTTP GET request is made [28]. The SensorData table structure is specified in Figure 2.

Field	Type	Null	Key	Default	Extra
id	int(10) unsigned	NO	PRI	*NULL*	auto_increment
sensor	varchar(30)	NO		*NULL*	
location	varchar(30)	NO		*NULL*	
temperature	varchar(10)	YES		*NULL*	
humidity	varchar(10)	YES		*NULL*	
ic	varchar(10)	YES		*NULL*	
idt	varchar(10)	YES		*NULL*	
itu	varchar(10)	YES		*NULL*	
reading_time	timestamp	NO		current_timestamp()	on update current_timestamp()

Figure 2. Sensor Data structure (MariaDB)

A script /esp-data.php was developed, which will display all the content stored in the database through a web page when an HTTP GET request is made, as shown in Figure 3.



ID	Sensor	Location	Temperature	Humidity	IC	IDT	ITU	Timestamp
3549	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:23:13
3548	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:22:43
3547	DHT11	Office	23.50	64.00	24.82	23.58	21.81	2020-06-18 22:22:13
3546	DHT11	Office	23.50	64.00	24.82	23.58	21.81	2020-06-18 22:21:43
3545	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:21:13
3544	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:20:43
3543	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:20:13
3542	DHT11	Office	23.60	64.00	24.88	23.69	21.90	2020-06-18 22:19:43
3541	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:19:13
3540	DHT11	Office	23.70	64.00	24.95	23.80	21.99	2020-06-18 22:18:42

Figure 3. esp-data.php script for viewing database content

Figure 4 presents the proposed and implemented architecture in this work. The IoT prototype collects the environment parameters and transmits them, using the Wi-Fi network, to the LAMP server installed on a Raspberry PI 3B+ that contains a MariaDB database and two scripts. The database stores all the information collected by the IoT prototype, enabling future analysis and consultation. The script /post-esp-data.php is responsible for inserting the information collected by the IoT prototype and storing it in the database. The visualization can be done through the script /esp-data.php, also using the Grafana and MindsDB tools. The /esp-data.php script requests all content stored in the database, which can be used to integrate the proposed solution with other existing systems.

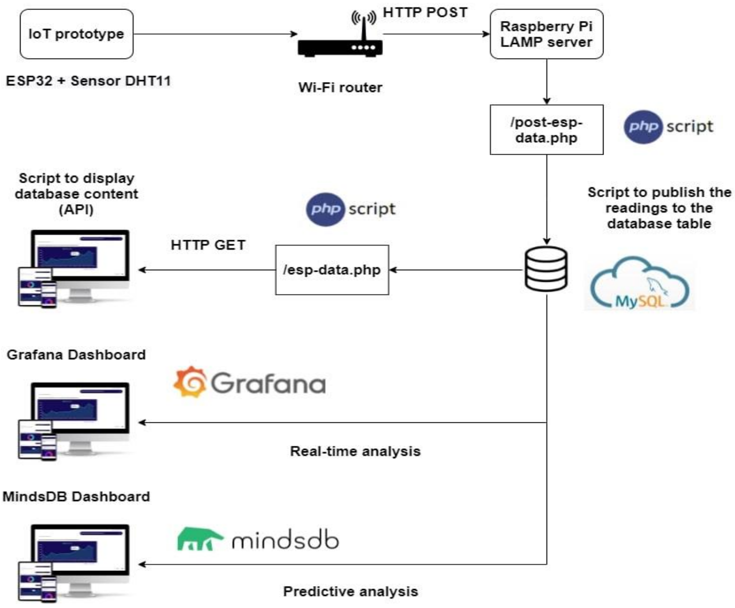


Figure 4. Project architecture

5. RESULTS

The presentation of data is carried out through the Grafana tool, which connects to the database, making it possible to consult the stored information and display them through tables, graphs and statistics. The Grafana platform is widely used for monitoring systems to generate real-time graphics using dynamic dashboards. In the tool settings it is also possible to configure alerts to notify the user in cases of abnormalities [29].

For the project, a Gmail account was created, enabled for connections from external programs to send alerts parameterized in the dashboard for each analysed environment. It was necessary to install the ssmtp package on the server, responsible for automating the transmission of e-mails via the command line so that the Grafana server triggers alerts. The dashboard was configured to average the temperature and humidity of the last fifteen minutes and trigger alerts whenever that the parameters are outside a pre-defined range. Figure 5 illustrates the Grafana dashboard and how the heat syndrome and the comfort level were determined, using the status values in each environment.

Figure 5. Grafana Dashboard containing graphs and statistics of collected data

The dashboard displays the minimum, average and maximum of each parameter as well as the graph of parameters as a function of time, according to the interval and environment defined by the user. The last data stored in each environment are also displayed through tables containing the date of collection (DD/MM/YYYY - HH:mm:ss), humidity (%), temperature (°C), the comfort indexes (°C), the heat syndrome and the status of each parameter according to Tables 4, 5 and 6.

Table 4. Relation of HI with possible physiological symptoms

HI[°C]	**Danger level**	**Heat syndrome**	**Status**
Less than 27℃	Absence of alert	Absence	4
From 27℃ to 32℃	Warning	Possible fatigue in cases of prolonged exposure	3
From 32℃ to 41℃	Very careful	Possibility of cramps, exhaustion and heat stroke for prolonged exposure and physical activity	2
From 41℃ to 54℃	Danger	Cramps, heat stroke and burnout likely. Possibility of brain damage (CVA) for prolonged exposure with physical activity	1
Greater than 54℃	Extreme danger	Heat stroke and impending cerebrovascular accident (CVA)	0

Table 5. Thermal discomfort level as a function of the TDI

TDI[°C]	**Danger level**	**Status**
Less than 21℃	No discomfort	5
From 21℃ to 25℃	Less than 50% of the population feels discomfort	4
From 25℃ to 28℃	More than 50% of the population feels uncomfortable	3
From 28℃ to 30℃	Most of the population feels discomfort	2
From 30℃ to 32℃	Everyone feels discomfort	1
Greater than 32℃	State of medical emergency	0

Table 6. Thermal comfort level as a function of the THI

THI[°C]	**Danger level**	**Status**
From 21℃ to 24℃	Comfortable	2
From 24℃ to 26℃	Slightly uncomfortable	1
Greater than 26℃	Extremely uncomfortable	0

MindsDB is an open source tool for automating machine learning projects that allows you to quickly and efficiently build and train predictive data models using standard SQL language [13], [30].

The tool allows you to import a file with the data, connect to a BI application or connect directly to a database to create artificial intelligence tables (AI Tables) that return forecasts through an SQL query, supporting the analysis of numbers, texts, categories, timestamps and certain types of images [13].

Predictions are obtained using AutoML algorithms, ensuring accurate results in a short period of time, requiring the developer to select the data and perform an initial configuration for MindsDB to automate the entire analytical process to generate the model [13], [30].

