A HYBRID SHAP-RNN MODEL FOR PREDICTING AND EXPLAINING D DOS ATTACKS ON IOT NETWORKS

Ahmad Mater Aljohani[1] and Ibrahim Elgendi[2]

[1, 2] Faculty of science and technology, University of Canberra, Canberra, Australia

ABSTRACT

This study focuses on applying explainable artificial intelligence approaches to solve the challenge of attack detection in Internet of Things networks. The paper emphasizes how the proposed intrusion detection platform might enhance the model ability of explaining and comprehending attacks that might take place. The study suggests theoretical and practical advancements that directly affect the intrusion detection in Internet of Things networks. To comprehend how the system operates and how attacks happen, this paper describes the architecture and structure of the introduced explainable model. The targeted contributions focus on identifying possible attacks and interpreting taken decisions. The findings of testing on the UNSW-NB15 dataset show a better performance of RNN-SHAP model when compared to other algorithms (GRU, SVM, and RL), in terms of numerous metrics such as accuracy (98.98%), precision (100%), recall (93.44%), and F1-score (95.63%).

KEYWORDS

Attack detection, explainable artificial intelligence (XAI), deep learning (DL), Recurrent neural networks (RNN), SHapley Additive explanation (SHAP), Distributed Denial of Service (DDoS)

1. INTRODUCTION

A network vulnerability refers to imperfections and shortcomings in the hardware, software, or overall operational procedures of a system. These types of network vulnerabilities are usually intangible and related to software or data.

The Denial of Service (DoS) attack typically involves overwhelming the service provider's resources and network bandwidth to deny legitimate users from accessing the network and its services. An example of a DoS attack is targeting wireless connections of sensors collecting data where attackers gain control of different parts of the network.

IoT Network analysis investigates the inter-connectivity between heterogeneous devices and humans. Over the last decade, artificial intelligence, especially explainable artificial intelligence (XAI) [1] rises as a suitable method for understanding, explaining, and interpreting the Artificial Intelligence (AI) decisions regarding different specific events and trends. Understanding, correlating, managing, and taking decisions regarding billions of individual IoT (Internet of Things) small parts of information is a significant technical issue.

The term XAI refers to a collection of methods and procedures for explaining how AI models make decisions. According to existing literature, machine learning (ML) models are vulnerable to attacks such as model inversion, model extraction, and membership inference. Such attacks focused on the learning model itself, or on the data used to train and build the model, depending

David C. Wyld et al. (Eds): DMML, CSITEC, NLPI, BDBS - 2024
pp. 107-119, 2024. CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.140606

on the circumstances and involved parties. However, when the owner of an AI model wishes to grant only black-box access and does not expose the model's parameters and architecture to third parties, solutions like XAI can notably enhance the susceptibility to model extraction attacks.

Indeed, attack detection systems for IoT networks may be constructed using machine learning methods, particularly XAI. Creating a realistic and efficient training dataset is a difficult problem when developing such systems with ML [2]. Since continuous data flow is the sole condition that allows attacks interception, the network data flow should be of high quality throughout an attack. When developing the ML model, a variety types of data are considered due to the operation of many heterogeneous devices in the network. It is possible to develop an appropriate threat detection system for an IoT context by overcoming these obstacles. To guarantee that IoT devices are safe from cyberattacks, security elements need to be considered from the outset of the design process.

Furthermore, it might be said that XAI models, despite its quick development, still immature and unproven area that frequently struggles with a lack of formality and accepted concepts. When analysing ML algorithms, it is critical to comprehend their operation since they have the potential to transform feeds into results without requiring direct coding. As a result, there is a highly explicit dilemma about explainability in respect to the compromise linked to discrimination, namely, the requirement for simple ML models. However, this is quite biased if the ML representations are straightforward.

Depending on the environment in which it is applied, the model's reductions and the assumptions that must be proven true for the framework to operate are significant in different ways. Hence, ML models that are simpler according to this criterion have higher needs and are more biased, which means that the bias decreases as model complexity rises.

It is also crucial to remember that research into explainable artificial intelligence is growing to provide outcomes that consumers will accept and comprehend. This is especially true when it comes to the examination of medical data to clarify the decisions made by ML models. Nonetheless, when Deep Learning (DL) [3] is integrated into more significant areas of society, including medical diagnosis, more details about what happens behind the scenes become available. Consequently, with the development of AI approaches with explainability capability, researchers will be able to understand a wide variety of findings in the most efficient and quick possible way, helping them to better understand the answers to AI results.

Since dealing with well-documented issues that have been handled by professionals for a lengthy period, it is not required for the method to be explainable, this does not imply that all ML models must be comprehensible. When the consequences of your poor choices are minimal, such as in the case of an AI that can learn to dance, it doesn't even make sense to explain them to an AI model. However, explainability is crucial for ML models that directly affect people's life, such as algorithms that determine who will be fired [4].

