International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Improved Explainable Artificial Intelligence for Intrusion Detection System

Saahira Banu Ahamed Maricar[*1], Anne Anoop[2], Betty Elezebeth Samuel[3], Anjali Appukuttan[4], Khalid Hasan Alsinjlawi[5]

Submitted: 09/12/2023 **Revised**: 20/01/2024 **Accepted**: 30/01/2024

Abstract: Cybersecurity professionals rely heavily on Intrusion Detection Systems (IDS) to identify as well as stop potential dangers. Networks may be better protected with the use of IDS. A variety of Machine Learning (ML) approaches are aimed at the development of successful IDSs. Ensemble methods in ML have a history of successful learning. This research proposes a cutting-edge IDS using ensemble methods of ML. This research used preprocessing data after the CICIDS-2017 dataset to enhance classification accuracy and suppress false positives. Using ML methods including Logistic Regression, XGBoost (XGB) ID classifiers, along with Light Gradient Boosting Machine (LGBM) classifiers, proposes an IDS. An ensemble technique classifier was applied after these models were trained, and accuracy was obtained. The suggested model also includes the Explainable Artificial Intelligence (XAI) algorithm Local interpretable model-agnostic explanation (LIME), which makes the for reliable ID easier to understand and explain. The XAI LIME is faster, more responsive, and easier to explain.

Keywords: *IDS; Ensemble techniques; CICIDS2017dataset; XAI; LIME; ML*

1. Introduction

Artificial Intelligence (AI) is the use of a representations set as well as techniques to extract useful information from a large dataset. Furthermore, people cannot put their faith in an AI system because of issues with data quality, complicated methodology, accountability, and the level of expertise of AI engineers [1]. Recent years have shown that ML-based IDSs are successful; in particular, deep neural networks advance the detection rates of IDS models. As models grow difficult, however, people have a much harder time understanding the reasoning behind their decisions [2].

To make this model more realistic, the use of well-known space data (i.e., CIA standards) and test it on an organisation disruption discovery experiment to validate this methods [3]. This paper tackles this problem by presenting an informative HPCDR ML framework. This technique locates the dangerous instructions inside an application by identifying and removing its most damaging temporary window. This allows the user to easily find the appropriate transparency [4]. A detailed examination of AI and XAI-based tactics applied in the production 4.0 scenario is offered in this article. It begins with a quick overview of the technology that enable Industry. [5].

The paper summarises recent advancements in the area and shows three classification tasks where it explains Deep Learning (DL) model predictions using LIME. In an effort to simplify them, LIME assesses these complicated models using three different categorization tasks [6]. This aims to highpoint the capabilities of using XAI for framework applications in this context. by going over the most typical problems that crop up when using XAI to these types of tasks. The next step is an analysis of the most current research trends and a review of the most recent publications on the topic [7]. It provides a taxonomy for XAI approaches that takes into account several security features and threat models related to cyber security [8].

The rest of the research is organized as follows. Section 2 reviews some of the previous studies that related to ID. The proposed method is detailed in section 3 then its results discussion is made in section 4. Section 5 gives the research conclusion followed by the references.

2. Literature Review

Abou El Houda et al. [9] provided a XAI-based outline to describe any critical DL-based choices for IoT-related IDSs. To identify disruptions associated with the Internet of Things (IoT), the system trusts on a IDS for IoT organizations, additionally this research develops using

[1]*Department of computer science, College of Computer Science and Information Technology, Mahalya Campus for Girls, Jazan University, Jazan 45142, Saudi Arabia, sahirabanuahamed@gmail.com, sahamed@jazanu.edu.sa*
[2]*Lecturer, Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia, aanup@jazanu.edu.sa*
[3]*Lecturer, Department of Computer Science and Information Technology Jazan University, Kingdom of Saudi Arabia, bsamuel@jazanu.edu.sa*
[4]*Lecturer, Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia, anarayanan@jazanu.edu.sa*
[5]*lecturer, Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia, Kalsinjlawi@jazanu.edu.sa*
** Corresponding Author Email: sahirabanuahamed@gmail.com*

deep brain organization. In addition to Deep Neural Network (DNN) based model, this structure also makes use of three basic XAI procedures: RuleFit, LIME, as well as SHapley Additive explanations (SHAP). The system may deliver both local as well as global details, which helps with the understanding of DL judgements.