Before training the models, a pre-processing of the data was done to prepare, organize and structure the data to optimize the final quality of the data to be analysed, consequently impacting the prediction model. The data collection was started on June 17th, 2020 and on October 20th, 2021 the SensorData table had 2038830 records with a size of 163.7 MB of storage. The first step was to remove non-numeric, null and noisy data. On April 17th, 2021, for example, the sensor installed in the office failed and collected noisy data that would compromise the quality of the data, illustrated in Figure 6.

Figure 6. Noisy data collection caused by sensor failure

Data was analysed with the average of each parameter per hour to increase efficiency, reduce the amount of data and processing costs for each model. Numeric values have also been added to represent the status of each parameter, as per WHO recommendations and Tables 4, 5 and 6.

Each machine learning model implemented has its own dataset that contains only the data relevant to predicting the specific parameter in each environment. The dataset used to predict heat syndrome, for example, contains only the date, temperature, humidity, heat index and heat syndrome. Table 7 contains the machine learning models implemented with the parameters used in each dataset and the predict parameter.

Table 7. Prediction of parameters in the developed machine learning models

AutoML models	**Dataset parameters**	**Predict parameter**
who_status	reading time, humidity, temperature and sensor	WHO status
hi_danger_level_heat_syndrome	reading time, humidity, temperature, IC and sensor	Danger level and heat syndrome
tdi_discomfort_level	reading time, humidity, temperature, IDT and sensor	Discomfort level
thi_comfort_level	reading time, humidity, temperature, ITU and sensor	Comfort level
humidity_temperature	reading time, humidity, temperature and sensor	Humidity and temperature

The who_status model uses Logistic Regression, k-Nearest Neighbors, Decision Trees, Support Vector Machine and Naive Bayes algorithms. For the other models that use multiclass classification, the tool uses the K-Nearest Neighbors, Decision Trees, Naive Bayes, Random forest and Gradient boosting algorithms. The models are trained using the different algorithms to assess which is the most efficient, using the time required for training and the accuracy to make the correct predictions [13]. Table 8 specifies the models implemented in each environment.

Table 8. Machine learning models developed for each environment.

Location	**Sensor**	**AutoML template**
Living room	DHT11-A	who_status_a, hi_danger_level_heat_syndrome_a, tdi_discomfort_level_a, thi_comfort_level_a, humidity_temperature_a
Kitchen	DHT11-B	who_status_b, hi_danger_level_heat_syndrome_b, tdi_discomfort_level_b, thi_comfort_level_b,

		humidity_temperature_b
Bedroom	DHT11-C	who_status_c, hi_danger_level_heat_syndrome_c, tdi_discomfort_level_c, thi_comfort_level_c, humidity_temperature_c
Bathroom	DHT11-D	who_status_d, hi_danger_level_heat_syndrome_d, tdi_discomfort_level_d, thi_comfort_level_d, humidity_temperature_d
Office	DHT11-F	who_status_f, hi_danger_level_heat_syndrome_f, tdi_discomfort_level_f, thi_comfort_level_f, humidity_temperature_f

To train the models in MindsDB, advanced configuration parameters using Raspberry Pi 3 B+ Graphics Processing Unit (GPU) were used. For advanced parameter configuration, the tool is informed that the dataset lines are sequentially associated by the reading time feature, the target parameters for each model, the number of previous samples that will be used to perform the forecast and the error margin of the sample set. In this project, 24 previous samples were verified and the margin of error of the sample set of 2° C for temperature and 5% for humidity, considering the use of the DHT11 sensor.

After calculating the predictor model, a report is generated specifying the number of samples used for the training and testing phase of the model, the relevance of each feature to the model, the confusion matrix and the model's accuracy. Figure 7 present the report generated in MindsDB after calculating the predictor model of humidity in the living room.

The confusion matrix is an NxN matrix that allows extracting metrics to assist in the evaluation of classification models, with N being the number of target classes. The matrix compares the actual data with the data predicted by the model and it is possible to identify the model's performance and the probability of success and error for each target class [13].

Accuracy describes how often the machine learning model correctly classifies a data point, that is, the rate at which the model makes a correct prediction. The greater the accuracy of an algorithm, the lower the risk for making decisions using the model. Precision is obtained by summing the number of times the model predicted the true positives and true negatives by the total number of predictions. Table 9 shows the accuracy of each predictive model calculated in each analysed environment [24].

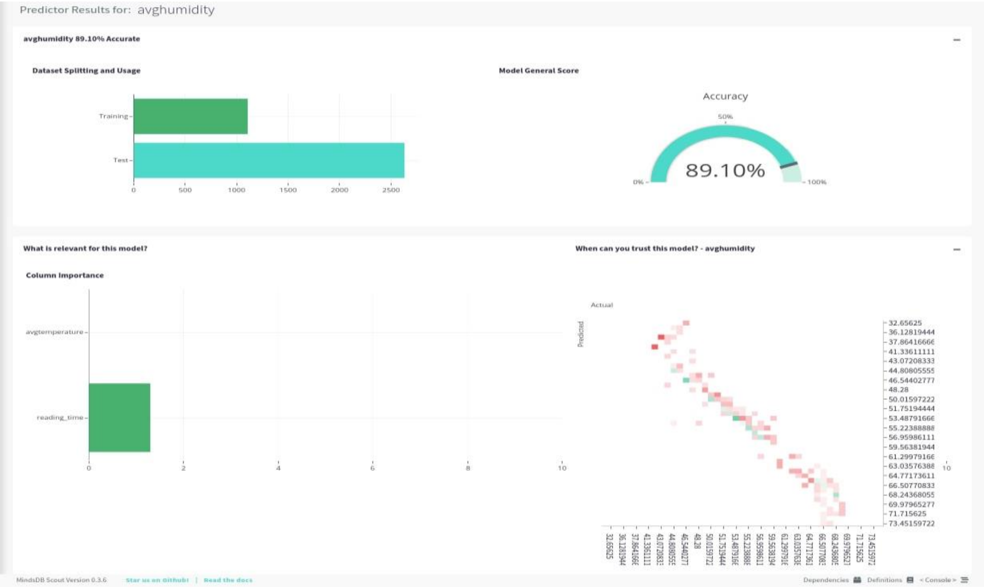


Figure 7. Results of the predictor model of the average humidity in the living room

Table 9. Machine learning models developed for each environment

Location	Target parameter	Accuracy
Living room	Who Status	93,8%
	Danger level and heat syndrome	92,1%
	Discomfort level	77,3%
	Comfort level	100%
	Relative air humidity (%)	89,1%
	Air temperature (°C)	64,5%
Kitchen	Who Status	99,6%
	Danger level and heat syndrome	86,9%
	Discomfort level	99,5%
	Comfort level	96,7%
	Relative air humidity (%)	76,6%
	Air temperature (°C)	76,0%
Bedroom	Who Status	91,6%
	Danger level and heat syndrome	100%
	Discomfort level	100%
	Comfort level	77,8%
	Relative air humidity (%)	90,4%
	Air temperature (°C)	80,9%
Bathroom	Who Status	98,3%
	Danger level and heat syndrome	100%
	Discomfort level	100%
	Comfort level	90,9%
	Relative air humidity (%)	48,0%
	Air temperature (°C)	81,9%
Office	Who Status	100%
	Danger level and heat syndrome	100%
	Discomfort level	65,3%
	Comfort level	100%
	Relative air humidity (%)	72,9%
	Air temperature (°C)	60,5%

The who_status_a model has positive results for uncomfortable environments according to the WHO recommendations, with numeric status 0, and negative results for comfortable places, with numeric status 1. During training, the relationships and the importance of each feature in the dataset, with moisture being classified as the most relevant parameter for the model, with an importance of 4.40 on a scale of 10.00. The confusion matrix of this model correctly predicted 99.00% of positive results and 93.00% of negative results with an overall accuracy of 93.80%.