A clear example of XAI is as follows: After being trained to identify different animal species, a computer will build up its understanding of body parts from the data to precisely identify, comprehend, and characterize the animal. The science and engineering of designing and constructing intelligent devices is known as AI computational equipment and programs. This proves and shows that certain AI systems mimic human cognitive processes, but it doesn't show how much of an influence this has.

The contributions of this study focus on proposing an explainable RNN-SHAP model for resolving the DDoS detection problem. In this RNN-SHAP model, the SHAP strategy should explain the decisions taken by the RNN network and enhance its performance in detecting and identifying the DoS attacks.

The rest of the paper shows the recent related works (section 2), proposes a modelling for the problem (section 3), illustrates the XAI-based RNN-SHAP model for the DDoS (section 4), details the results (section 5), then interprets (section 6), and concludes (section 7) the study.

2. LITERATURE REVIEW

Authors in [5] proves that there is an opportunity with commercial XAI frameworks such Local LIME, which assigns sensitivity in an ML local prediction instance to specific input features, there is a chance to increase interpretability in ML-based cyber analytics. These technologies have mostly been tested in computer vision-related fields where spatially connected features and concrete classification problems exist. Because of the competitive character of attacks and the proof they leave in disparate data streams, these scenarios vary dramatically from many other cyber applications. Authors suggest defining different vulnerability classes, but they neglect the issues of local ML explanations in the field of cyber security,

The work in [6] has presented a first step in adapting XAI methodologies in smart monitoring by drawing an overview on existing ML solutions for smart monitoring. The incomprehensibility of many ML algorithms used in smart monitoring has been examined, and XAI techniques have been put out as a solution. It is concluded that depending on their goal and the targeted human audience, ML algorithms may need varying degrees of explanations for smart monitoring. The study also concludes that a thorough examination of the explainability and degrees of explanation for ML algorithms is necessary for promoting the development of smart monitoring.

The study in [7] provides a thorough analysis of recent and upcoming advances in XAI technology for smart cities. It also emphasizes the technical, industrial, and sociological developments that spur the development of XAI for smart cities. It provides a detailed explanation of what is essential to implementing XAI solutions for smart cities. Numerous XAI application cases, problems, applications, potential solutions, present and future research advancements are also covered in the paper. Detailed descriptions of research initiatives and activities, such as efforts to standardize the development of XAI for smart cities, were provided.

Another investigation in [8] gave a thorough analysis of XAI in intelligent connected vehicles (ICVs) for the purpose of intrusion detection and mitigation. Because of vulnerabilities in linked devices, the Internet of Vehicles (IoV) is an expanded use of the IoT in smart transport systems (ITSs). However, because most detection systems use AI in a "black box" fashion, transparency remains a challenge, necessitating the development of explainable AI. A thorough overview of the current XAI frameworks, and their utilization to ICV security are covered in this study. To promote rule based XAI's acceptance in crucial areas like the automotive industry, XAI developers must also address the problem of bias caused by this technology. Moreover, an interesting research topic is to recognize the requirements for dependability, the minimal needed computing complexity, and the inclusion of user-friendly XAI modules.

In [9], the use of XAI in cybersecurity is thoroughly reviewed in the current paper. System, network, and software protection from various threats is made possible by cybersecurity. The potential for using XAI to anticipate such attacks is enormous. This study offers a succinct introduction of cybersecurity attack types. The use of conventional AI techniques and the difficulties connected with them is then examined, which pave the way for the use of XAI in a variety of applications. Additionally, the XAI implementations of numerous academic projects and business are shown.

In the same regard, the study in [10] introduces the *IoTBoT-IDS* architecture, to defend IoT-based intelligent infrastructure against botnet attacks, as a revolutionary probabilistic learning-based botnet detection system. *IoTBoT-IDS* uses statistical learning-based approaches such as the *Beta Mixture Model* (BMM) to represent the typical behaviour of IoT networks. Any departure from the expected pattern of behaviour is recognized as an abnormal occurrence. Three comparisons set of data produced from IoT networks were utilized to assess IoTBoT-IDS.

In the framework of a project financed by the European Commission [11], an Explainable AI solution utilizing DL and semantic web techniques is suggested to create a hybrid classifier for the application of smart cities flood monitoring. In this mixed model, the DL component determines the existence and degree of object coverage while the categorization is done using semantic rules that were carefully developed with experts. The experimental findings, which were presented with a practical application, revealed that this hybrid technique to image classification performs on average 11% better (F-Measure) than DL-only classifiers.