Barnard et al. [10] provide a two-stage for the detection of network intrusions. By explaining supervised ID model, XGB, which uses extreme gradient boosting using the SHAP framework. In the second phase, an auto-encoder is trained using these explanations to distinguish between previously observed and new assaults. Zhang et al. the eleventh current literature on XAI techniques for digital protection applications is comprehensively reviewed in this overview. Artificial Intelligence, including ML and DL, commonly used in the domains of network including interruption finding, virus identification, and spam sorting, as a result of the quick growth of Internet-connected systems and AI in latest times. The survey's goal is to discourse by offering a thorough and current examination of XAI solutions that may be used to address issues in the field of cyber security.

Javed et al. [12] Additionally the idea of XAI for smart cities, many use cases of XAI technology, applications, challenges, another results, as well as ways to improve research. There is a lot of information on the research and activities that have gone into developing XAI for smart cities, including the attempts to standardize it. It reviews several technical challenges and summarize the lessons learned from cutting-edge research to illuminate fresh avenues for future.

Reyes et al. in [13] highlighted a wireless network IDS (WNIDS) based on ML to efficiently identify assaults on Wi-Fi networks. In the proposed WNIDS, there are two interdependent steps. At each level, a ML model is created to categorize the organization's information into either the usual or specific attack classes. The WNIDS ML model is trained and validated by the Aegean Wi-Fi Intrusion Dataset (AWID). Some story selection algorithms were used to choose the ideal set of features for the WNIDS.

Liu et al. [14] provided FAIXID, a framework that may simplify ID alerts via the use of XAI as well as data cleaning techniques. The ability to swiftly remove false positives aids cyber analysts in drawing extra learnt conclusions. Members of the Fernandez et al. [15] outlined a thorough evaluation of the evolving fuzzy systems research field is the goal of this contribution. Doing so requires asking and answering the "4 W" questions designed to shed light on the relevance and situation of the topic at hand. It will be specifically described how evolving fuzzy systems may be explained, when they were formed, what they are recycled for, as well as where researchers should concentrate their efforts in this field in the future.

Thakker et al. [16] took into consideration Semantic Web technologies and investigates the notion of "XAI" as a subset of the " XAI" issue. It is shown using a smart city flood observing application within the framework of a plan financed by the European Commission. In order to keep an eye on potential flood problems, it is necessary to track drainage and gullies in key geographical regions. One common method for dealing with this problem is to use cameras to capture real-time images of the impacted regions as they are being covered by different items, such as plastic bottles, leaves, and so on. and develop a DL classifier to detect obstructions and identify these things using the photos' coverage and existence.

2.1. Significance of the Study

This research used the CICIDS-2017 dataset to train an IDS. An important aspect of ID is selecting appropriate characteristics. The literature found that certain characteristics are necessary for all assaults, while others are either partly necessary, are not necessary at all, or are only necessary for some attacks. The CICIDS-2017 dataset contains 76 characteristics that may be used for IDS training and testing purposes. In an effort to enhance classification accuracy, the CICIDS-2017 dataset down to ten characteristics based on this research. In order to choose amongst trust measurement models, improve untrustworthy models, and get insights into text domain predictions, explanations have proved effective for both experts and non-experts. Therefore, after the application of the ML models, the authors generated LIME observations for Logistic Regression, XGB ML classifier and the LGBM classifier. Despite these efforts, there are still problems with attack detection and prediction. So, here are some things that this effort has contributed:

- Train several ML classifiers, such as LR, LGBM, and XGB, using an IoT network detail dataset.
- All of the ML models should be trained using the dataset.
- Determine the accuracy of each ML model in detecting malware in IoT networks.
- Following a thorough evaluation of each ML model, the XGB model emerged victorious. Therefore, the XGB model successfully identifies malware in IoT networks, according to research.
- The XAI algorithm LIME enhances understanding as well as explainability, which is the last step towards trustworthy ID.