During the capture period, the hi_danger_level_heat_syndrome_a model registered only the environment with no danger level and heat syndrome, with numerical status 4, and the environment in a state of attention with possible fatigue in cases of prolonged exposure, with numerical status 3. Temperature was rated as the most relevant parameter for the model, with importance 1.50 on a scale of 10.00. This model's confusion matrix correctly predicted 100.00% of numeric 3 statuses and 98.00% of numeric 4 statuses with an overall precision of 92.10%.

The tdi_discomfort_level_a registered only the environment without discomfort, with the numeric status 5, and the uncomfortable environment for less than half of the people, with the numeric status 4. The confusion matrix of this model correctly predicted 95.00% of the numeric status 4 and 98.00% of numeric status 5 with an overall precision of 77.30%.

The thi_comfort_level_a model registered the comfortable environment, with status 2, the slightly uncomfortable environment, with status 3, and the extremely uncomfortable environment, with status 0. Temperature was classified as the most relevant parameter for the model, with importance 5.00 on a scale of 10.00. The confusion matrix of this model correctly predicted 100.00% of the samples in different statuses with an overall accuracy of 100.00%.

The humidity_temperature_a model recorded the average air humidity between 44.00% and 83.00% and the average ambient temperature between 19.30℃ and 33.00 ℃. The time of collection was rated as the most relevant parameter for the model with importance 1.30 and 6.40 on a scale of 10.00 for humidity and temperature respectively. The overall accuracy for forecasting humidity is 89.10% and for temperature 64.50%.

Detailed predictive analytics of environmental parameters in the kitchen, bedroom, bathroom and office can be achieved using the reports generated in MindsDB and the same procedure described for the living room.

Figure 8 contains a comparison between actual and predicted values of temperature and humidity in the living room as of September 6th, 2021. The average difference between actual and predicted moisture was 1.77%, with a maximum variation of 6.09%. For temperature, the average difference was 0.67℃ with the maximum variation of 2.54℃. In this context, it is important to note that this predictive analysis was performed in an environment with no barriers to protect from external temperature or humidity sudden variations. This is the main reason why there is some deviation from observed actual data.

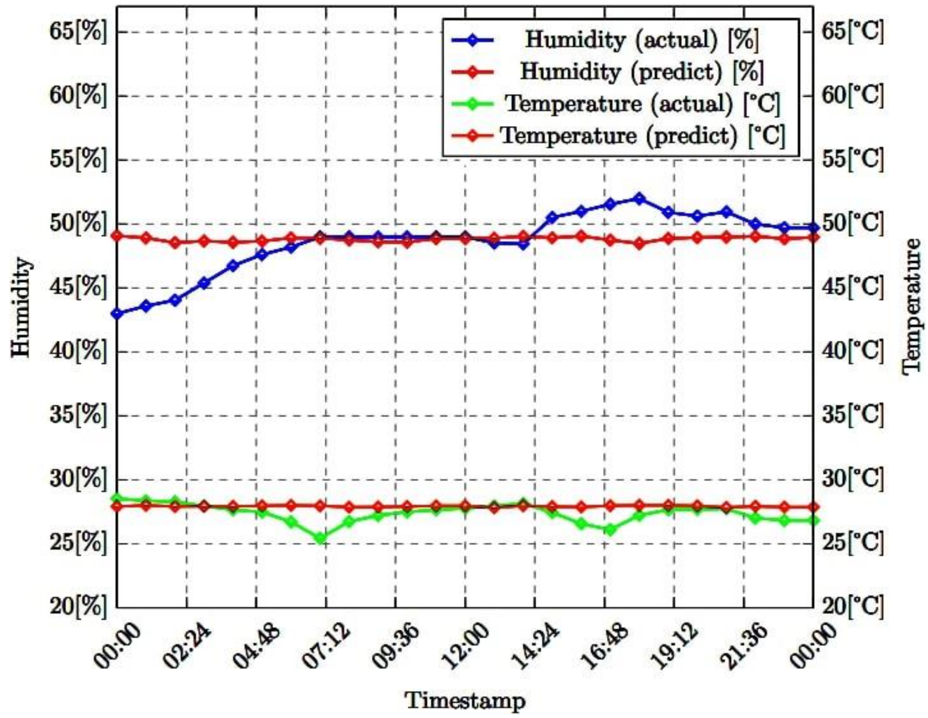


Figure 8. Comparison between predicted and actual humidity/temperature on September 6th, 2021

6. CONCLUSION

Data analysis becomes more complex as the amount of stored information increases, making it necessary to use elements of visual communication to assess the comfort of the environment in an intuitive and simple way.

The tools used promote the traceability and organized visualization of the information collected in each environment, synthesizing the relevant parameters in graphics, indicators and alerts so that the user can understand the results obtained.

Comfort in an indoor environment is analysed using relationships between air temperature (°C) and relative humidity (%) to assess environmental conditions and alert places that may pose risks to the health of exposed people, using the indices of thermal comfort.

The results of the environmental analysis are illustrated using the Grafana dashboard to assess comfort in real time and the MindsDB dashboard to predict the variables related to thermal comfort, including the comfort level and possible physiological causes for an individual who remains in the environment.

Machine Learning prediction was performed and the obtained data indicate that it is possible to reliably anticipate the evolution of the thermal comfort in internal environments. Therefore, it is possible to both take preventive actions to avoid future deviations of temperature and humidity as well adopt energy profiles to reduce the power consumption in a ecological-friendly system.

For future work, the number of IoT devices will be increased and the current LAMP server replaced by a cloud server to ensure there is no reduction in performance while increasing demand for the server. Thus, server resources will be increased during the calculation of machine learning models and also scaled according to the number of devices and sensors.

REFERENCES

[1] A. Datta, R. Suresh, A. Gupta, D. Singh, P. Kulshrestha, "Indoor air quality of non-residential urban buildings in Delhi, India" *Int. Journal of Sustainable Built Environment*, vol. *6, no*. 2,pp. *412,* Dec. 2017, DOI: 10.1016/j.ijsbe.2017.07.005.

[2] F. H. Mohamed Salleh, M. B. Saripuddin and R. bin Omar, "Predicting thermal comfort of HVAC building using 6 thermal factors," in *2020 8th Int. Conf. on Information Technology and Multimedia (ICIMU)*, Selangor, Malaysia, 2020, pp. *170–175.* DOI: 10.1109/ICIMU49871.2020.9243466.