Table 1. Summarization of the recent relevant research studies

[ref]	**Methodology**	**Application**	**Findings**
[5]	XAI with ML (Local LIME)	commercial XAI frameworks	Adding a local prediction instance to specific input features increases interpretability in ML
[6]	XAI-ML	smart monitoring	ML algorithms may need varying degrees of explanations for smart monitoring
[7]	XAI-DL	XAI for smart cities	The paper provides a detailed explanation of what is essential to implementing XAI solutions for smart cities
[8]	XAI-ML	XAI in intelligent connected vehicles	To promote rule based XAI's acceptance in automotive industry, XAI developers must address the problem of bias caused by this technology.
[9]	XAI-ML	XAI in cybersecurity	XAI implementations of numerous academic projects and business are shown
[10]	XAI-ML	*IoTBoT-IDS* architecture	A revolutionary probabilistic learning-based botnet detection system (*IoTBoT-IDS*) is proposed.
[11]	XAI-DL	smart cities flood monitoring	Hybrid technique (DL-XAI) in image classification performs about an average of 11% better than DL-only classifiers (in F-Measure)

The latter studies (previously discussed, and in Table 1) on ML interpretability techniques that has been done in recent years has shown that there is still potential for development. These studies highlight the advantages and improvements that XAI techniques may offer to current machine learning workflows, but they also highlight the shortcomings and performance gaps in these methods.

3. MODELLING OF THE PROPOSED SYSTEM

The aim is to resolve the problem of detection of DDoS attacks in IoT networks. Hence the following mathematical formulation is suggested: Let ***Nd*** be a node in ***Nwk***. ***Nwk*** represents the set of nodes Nd_i in the IoT network. The considered objective functions are as follows:

$$\boldsymbol{f1}\textbf{: Maximize } \sum Pv_i \,,\ \forall\ Nd_i \in Ntw \qquad \textbf{(1)}$$

f1 indicates that ***Pv***$_i$, the degree of privacy achieved at the level of each node ***Nd***$_i$, should be optimized (by maximization).

$$f2\text{: Maximize } \sum Cf_i \text{ , } \forall\ Nd_i \in Ntw \qquad (2)$$

f2 indicates that ***Cf***$_i$, the degree of confidentiality achieved at the level of each node ***Nd***$_i$, should be optimized (by maximization).

$$f3\text{: Maximize } \sum Sc_i \text{ , } \forall\ Nd_i \in Ntw \qquad (3)$$

f3 indicates that ***Sc***$_i$, the degree of security achieved at the level of each node ***Nd***$_i$, should be optimized (by maximization).

$$f4\text{: Maximize } \sum Tr(Pv_i, Cf_i) \text{ , } \forall\ Nd_i \in Ntw \qquad (4)$$

f4 indicates that ***Tr(Pv***$_i$***, Cf***$_i$***)***, the degree of trade-off between privacy and confidentiality for a node ***Nd***$_i$, should be optimized (by maximization).

4. METHODOLOGY

4.1 XAI Methodology

Organizations may acquire access to AI technology's underlying decision-making and make changes via explainable AI and interpretable ML. Explainable AI may improve a product's or service's user experience by increasing the end user faith that AI provides sound judgments. As AI advances, ML processes must be understood and regulated to ensure the accuracy of AI model outcomes. Distinctions between AI and XAI, as well as the methods and strategies utilized to convert AI to XAI are shown in Figure 1.

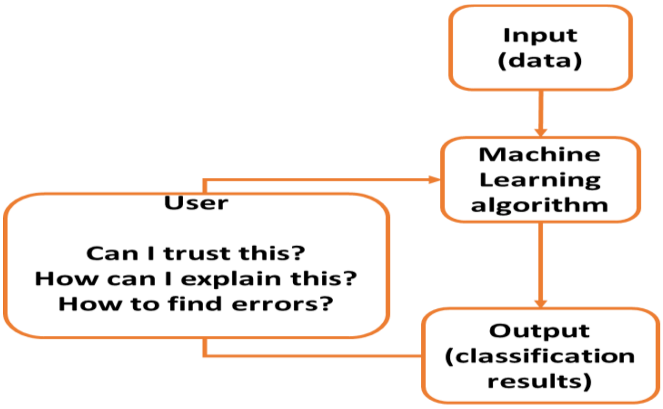


Figure 1. How XAI works?

4.2 RNN-SHAP

To explain a recurrent neural network (RNN) using SHAP [12], we can apply the following general process:

- Train the RNN on your data.
- Compute the SHAP values for each input feature and for each output of the RNN.

- Visualize SHAP values to understand how much each feature contributed to the model's output at a time step.

In the case of an RNN, there are some specific considerations to keep in mind, as the model has an internal hidden state that is used to propagate information through time.

One way to approach this is to unroll the RNN over time, so that we have a separate set of SHAP values for each time step. This lets us see how the contribution of each feature changes as the hidden state (and therefore, the input to the model) evolves over time.

It's also important to ensure that the SHAP values are being computed in a way that considers the sequential nature of the input data. One approach is to use the "masking" technique, which involves setting the SHAP values for any inputs that are padded with zeros to zero as well. This helps to ensure that the SHAP values are only reflecting the information that's relevant to the model's output.