3. Methodology

3.1. Dataset

Information on network traffic is included in the CICIDS-

2017 dataset, which is composed of network data. The dataset includes both normal and abnormal network data. This method involves merging four files from this dataset. Both normal network data and network data containing an anomaly will be found in each of these four files. Data for four distinct types of anomaly are included in each of the four files. This dataset is used since the network trained using this data set performs admirably in identifying networks that include abnormalities [17]. The dataset is split into training and testing data as output into data loading.

3.2. Data Loading

The data loading uses the input which is split into training data and testing data. The input of raw data is the first stage of any ML study. A dataset made up of log files or a database may include this raw information. It began by merging all eight files that made up the dataset into a single one. Pandas also provides a full set of tools for working with the data that will be loaded, including functions made by sci-kit-learn, a Python open-source ML package, and simulated data. The data loading output is fed to the preprocessing data.

3.3. Preprocessing of Data

The preprocessing data gets input from data loading. The CICIDS dataset was standardised using StandardScaler, following the principle of "garbage in, garbage out." When working with Pandas, you can simply use None or NaN to indicate that data is missing. Therefore, it is necessary to remove null values from a data frame. Due to the large volume and high number of null values in the dataset, the drop.null() method was used to eliminate the rows and columns that included these values.

In order to eliminate any unnecessary data, the dataset will undergo pre-processing. Both the efficiency and speed of training the XGB model will be enhanced as a consequence. Before training the XGB model, the dataset must undergo pre-processing to transform the data into a.csv file. Additionally, the dataset must have all superfluous data deleted. The dataset has the unnecessary labels removed. After the.csv file is formed, a new column called "label" is added to it. The information in this column determines whether the network is malicious or not. There will be a.csv file with rows representing the various networks' data and columns representing their characteristics. Since it indicates if a network has an anomaly or is benign, the 'label' column will have various values. Because of its machine-like nature, the ML model performs best when fed numerical data during training. The 'label' column's string values are thus transformed into numerical ones. Rows with the value 'benign' in the 'Label' column will be changed with 0, as well as rows with names of malware in the 'Label' column will be replaced with 1, i.e., rows representing data from a network that contains malware. Next, eliminate empty columns and identify and remove null values from rows and columns. Since some of the preprocessing data may include values in text or string format, also it need to convert them to numerical values. The classification report is calculated from preprocessing data prediction.

Data loading, preprocessing, and AI model implementation are all parts of the ML. Figure 1 displays three classic learning methods and a classifier ensemble method.

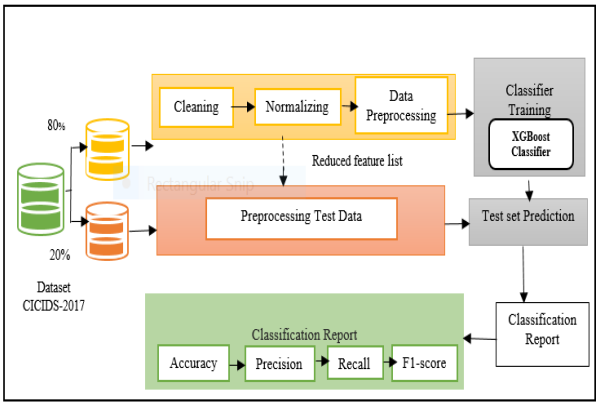

Fig .1. Model framework for the proposed work

3.4. Machine Learning

The output obtained after preprocessing is used by all the ML models as input. This research uses three distinct classifiers—Logistic Regression, the XGB ML classifier, and the LGBM classifier work on 20% testing pre-processed data as well as 80% training pre-processed data. A set of train data was recycled to train the model parameter, and one set of unseen test data was used to test it. In this approach, it controls the parameters and avoid overfitting the model.