[3] J. K. Page, "The thermal environment," "The air quality environment," in World Health Organization - Environmental Health in Rural and Urban Development and Housing Unit, 1st ed. Geneva, Geneva, Switzerland: WHO, 1990, ch. 1 and 2, pp. 15–64. [Online]. Available: https://apps.who.int/iris/handle/10665/62723.

[4] K. Fabbri, "Ergonomics of the thermal environment. Human body and clothes," in *Indoor Thermal Comfort Perception, 1*st ed, Cesena, Italy: Springer, Cham, 2015, ch. 3, sec. 1, pp. *25–63*. DOI: 10.1007/978- 3-319-18651-1.

[5] G. Gao, J. Li and Y. Wen, "DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning," *IEEE Internet of Things Journal*, vol. *7, no*. 9,pp. *8472–8484,* Sep. 2020, DOI: 10.1109/JIOT.2020.2992117.

[6] K. Nkurikiyeyezu, "An efficient thermal comfort delivery in work- places," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Kyoto, Japan, 2019, pp. *427–428.* DOI: 10.1109/PERCOMW.2019.8730680.

[7] M. R. Alam, M. B. I. Reaz, M. A. M. Ali, "A review of smart homes: past, present, and future," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. *42, no*. 6,pp. *1190– 1203*, Nov., 2012, DOI: 10.1109/TSMCC.2012.2189204.

[8] H. Singh, V. Pallagani, V. Khandelwal and U. Venkanna, "IoT based smart home automation system using sensor node," in *2018 4th Int. Conf. on Recent Advances in Information Technology (RAIT)*, Dhanbad, India, 2018, pp. *1–5.* DOI: 10.1109/RAIT.2018.8389037.

[9] G.M. Barbirato, L. C. L. Souza and S. C. Torres, "Clima e planejamento urbano: por que é importante obter informações sobre o clima," in *Clima e cidade: a abordagem climática como subsídio para estudos urbanos, 1*st ed. Maceió, AL, BR: EDUFAL, 2007, ch. 6, pp. *133–148.*

[10] E. N. Barboza, C. R. A. Caiana, F. C. B Neto, A. G. Maia, C. J. L. Lima, P. B. Maracaja, "Análise do índice de calor (IC), índice de conforto térmico (IDT) e índice de temperatura e umidade (ITU) na cidade de Iguatu/CE a partir de dados históricos," *Revista Brasileira de Gestão Ambiental*, vol. *13, no*. 3,pp. *2–5*, Jul, 2019.

[11] M. A. Borges, P. B. Lopes, L. A. Silva, M. de O. Igarashi, G. M. F. Correia. "An architecture for the internet of things and the use of big data techniques in the analysis of carbon monoxide," in *2017 IEEE International Conf. on Inf. Reuse and Integration (IRI)*, San Diego, USA, 2017, pp. *184-191*.

[12] B. P. Santos, L. A. Silva, C. S. Celes, J. B. Borges, B. S. P. Neto, M. A. M. Vieira, L. F. M. Vieira, O. N. Goussevskaia, A. A. Loureiro, "Internet das coisas: da teoria a prática," in *Minicursos SBRC - Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuıdos*, Porto Alegre, Brazil, 2016, pp. *3-27.*

[13] I. Guyon, Y. Hu, Y. Li, W. Tu, Q. Yang, Y. Yu, "Taking human out of learning applications: A survey on automated machine learning," *ArXiv*, vol. abs/1810.13306, pp. *1–18*, Dec, 2019. Accessed on: June, 21, 2021, [Online]. Avalaible: https://arxiv.org/pdf/1810.13306.pdf.

[14] P. Backman, "Deploy machine learning models directly in your database with AI Tables," in *MindsDB Documentation*. Los Angeles, CA, USA: MindsDB, 2020, pp. *1–6*. [Online]. Available: https://mindsdb.com/wp-content/uploads/2020/12/MindsDB-AI-Tables-Whitepaper-2020.pdf.

[15] American Society of Heating, "Thermal comfort," in ASHRAE Hand- book Fundamentals, Inch-Pound ed., Atlanta, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2009, ch. 9, sec. 1.

[16] Y. Demchenko, C. Ngo, C. Laat, P. Membrey, D. Gordijenko, "Big security for big data: Addressing security challenges for the big data infrastructure," in *10th VLDB Workshop on Secure Data Management*, vol. *2*, Cham, Switzerland, 2014, pp. *3-5*. DOI:10.1007/978-3-319- 06811-413.

[17] K. Blazejczyk, Y. Epstein, G. Jendritzky, H. Staiger, B. Tinz, "Comparison of UTCI to selected thermal indices" *Int. journal of biometeorology,* vol. *56, no*. 3,pp. *515-517,* May 2011, DOI: 10.1007/s00484-011-0453- 2.

[18] T. A. Yousif, H. M. M. Tahir, "Application of Thom's thermal discomfort index in Khartoum state, Sudan" *J Forest ProdIndust,* vol. *2, no*. 5,pp. *36-38,* Sep. 2013.

[19] S. Greenland, "The internet of changes everything," in *The Internet of Things, 1st ed*. Cambridge, Massachusetts, USA: MIT Press, 2015, ch. 1, 2, 4 and 5, pp. *11–24* and *33–47.*

[20] H. Garg, M. Dave, "Securing IoT devices and securely connecting the Dots using REST API and middleware," presented at the *4th Int. Conf. on Internet of Things: Smart Innovation and Usages (IoT-SIU),* Ghaziabad, India, Jul. 29, 2019. DOI: 10.1109/IoT-SIU.2019.8777627.

[21] C. Savaglio, G. Fortino, M. Zhou, "Towards interoperable, cognitive and autonomic IoT systems: An agent-based approach," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT),* Reston, VA, USA, 2016, pp. *58–59*. DOI:10.1109/WF-IoT.2016.7845459.

[22] R. R. F. Dias, "IoT sem mistérios," in *Internet das Coisas sem mistérios: Uma nova inteligência para os negócios, 1*st ed. City of Publisher, São Paulo, Brazil: Netpress Books, 2016, ch. 4 *and* 5, pp. *51–54* and *71–92.*

[23] IBM Cloud Education, "What is machine learning?", in IBM Cloud Education, Armonk, New York, USA: IBM Cloud Learn Hub, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/machine-learning.

[24] L. G. Serrano, "What is machine learning? It is common sense except done by a computer," in *Grokking Machine Learning,* 1st ed. Greenwich, Connecticut, USA: Manning Publications, 2021, ch. 1. [Online]. Available: https://livebook.manning.com/book/grokking-machine-learning.

[25] J. Waring, C. Lindvall, R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artificial Intelligence in Medicine,* vol. *1, no*. 10,pp. *1–5,* Fev., 2020, DOI. 10.1016/j.artmed.2020.101822.

[26] *ESP32-WROOM-32 Datasheet, 3.7* ed., Espressif Systems, Pudong, SH, China: 2021, pp. *8–48*.