4.3 RNN-SHAP for DDoS

The attack of distributed DoS (DDoS) is a malevolent endeavor to impede regular network, server, or service traffic by flooding the victim or its surroundings with excessive amounts of Internet data. It has seriously compromised the network environment's security.

DDoS attacks are a type of denial-of-service attack when the attacker targets a specific victim by using an IP address belonging to an authorized user. DDoS attacks come in a variety of forms, including DNS Reflect, ICMP Flood, SYN-flood, ACK-flood, UDP-flood, Connection DDoS, and so on. The attackers' primary goal is to clog the resources so that receivers cannot get services. To achieve this, attackers can employ a set of techniques, including flooding the network with fake requests. The DDoS attack is dispersed by the attacker using many computers. DDoS attacks pose an immense threat to the Internet, and many defense mechanisms have been proposed to combat the problem.

Various techniques were employed to address DDoS. It is possible to get insight into the elements that are most crucial in identifying whether an incoming traffic flow is malicious by employing SHAP to explain the predictions of an RNN-based DDoS detection model.

The following actions can be taken to put RNN-SHAP for DDoS into practice:

1. Prepare the data: The dataset must be prepared to extract pertinent characteristics and convert them into a numeric format that the RNN model can use.
2. Train the model: Using the preprocessed dataset, train an RNN model.
3. Calculate SHAP values: To determine the SHAP values for each feature and each RNN output, use the SHAP implementation, such as `shap.Explainer}.
4. Visualize SHAP values: Use the SHAP values to determine which attributes are most crucial for determining the likelihood of malicious activity in each traffic flow. This can assist us in recognizing trends and actions that point to a denial-of-service assault, such an abrupt spike in traffic coming from a certain IP address.
5. Improve the model: To enhance the RNN model's capacity to identify DDoS assaults, adjust based on the knowledge gleaned from the SHAP values.

In conclusion, only a few numbers of studies, such as [13], investigate the optimization of RNN by elucidating it using SHAP. Some research, such as [14], concentrates on the application of RNN for DDoS.

The use of SHAP to describe the RNN employed to address the DDoS problem is then our primary contribution. We may learn more about the variables influencing the model's predictions and enhance its efficacy in identifying these kinds of cyberattacks by utilizing RNN-SHAP for DDoS.

4.4 Designed Artifact for RNN-SHAP based DDoS Detection.

As shown in Figure 2, there are several important processes involved in designing an artifact for Recurrent Neural Networks (RNN) improved by Explainable AI for DDoS (Distributed Denial of Service) detection. Here's a high-level rundown of the design process. The investigated research challenge is to use RNN-based models to identify DDoS attacks in network traffic data efficiently while ensuring that AI's decision-making process is clear and understandable for security experts.