3.4.1. Logistic Regression

Anomaly detection using ML methods has become quite common in current years [18][19]. Network ID has found success using ML, and Anomaly-based ID is fundamentally a classification issue. The capacity for computers to learn new things without being explicitly taught is a key component of ML, a subfield of AI [19]. When it comes to ID in contemporary data settings, traditional ML approaches aren't up to the task because of their shallow design [20]. One method for making predictions is logistic regression, which uses the idea of probability. Classification difficulties are its primary use. Binary classification using a sigmoid function, or logistic function, for prediction is its main purpose. Logistic regression is really a classification method, despite what its name suggests.

3.4.2. LGBM Classifier

One histogram-based decision tree approach that boosts model performance while decreasing machine execution time and memory use is LGB. According to [21], compared

to added boosting ensemble decision tree algorithms, LGBM is much more optimised. The algorithm for learning has been significantly enhanced, making it quicker, more dispersed, and more powerful. LGBM can effectively handle large-scale data flow [22]. To train decision trees and calculate the superlative split, LGBM employs a histogram-based approach and a pre-sorted technique, as do many other boosting algorithms [23]. The two novel techniques that LGBM uses are Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling (GOSS). GOSS divides the data samples in order to find a split value by down-sampling the instances according to the magnitude of the gradients. The model is designed to focus on big gradient data and exclude small gradient samples. Undertrained samples have big gradients, while well-trained samples have tiny ones. In terms of accuracy, this technique outperforms uniform random sampling. Conversely, EFB overcomes the drawbacks of conventional algorithms that rely on histograms. Unlike other decision trees, LGBM grows by adding nodes, or leaves, at the top (best first). Figure 3 shows the decision tree's development, leaf by leaf. The nodes that are green in this picture indicate the places where trees grow. The greatest possible delta value should be used for the leaf-wise expansion. Several ML issues, including classification, regression, and decision making, may be solved using LGBM [24]. There is less room for mistake or loss when trees are grown leaf-wise as opposed to level-wise.

3.4.3. XGB Classifier

The observations show that the XGB method uses very few resources for computing and takes very little time for training. Malware detection was also shown to be successful. Despite its apparent usefulness, the XGB technique has been used in very few ML classifiers. This system's suggested usage of the XGB algorithm to identify malware in an IoT setting will allow researchers to test the algorithm's limits. In order to train the XGB ML classifier will be using the CICIDS-2017, which achieved excellent results when applied to nine networks for anomaly detection ML model training. In order to prepare the datasets for training the XGB model will preprocessing the data and choose the most relevant preprocessing data.

These two desktop apps take the preprocessing data as input, detect when a button is pressed, and then feed the inputs to the trained XGB classifier that is loaded. The classifier then creates an output. The detection result will be shown on the desktop app's UI based on the classifier's output. The output will show the words "Malware detected" if the preprocessing data input corresponds to a network with malware, and "Normal" if the network is malware-free.