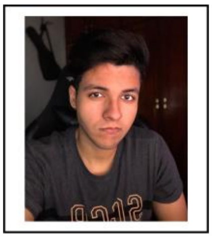
[27] *DHT11 - Humidity and Temperature Sensor, 1st* ed., SUNROM Technologies, Isanpur, AMD, India: 2012, pp. *1–7*.

[28] R. Santos, "ESP32/ESP8266 Insert Data into MySQL Database using PHP and Arduino IDE"," *Random Nerd Tutorials*. Acessed: Oct. 25, 2020. [Online]. Available: https://randomnerdtutorials.com/esp32-esp8266-mysql-database-php.

[29] Grafana LABS, "What is Grafana?," in *Grafana Documentation*. New York, NY, USA: Grafana LABS, 2020. [Online]. Available: https://grafana.com/docs/grafana/latest/getting-started/what-is-grafana/.

[30] G. Leopold, "MindsDB, AutoML Startup, Gains Seed Fund- ing"," *Datanami*. Acessed: Oct. 8, 2020. [Online]. Available: https://www.datanami.com/2020/04/20/mindsdb-automl-startup-gains-se ed-funding/.

AUTHORS



Matheus G. do Nascimento was born in São Paulo, Brazil, in 1997. He received the B. Sc. degree in electrical engineering from the Mackenzie Presbyterian University, São Paulo, Brazil, in 2019.
He is currently a scholarship student in the Graduate Program in Electrical Engineering and Computing (PPGEEC), Mackenzie Presbyterian University, São Paulo, Brazil. His current research interests include big data, internet of things, machine learning and thermal comfort.
From 2018 to 2019 he worked at Valid S.A., a Brazilian company, working as an RFID intern in the IoT area. In 2020 he was hired as a new product analyst, developing and implementing IoT and RFID solutions.

Paulo B. Lopes has received the B.Sc. and M.Sc. in EE from Federal University of Rio de Janeiro, Brazil, in 1978 and 1981, respectively, and the Ph.D. in EE from Concordia University, Montreal, Canada, in 1985.
From 1985 to 1988, he was with Elebra and CMA, two Brazilian companies, working on the design of several communication equipments. From 1988 to 1999, he was with Texas Instruments as a DSP specialist. In 1999, he moved to Motorola-SPS (later to become Freescale Semiconductor) as a Sales and Application Manager.
Since 2009, he has been with Mackenzie Presbyterian University as a professor in the School of Engineering. His research interests are Circuit Theory, Digital Signal Processing, Analog Circuit Design and Communication Theory and Internet of Things.

TEMPORAL-SOUND BASED USER INTERFACE FOR SMART HOME

Kido Tani and Nobuyuki Umezu

Mechanical Systems Engineering, Sci. and Eng., Ibaraki University, Japan

ABSTRACT

We propose a gesture-based interface to control a smart home. Our system replaces existing physical controls with our temporal sound commands using accelerometer. In our preliminary experiments, we recorded the sounds generated by six different gestures (knocking the desk, mouse clicking, and clapping) and converted them into spectrogram images. Classification learning was performed on these images using a CNN. Due to the difference between the microphones used, the classification results are not successful for most of the data. We then recorded acceleration values, instead of sounds, using a smart watch. 5 types of motions were performed in our experiments to execute activity classification on these acceleration data using a machine learning library named Core ML provided by Apple Inc.. These results still have much room to be improved.

KEYWORDS

Smart Home, Sound Categorizing, IoT, machine learning.

1. INTRODUCTION

IoT devices are widely used for controlling home appliances via Internet. Remote controllers, however, are still the most commonly used technology to operate home appliances. Remote control with Infrared rays has been in use since 1975 in Japan and is a reliable method with less small false detections.

Though remote controllers are commonly used today, there are two major problems. The first problem is that users need to use a particular remote controller corresponding to a specific home appliance. For example, when users try to turn on the air conditioner, they need to find and use a remote controller for the air conditioner. They can't use a remote controller for the TV instead. Such a number of remote controllers are confusing and take up space.

Another problem is that remote controllers are easily left behind without being sanitized. According to a report on COVID19 health crisis (Fig. 1) [1], a large number of viruses were found on remote controllers in a huge cruise ship case. In contrast,COVID19 are detected few from the light switch or the door knob because they might be considered as items which are need to be sanitized.

To solve these problems, we aim to create a new home appliance control device that is easy to manage and users don't need to touch directly. So, we propose a gesture-based interface to control a smart home. Our system replaces existing physical controls with our temporal sound commands.

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 67-76, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112107

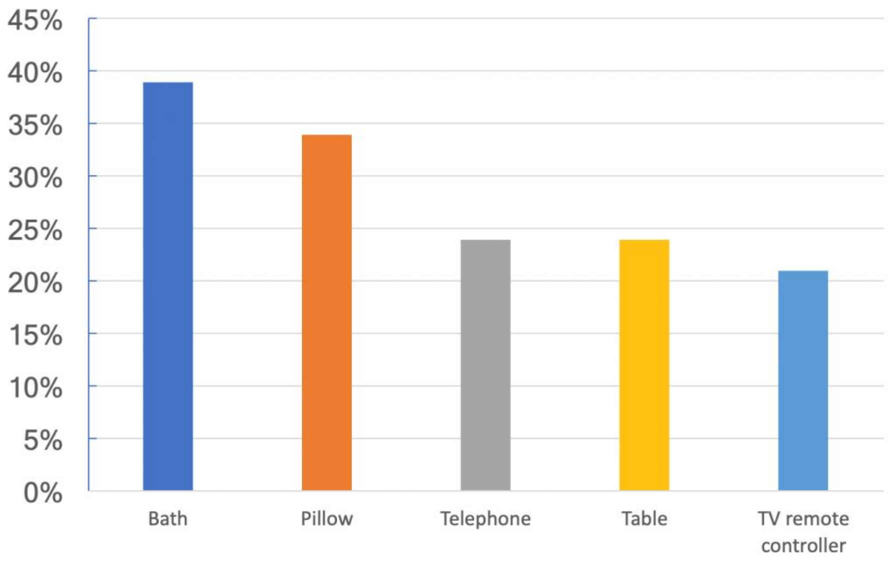


Figure 1. Detection frequency of COVID19 on the Diamond Princess [1]

2. RELATED WORK

2.1. Sensor array for smart home

An array of sensors has been proposed for recognizing various activities in a home [2]. Their system called SYNTHETIC SENSORS includes multiple sensors such as a thermometer, hygrometer and microphone on its single board design (Fig. 2). When you twist the faucet in the kitchen to get water, various information, such as the sound with twisted faucet with water falling, and the humidity raised by this falling water, would be detected by these sensors. Activity classification is performed on these sensor values.In addition, the use of two SYNTHETIC SENSORS enables more advanced sensing, such as counting how much water flows.