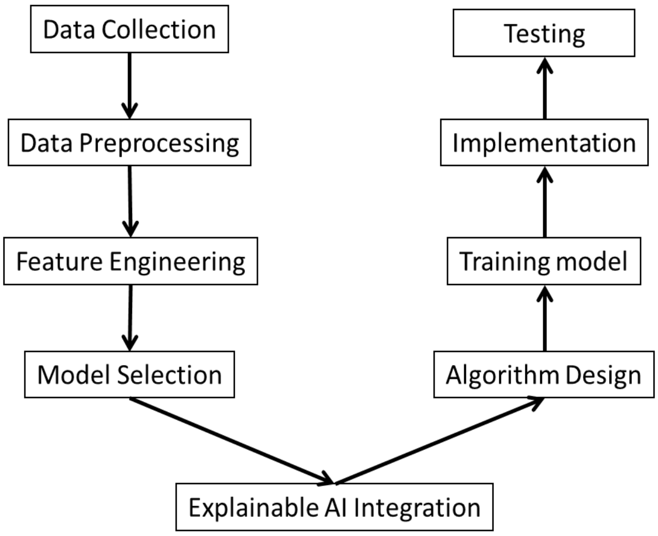


Figure 2. Artifact Design of the model

Data collection is the process of compiling information about network traffic from multiple sources, such as network logs and intrusion detection systems. To train and evaluate the model, the data should contain both regular and attack traffic. **Data Preprocessing** prepares the data to make it acceptable for RNN input, reduce noise, and anonymize sensitive data. This might entail encoding, scaling, and normalizing the data. **Feature Engineering** focuses on obtaining pertinent elements from the data, like source/destination IP addresses, packet sizes, rates, and more, to accurately depict network traffic. **Selecting the model** denotes selecting an appropriate RNN architecture for processing data sequentially. For time-series data, such as network traffic, Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) networks are frequently utilized. **Integrating the explainable** which refers to the incorporation of XAI techniques into the model. One approach is to use attention mechanisms in the RNN to highlight the most critical features contributing to the decision. **The algorithm Design** includes designing the RNN model architecture, including the number of layers, units, and activation functions; Incorporating attention mechanisms or other explainable AI techniques within the RNN architecture to provide interpretable results. **The training** involves the training dataset (Splitting the data into training, validation, and test sets.); Training the model (Train the RNN model using the training data and monitor its performance and fine-tune hyper-parameters as needed.); and evaluating the model's performance on the test dataset using metrics such as accuracy, precision, recall, and F1-score.; Also Validating the explainability component by analyzing which features the model focuses on during its decision-making process. **Implementing the environment** by using a suitable programming language and deep learning framework like Python and Tensor Flow/Keras or PyTorch. Moreover, depending on the dataset's size and complexity, you may require GPUs or TPUs to expedite training. The final phase, **testing**, includes simulating DDoS attacks (Test the model with simulated DDoS attacks to evaluate its ability to detect and classify these attacks accurately.); Real-World Testing (Deploying the model in a real network environment to assess its effectiveness and accuracy in identifying DDoS attacks.); Explainability Testing (Analyzing the model's explanations and ensure they make sense to security professionals.).

The created artifact creates a model that is transparent in its decision-making process and efficient in recognizing attacks by fusing explainable AI approaches with RNN-based DDoS detection. Thorough testing is essential to confirm the artifact's efficacy in improving network security, particularly in real-world situations.

5. RESULTS

In this section, the performance of the proposed SHAP-RNN is assessed on UNSW_NB15 datasets, and compared to other paradigms such as RNN, RL and SVM.

5.1 Used Dataset

We use data driven solutions. The source of our data is a well-known dataset from the literature. Sampling techniques are not needed since existing data in **UNSW_NB15** [15] is structured, documented, and used in the state-of-the-art. Introduced by the Australian Centre for Cyber Security (ACCS), the UNSW-NB15 dataset collects network packet information generated by a mix of real-world modern normal activities and contemporary synthetic attack behaviors. This data was produced via the IXIA PerfectStorm tool in the Australian Centre for Cyber Security's Cyber Range Lab. The Tcpdump tool was utilized to capture 100 GB of the raw traffic as Pcap files. It contains nine distinct types of attacks, including worms, shellcode, reconnaissance, generic, exploits, DoS, backdoors, and analysis. To develop the dataset, the Argus and Bro-IDS tools are employed, and a total of 49 features with class labels were generated across twelve different algorithms. Researchers often leverage this dataset to develop and evaluate machine learning models for the detection of network-based attacks and to improve cybersecurity measures.

To assess the proposed RNN-SHAP, different metrics can be used on UNSW_NB15, such as precision, F1 value and recall (to assess the performance of the DDoS traffic of the feature/class), accuracy (to assess the classifier performance), AUC, and ROC.

Stemming from the dataset UNSW-NB15 [16], Table 2 shows the number of records in UNSW-NB15 of DoS and some other types of attacks.

Table 2. Types of UNSW-NB15 records.

Type	**Training records**	**Testing records**
Worms	130	44
Normal	56000	37000
DoS	**12264**	**4089**
Exploits	33393	11132
Backdoors	1746	583
Total	175341	82332

5.2 Design of Accuracy and other Metrics

The accuracy metric relies on the formula A = Σ(TN+TP) / Σ(FP+FN+TN+TP) (xx) the abbreviations in A are True Negative (TN), True Positive (TP), False Positive (FP), and False Negative (FN). The accuracy rates for 100 epochs are shown in Figure 3.

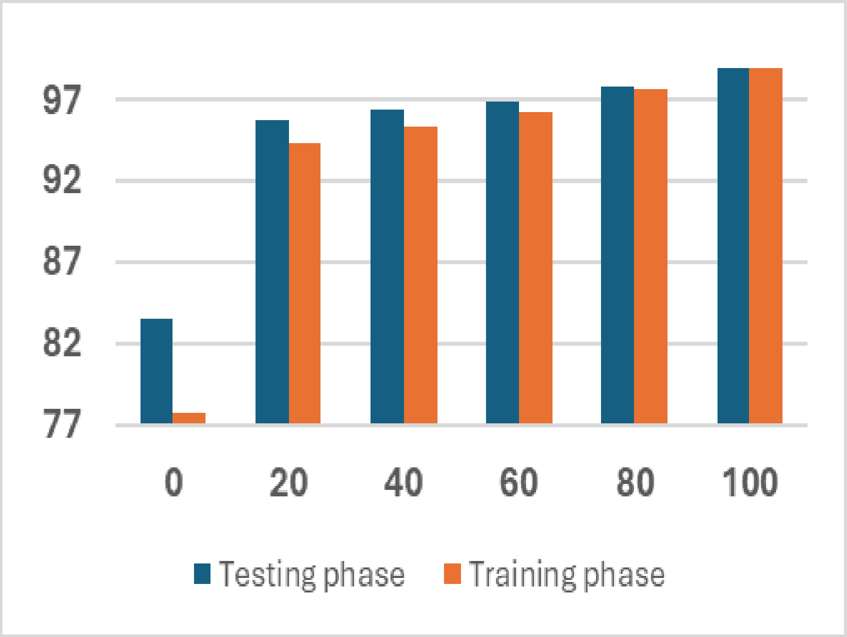


Figure 3. Evaluation of accuracy metric for RNN-SHAP

F1 score formula is **F1** = (2TP) / (2TP+FP+FN) (xxx).
Precision formula is **P** = TN / (TN+FP) (xxx).
Recall formula is **R** = TP / (TP+FN) (xxx).
The confusion matrix of testing instances for the RNN-SHAP (Table 3) indicates a high performance of correct predictions.

5.3 Comparisons of Training and Testing Phases

Table 3. Confusion matrix of testing instances using RNN-SHAP.

	Anomaly	**Normal / correct**
Anomaly	2353	108
Normal / correct	72	1556

All the metrics values for all the tested algorithms are detailed in Table 4 and Table 5.

Table 4. Values of the metrics on the training phase.

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
RNN	96.94	89.21	92.38	90.27
RNN-SHAP	98.91	92.78	93.93	95.24

GRU-RNN	97.27	88.29	92.81	93.49
SVM	92.38	89.32	91.33	94.28
Random Forest (RF)	94.26	91.35	92.16	94.33

Table 5. Values of the metrics on the testing phase.

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
RNN	97.43	89.92	94.21	91.39
RNN-SHAP	98.98	100	93.44	95.63
GRU-RNN	97.56	89.65	100	94.84
SVM	93.45	89.93	93.28	94.89
RF	94.78	100	92.77	95.65

5.4 AUC and ROC Analysis

Another metric, named AUC, indicates the rate of correct predictions from all records values. An AUC equal to 1 implies fully correct records. Table 6 shows that RNN-SHAP has the highest AUC values, then, has the best prediction rate.

Table 6. AUC values for RNN, RNN-SHAP and RF.

	RNN	**RNN-SHAP**	**RF**
AUC	92.34%	96.69%	91.36%

The ROC analysis is another test relying on the ROC curve used to determine the behaviour of used algorithms. The results are illustrated in Figure 4. indicate the following: The ROC of RF illustrates that for a prediction having 15% of error, a 95% of TP values is achieved. The ROC of RNN illustrates that for a prediction having 20% of error, a 80% of TP values is achieved. The ROC of RNN-SHAP illustrates that for a prediction having 10% of error, a 95% of TP values is achieved.

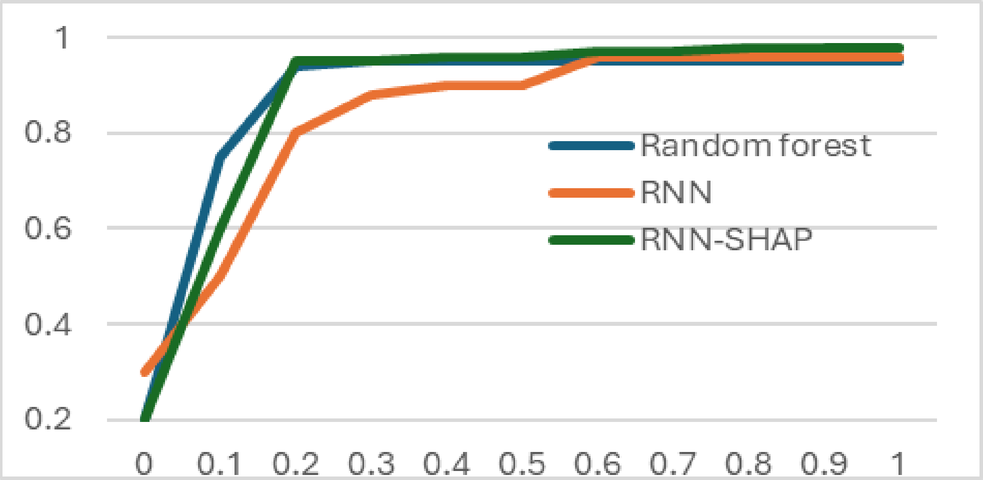


Figure 4. ROC model for the used algorithms

5.5 Temporal Complexity Analysis

The training time is also a metric that is used to assess the performance of algorithms. From Figure 5, it can be concluded that, despite the previous proved performance of RNN-SHAP, its training time is not better than other paradigms.