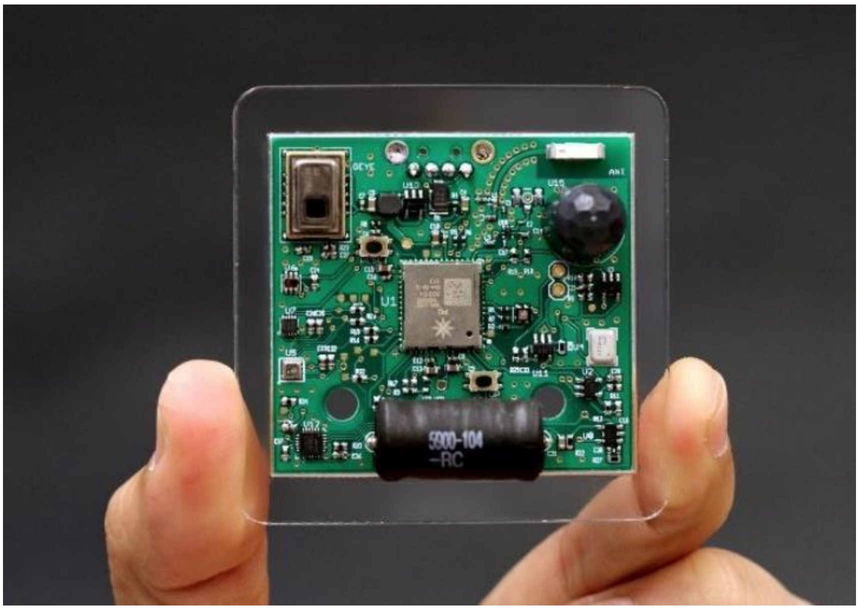


Figure 2. SYNTHETIC SENSORS [2]. Multiple sensors are mounted on a single board.

2.2. Controlling with knocking

A small IoT device called "Knocki" shown in Fig. 3 can be mounted on a surface such as a wall [3]. By knocking the wall around this Knocki for a specific number of times, users can turn on or send messages to linked appliances such as lights, TVs, and air conditioners. This device can be used in noisy environments because it does not relay on a microphone.

Figure 3. Knocki attached to a wall [3].

3. TEMPORAL-SOUND BASED USER INTERFACE

3.1. Overview of the proposed system

In our method, we obtain signals from an accelerometer attached on the user's wrist to record its vibration. If the recorded signals match one of pre-defined commands for home appliance operation, the system sends a corresponding infrared command to control that specified appliance. We use simple gestures, such as knocking on or scratching the table to control home products.

The processing flow of the proposed system is shown in Fig 4.

1: The user wears the device,
2: performs a specific action to generate vibration.
3: The system recognizes the action from sensor information,
4: requests the remote controller to control appliances.
5: Home appliances react to that command.

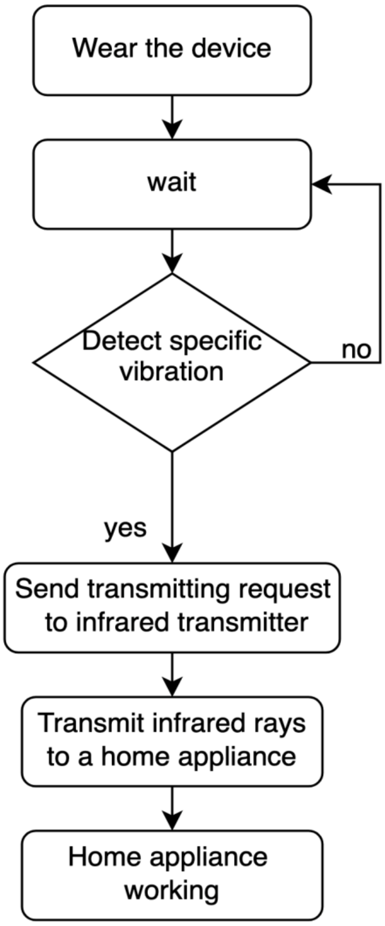


Figure 4. The flow of the system

3.2. Recording Temporal-sounds

For a preliminary experiment, we developed a prototype system that has a sound sensor and microphone, instead of an accelerometer. User gesture for our prototype are recognized not with the vibration generated by a gesture, but with the sound by that gesture. We collected a number of sound recordings used for learning gesture classification. These sounds were generated on a wooden desk and recorded with a USB microphone at a distance of 3 (cm) from the desk surface. We recorded 150 sound files for each of the following 6 types: 1) single mouse click, 2) double mouse click, 3) single knock on the desk, 4) double knock on the desk, 5) triple knock on the desk, and 6) double claps. Every waveform image shows a characteristic waveform (Fig. 5). Although mouse clicks are not used as the operation command at last, we add them as a test to confirm the operation of the system.

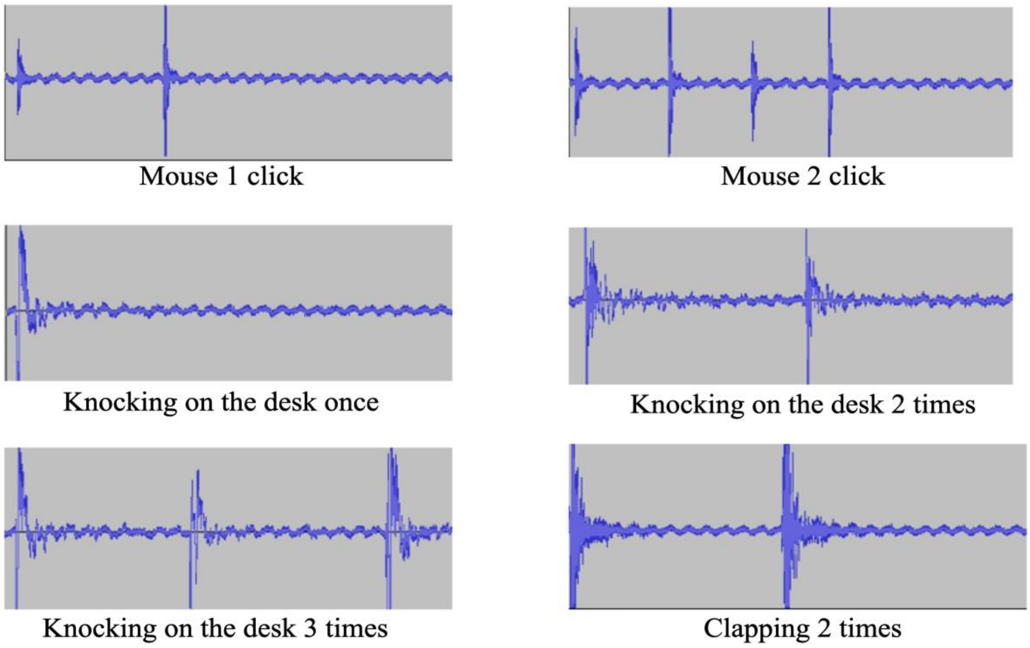


Figure 5. Wave images of 6 types of sounds

3.3. Converting to Spectrogram

Because it is difficult to perform deep learning with one dimensional sound signals, we converted them into two-dimensional spectrogram images [4]. A spectrogram is an image that represents the amplitude, frequency, the time of a sound by performing frequency analysis and using color shading. In the example of Fig.6, the closer the color of each point is to red, the higher the amplitude of the frequency.