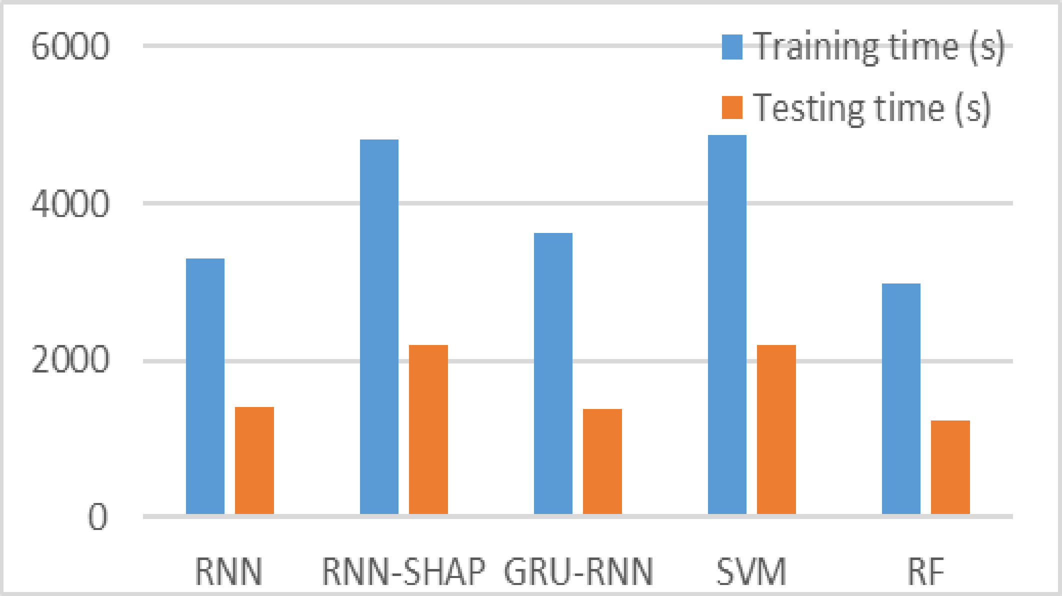


Figure 5. Time of training for the different used algorithms

6. Interpretation and Deductions

By creating the foundational explanatory AI-oriented models that enable the description of real operations, XAI enables AI-oriented solutions to explain the context in which intelligent devices operate. Ultimately, it is contradictory to grow AI technology into a divine force that people would desire without establishing a cause-and-effect relationship. However, it is unfeasible to ignore the computerized intelligence that technology provides to society. In brief, it is imperative to create AI models that are both interpretable and flexible enough to facilitate cooperation with experts and scholars from a variety of subjects.

Thus, XAI will be essential to grasp, handle, and be able to effectively articulate their artificial thinking, logic, method, and to communicate an acceptable comprehension of how the model function, and ultimately believe the next wave of AI-oriented machines.

Like many intrusion detection datasets, UNSW-NB15 may exhibit class imbalance, meaning that some attack classes are less common than others. This imbalance can affect the training and evaluation of machine learning models.

In terms of detecting IoT attacks, XAI still offers a lot of undiscovered possibilities that will be unlocked in the upcoming years, based on the recent state-of-the-art. It is important to remember that, to understand and interpret machine learning algorithms, there has been a paradigm shift from "traditional programming," in which all heuristics had to be explicitly passed, to a new idea in which, instead of saving every operation the model performs, multiple instances are provided, and machine learning is allowed to determine the optimal plan of action.

7. CONCLUSION AND FUTURE DIRECTIONS

The goal of this research is to apply explainable artificial intelligence approaches to address the challenge of threats detection in IoT networks.

To construct DL IoT threat detection systems that fully leverage text mining, large-scale datasets for attack detection must be produced. Bridging the gap between ML expertise and IoT application security experience is necessary to establish ML models that are especially suited for online threat detection.

The networking community, users of distributed systems, and any kind of network service involving a group of stakeholders exchanging documents, data, and resources will find value in the proposed research's conclusions.

This means that more research must be done on the use of explainable AI to networking and security applications. Deep learning (DL) approaches are finding their way into intrusion detection systems. These techniques should be accompanied by a platform or process that explains and enhances the conclusions made by the DL.
Moreover, the CIC-DDoS2019 dataset [17] can be used and compared with the results on UNSW-NB15.

REFERENCES

[1] Marwa Keshk, Nickolaos Koroniotis, Nam Pham, Nour Moustafa, Benjamin Turnbull, Albert Y. Zomaya. An explainable deep learning-enabled intrusion detection framework in IoT networks, Information Sciences, Vol 639, 2023, https://doi.org/10.1016/j.ins.2023.119000.

[2] A. A. Alashhab, M. S. M. Zahid, M. Abdullahi and M. S. Rahman, "Real-Time Detection of Low-Rate DDoS Attacks in SDN-Based Networks Using Online Machine Learning Model," 2023 7th Cyber Security in Networking Conference (CSNet), Montreal, QC, Canada, 2023, pp. 95-101, http://doi.org/10.1109/CSNet59123.2023.10339791.