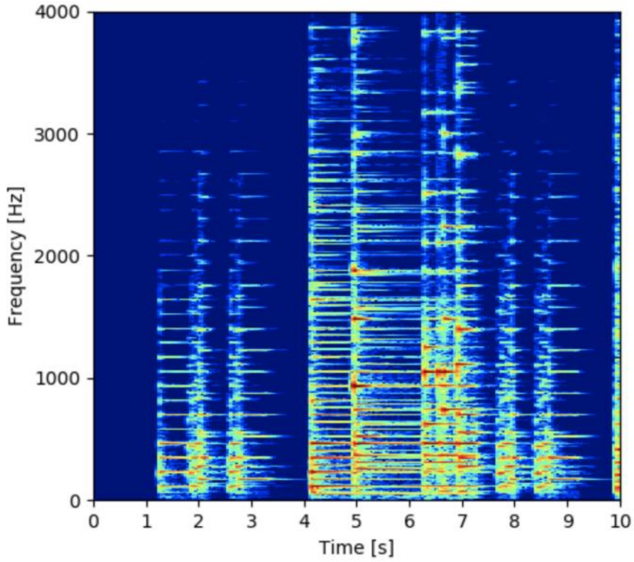


Figure 6. Example of the spectrogram

3.4. Learning spectrogram

We converted 150 sounds for each of 6 types into 900 images (Fig.7). We used a CNN model for learning spectrogram. Neither data enhancement of the training image nor drop-out is used. 30 of 150 spectrogram images for each sound are used as test data to improve accuracy. The training time was approximately 5 hours performed on a PC with AMD Ryzen 3900x, 32GB main memory, GeForce RTX2070 SUPER 8GB.

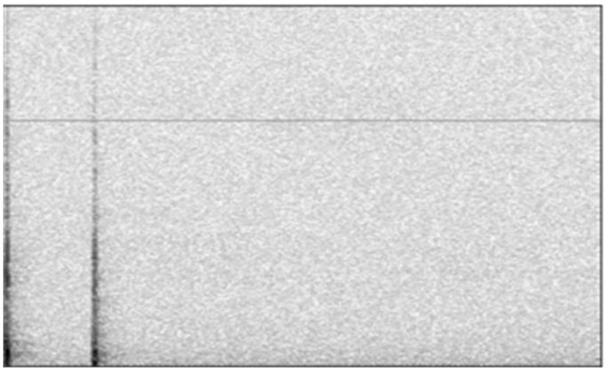


Figure 7. Converted image of spectrogram

3.5. Developing a remote controller

We developed a remote controller using Raspberry Pi [5], a popular System on a Chip (SoC) computer system. We implemented learning processes with infrared (IR) rays from other controllers and an IR transmission circuit on this small controller (Fig. 8). Our learning process with IR signals is based on a library of “IR Record and Playback” by pigpio.

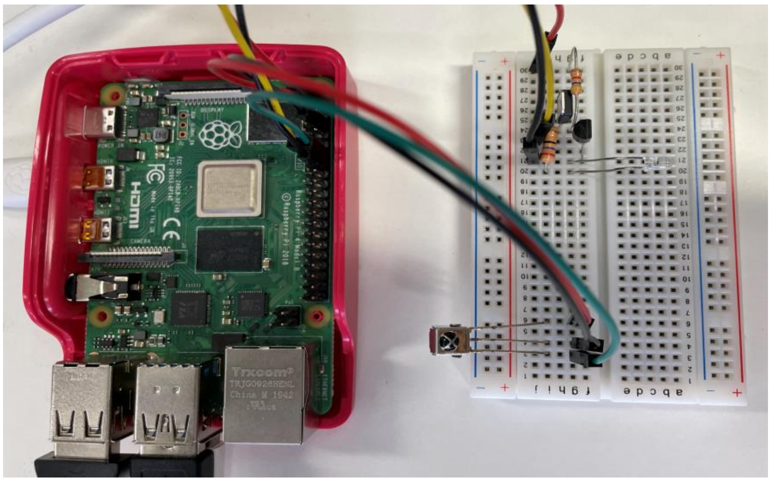


Figure 8. Remote controller using Raspberry Pi

3.6. Acquisition of acceleration data

In this experiment, we collected acceleration data using Apple Watch Series3[6]. We use CMMotionManager [7] to acquire motion data, which can acquire the motion sensor data in the device published by Apple, when the button displayed in the Apple Watch was tapped, the program acquires motion data including acceleration data. Five types of gestures were collected (resting, knocking twice on the desk, knocking three times, clapping twice and knocking with a thump-thump thump rhythm).We recorded about 100 times for each gesture. The gesture occurred three times in one data. Figure 9 shows a graph of the transition of acceleration for two knocks.

It can be seen that the X-axis and Z-axis reacted twice.

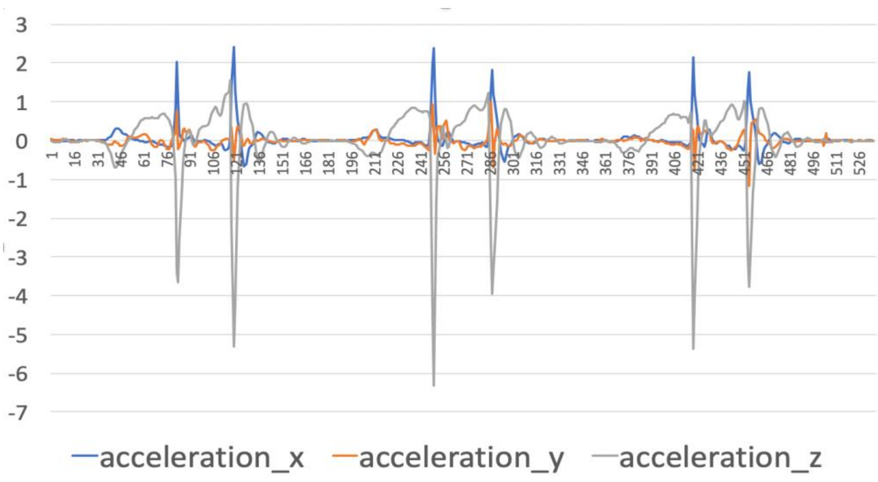


Figure 9. Acceleration for knocking two times on the desk

Core ML [8], provided by Apple Inc., is used to train the collected acceleration data.It can perform machine learning on-device by simply preparing the specified data and there is no need to build a learning layer or set up an activation function. As shown in Fig. 10, it supports various types of learning. In this research, we use activity identification. In this research, we use Activity Classification.

3.7. Learning acceleration data

We used Core ML's activity identification to learn. There are 16 accelerations that can be used for learning, acceleration_x,acceleration_y, acceleration_z, attitude_pitch, attitude_roll, attitude_yaw, gravity_x, gravity_y, gravity_z, quaternion_w, quaternion_x, quaternion_y, quaternion_z, rotation_x, rotation_y, rotation_z. In this study, we use only acceleration_x, acceleration_y, and acceleration_z because there is no significant difference between using all of acceleration data and using only 3 accelerations.