[3] G. Karatas, O. Demir and O. Koray Sahingoz, "Deep Learning in Intrusion Detection Systems," 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), Ankara, Turkey, 2018, pp. 113-116, http://doi.org/10.1109/IBIGDELFT.2018.8625278.

[4] Mohamed Lahby, Utku Kose, Akash Kumar Bhoi. Explainable Artificial Intelligence for Smart Cities. 1st Edition, November 2021, CRC Press, https://doi.org/10.1201/9781003172772.

[5] Alperin K. B., Wollaber A. B., Gomez S. R. Improving Interpretability for Cyber Vulnerability Assessment Using Focus and Context Visualizations, 2020 IEEE Symposium on Visualization for Cyber Security (VizSec), Salt Lake City, UT, USA, 2020, pp. 30-39, http://doi.org/10.1109/VizSec51108.2020.00011.

[6] Luckey, D., Fritz, H., Legatiuk, D., Dragos, K., Smarsly, K. (2021). Artificial Intelligence Techniques for Smart City Applications. In: Toledo Santos, E., Scheer, S. (eds) Proceedings of the 18th International Conference on Computing in Civil and Building Engineering. ICCCBE 2020. Lecture Notes in Civil Engineering, vol 98. Springer, Cham. https://doi.org/10.1007/978-3-030-51295-8_1.

[7] Javed, A.R.; Ahmed,W.; Pandya, S.; Maddikunta, P.K.R.; Alazab, M.; Gadekallu, T.R. A Survey of Explainable Artificial Intelligence for Smart Cities. Electronics 2023, 12, 1020. https://doi.org/10.3390/electronics12041020.

[8] Nwakanma, C.I.;Ahakonye, L.A.C.; Njoku, J.N.; Odirichukwu, J.C.; Okolie, S.A.; Uzondu, C.; Ndubuisi Nweke, C.C.; Kim, D.-S. Explainable Artificial Intelligence (XAI) for Intrusion Detection

and Mitigation in Intelligent Connected Vehicles: A Review. Appl. Sci. 2023, 13, 1252. https://doi.org/10.3390/app13031252.

[9] Srivastava, Gautam et al. "XAI for Cybersecurity: State of the Art, Challenges, Open Issues and Future Directions." ArXiv abs/2206.03585 (2022).

[10] Ashraf Javed , Keshk Marwa, Moustafa Nour, Abdel-Basset Mohamed, Khurshid Hasnat, D. Bakhshi Asim , R. Mostafa Reham. IoTBoT-IDS: A novel statistical learning-enabled botnet detection framework for protecting networks of smart cities, Sustainable Cities and Society, 72, 2021, https://doi.org/10.1016/j.scs.2021.103041.

[11] Dhavalkumar Thakker, Bhupesh Kumar Mishra, Amr Abdullatif, Suvodeep Mazumdar and Sydney Simpson. Explainable Artificial Intelligence for Developing Smart Cities Solutions. Smart Cities 2020, 3, 1353–1382; http://doi.org/10.3390/smartcities3040065.

[12] RNN-SHAP: https://github.com/shap/shap/issues/213 ; available : September 2023.

[13] Truong Pham N., Dzung Nguyen S., Song Thuy Nguyen V., Hong Pham B. N., Minh Dang D. N. (2023) Speech emotion recognition using overlapping sliding window and Shapley additive explainable deep neural network, Journal of Information and Telecommunication, 7:3, 317-335, https://doi.org/10.1080/24751839.2023.2187278.

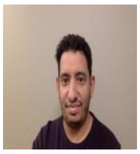
[14] Nadeem M. W., Goh H. G., Aun Y. and Ponnusamy V. "A Recurrent Neural Network based Method for Low-Rate DDoS Attack Detection in SDN," 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2022, pp. 13-18, https://doi.org/10.1109/AiDAS56890.2022.9918802.

[15] UNSW_NB15 dataset: https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15 ; available : May 2023.

[16] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, 2015, pp. 1-6, http://doi.org/10.1109/MilCIS.2015.7348942.

[17] S. S. Kona, Detection of ddos attacks using rnn-lstm and hybrid model ensemble. Ph.D. dissertation, Dublin, National College of Ireland, 2023.

AUTHORS



Ahmad Mater Aljohani was a teacher at the University of Tabuk (Computer Sciences of Tabuk, Applied College, Tabuk, Saudi Arabia). He achieved his Master degree from the University of Bedfordshire, UK and he is currently a PhD student at the University of Canberra, Australia. His research interest focuses on security of IoT networks, attacks detection, trust, privacy and DDoS attacks on IoT devices.

Ibrahim Elgendi holds a Ph.D. in Information Technology from University of Canberra, Australia. He is currently a lecturer in Networking and Cybersecurity. His research focuses on Mobile and Wireless Networks, Internet-of- Things, Machine Learning, and Cyber-Physical-Security.

© 2024 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.