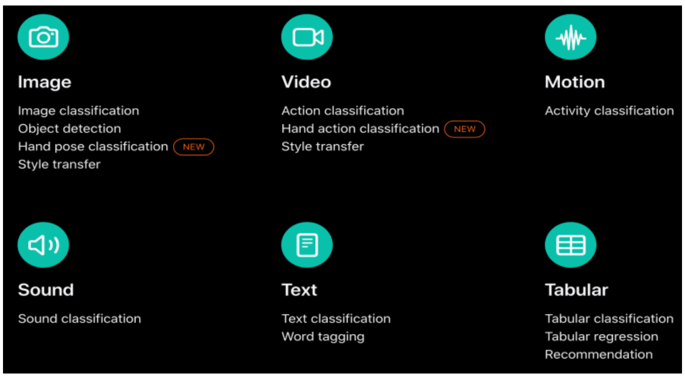


Figure 10. Available model for Core ML framework from Apple Inc. [7]

4. Classification Experiments

4.1. Classification experiment using Raspberry Pi

We conducted a series of experiments to evaluate recognition results based on our CNN learning. These experiments were performed using a microphone connected to the Raspberry Pi. We made 20 trials for each sound to rate the recognition precision of the proposed method.

The experiment results are shown in Table 1. Sounds of clicking on mouse once and twice are not included in Table 1 because these two sounds are too quiet to detect.

All 4 sounds have poor accuracy. It might be we because we used sounds recorded using microphone attached to a computer, not a Raspberry Pi when we train the model.

Table 1. Result of the experiment using Raspberry Pi. Columns shows the input. Rows shows the result of the classification.

result input	Knocking on the desk (once)	Knocking on the desk (twice)	Knocking on the desk (three times)	Clapping (Twice)
Knocking on the desk(once)	40%	60%	0%	0%
Knocking on the desk(twice)	0%	85%	15%	0%
Knocking on the desk (three times)	0%	40%	60%	0%
Clapping (Twice)	10%	25%	10%	55%

4.2. Classification experiment using Apple Watch

We conducted experiments to measure the classification accuracy of Core ML model. We set the parameters of Core ML: prediction windows size (130), sample rate (50). After wearing the Apple Watch (Fig.11), we made 20 trials for each gesture to rate the recognition precision of the proposed method. Table 2 shows the result of the experiments. Almost all the trials are classified as the knocking on the desk 3 times. This might be practically because the duration of knocking on the desk 3 times surpasses the time window used for the recognition process, which requires further investigation.

Figure 11. Running the classification program. The Japanese sentence in the screen means motion classification.

Table 2. Result of the experiment using Apple Watch. Columns shows the input. Rows shows the result of the classification.

result / input	Neutral	Knocking on the desk (twice)	Knocking on the desk (three times)	Knocking on the desk (2-1)	Clapping (Twice)
Neutral	0%	0%	100%	0%	0%
Knocking on the desk (twice)	0%	0%	100%	0%	0%
Knocking on the desk (three times)	0%	0%	100%	0%	0%
Knocking on the desk (2-1)	0%	5%	95%	0%	0%
Clapping (Twice)	0%	5%	95%	0%	0%

5. Conclusion

Although the remote controllers are commonly used in daily life, they have two major problems. The first problem is that users need to use a particular remote controller corresponding to a specific home appliance. A number of remote controllers are confusing and take up space. Another problem is that remote controllers are easily left behind without being sanitized. According to a report on COVID19 health crisis, a large number of viruses were found on remote controllers in a huge cruise ship case. To solve these problems, we proposed temporal-sound based user interface that is easy to manage and don't need to touch directly.

For a preliminary experiment, we developed a prototype system that has a sound sensor and microphone. After recoding command sounds for controlling home appliances, we converted them into spectrogram and conduct CNN learning using the images. We recorded the acceleration with Apple Watch and we conduct activity classification for the data using Core ML.

To evaluate the CNN model, we conduct the classification experiment using Raspberry Pi. The accuracy is poor for the most part. We also conduct experiments to measure the accuracy rate of Core ML model using Apple Watch. Almost all the results are classified as knocking on the desk 3 times. Both models have much room to be improved.

6. Future Work

In our experiments, we used spectrogram images recorded with a microphone connected to a PC for learning, and those images with Raspberry Pi for recognition tasks. The recognition accuracy might be improved if we use spectrogram images recorded Raspberry's microphone for both learning and recognition. We are conducting experiments with our recognition model using sound files with the microphone connected to Raspberry Pi.

Alexa [9] is one of the most popular speech recognition interfaces and a number of appliances have compatibility [10] with it. Our approach could solve some of disadvantages in speech interfaces such as longer voice commands, and effects by background noises.

To replace existing remote controllers with our system, more operation commands are needed, such as turning up/down the volume, changing the channel. We are planning to increase and regulate the gestures to correspond more commands.

We are planning to conduct user experiments and questionnaire to measure the quality of our system. First, the participants in the experiment sit in a chair and type the instructed text into the computer. While they are typing the text, they are instructed to turn on/off the light four times at random time points. We measure and compare the time between the instruction and actual operation when participants use the remote controller or our device. The questionnaire is based on the following items: ease of movement, ease of understanding gesture commands, interest in using it on a daily life, differences from the remote controllers and comment. We collect answers to the questions on a 5-point scale from 1: dissatisfied to 5: satisfied.

REFERENCES

[1] National Institute Of Infection Disease, "Report on the Diamond Princess Environmental Inspection", https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019- ncov/2484-idsc/9849-covid19-19-2.html, (Accessed 2021.5.1)

[2] GIERAD LAPUT, "SYNTHETIC SENSORS", https://www.gierad.com/projects/supersensor/%3E, (Accessed 2020.10.20)

[3] Knocki, "Knocki", https://knocki.com/, (Accessed 2020.11.13)

[4] MathWorks, "ドキュメンテーション Spectrogram", https://jp.mathworks.com/help/signal/ref/spectrogram.html, (Accessed 2021.7.6)

[5] Raspberrypi, " Raspberry Pi 4", https://www.raspberrypi.org/products/raspberry-pi-4-model-b/ , (Accessed 2021.7.6)

[6] Apple, "Watch –Apple", https://www.apple.com/jp/watch/, (Accessed 2021.11.25)

[7] Apple, "CMMotionManager", https://developer.apple.com/documentation/coremotion/cmmotionmanager, (Accessed 2021.12.8)

[8] Apple, "Create ML", https://developer.apple.com/machine-learning/create-ml/, (Accessed 2021.11.25)

[9] Amazon.com, "Alexa", https://www.amazon.com/b?ie=UTF8&node=21576558011, (Accessed 2021.12.6)

[10] Hamilton Beach, "Smart 12 Cup Coffee Maker – Works with Alexa® certified", https://hamiltonbeach.com/smart12-cup-coffee-maker-works-with-alexa-49350, (Accessed 2020.11.13)

AUTHOR INDEX