Algorithm 1 Proposed XGB-IDS

Input: CICIDS dataset

Output: Predict intrusions

1. Begin
2. For individually rows in dataset
3. Do Data Cleaning
4. Eliminate redundant, unrelated and fewer valuable cases
5. Fill the absent rate
6. Step 2: Data Transformation
7. If (null or non-numeric values)
8. Then
9. Transform categorical data into numeric
10. Else
11. One-hot encoder ()
12. Step 3: Data Pre-processing
13. Remove less useful or irrelevant pre-processing data
14. Step 4: Data Evaluation as well as Training
15. For collectively Learning Process Do
16. Divide the dataset into Test data along with Train data
17. Train input data
18. End for
19. Compare prediction results
20. Step 5: Final Model
21. Predict final results
22. end

3.5. X-AI Model (LIME)

To make the model more explainable, the LIME model was integrated into the ML pipeline that was created. LIME can accurately describe several ML algorithms for regression predictions by using the change in feature values of a data sample. This turns each article value into the predictor's contribution. For each data sample, an interpreter may provide a local perspective. Logistic Regression, the XGB ML classifier, and the LGBM classifier are often suggested as advanced methods to be used when accuracy has to be increased. (25, 26, 27).

Following the use of LIME, the model clarifies classifier issues. By manipulating the input data samples as well as seeing the resulting changes in predictions, the LIME method provides insight into a black-box model used in ML. It is possible to understand the predictions made at one place using the simple model by simulating the complicated

model's behaviour at another place using the LIME model. With LIME, you can work with Tableau, text, and picture data types. The results shows how LIME interpreted the CICIDS 2017 data.
The LIME algorithm described.

- There must be n instances of disruption without a little change in value for an explanation to be necessary. In order to create a local linear model around the altered observation,
- LIME uses this fabricated data.
- The results of the data that has been altered are anticipated.
- Find out how far away each affected observation is from the first observation.
- Find the score of similarity using the distance.
- The next step is to figure out how to best depict the altered data predictions using the preprocessing data.
- Using the preprocessing data that was chosen, a model is fitted to the perturbed data.
- The results are defined by the basic model's coefficients, which are also called weights.

4. Results and Discussion

4.1. Confusion Matrix

For ML classification situations where the output can be two or more modules, a confusion matrix is a performance metric.
A prediction is created once the ML model has been trained with the provided dataset as well as dealt with the numerous attacks characterized through confusion matrices using a Logistic Regression, XGB, and LGBM classifiers, as well as a classifier. Information from a confusion matrix is used to compute the evaluation reports for each classifier, including accuracy, precision, recall, and f1 score.

4.2. Accuracy

The following formula may be used to calculate accuracy using a confusion matrix in equation (1):

$$\text{Acc} = \frac{TN+TP}{\text{TN+TP+FN+FP}}$$

4.3. Precision

Using confusion matrix precision is designed, as follows in (2):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

4.4. Recall

The recall calculation is finalised using the confusion matrix, with the following procedure in (3):

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

4.5. F1 Score

The F1 Score metric, a harmonic ratio, assists in achieving a balance between recall and accuracy. (4) equates the F1 score.

$$\text{F1 score} = 2 * \left(\frac{\text{precision}*\text{recall}}{\text{precision+recall}}\right)$$

4.6. Discussion

Python will be used to implement the system that is being presented. Information retrieved from the Python 'pandas' package. The CICIDS-2017 dataset underwent feature selection using the 'SelectKBest' method imported from the 'sklearn' package. The function 'XGBClassifier ()' from the 'sklearn' library is imported to load.

The suggested model is contrasted with other classifiers in this section. Table 1 displays the outcomes of the comparison. Logistic regression, XGB, and LGBM were the three ML models used in this experiment. Almost identical in accuracy, with only subtle variations, were all of the models. Table 1 shows accuracy of 0.94 using an XGB classifier, as shown in the comparison of data below.

Table 1. Comparison Table

Method	Classes	precision	recall	F1-score	Support
Logistic Regression	0	0.67	0.76	0.72	7384
	1	0.78	0.7	0.74	9083
	Accuracy			0.73	16467
	Macro avg	0.73	0.73	0.73	16467
	Weighted avg	0.73	0.73	0.73	16467
LGBM	0	0.9	0.96	0.93	7384
	1	0.96	0.91	0.94	9083
	Accuracy			0.93	16467
	Macro avg	0.93	0.94	0.93	16467
	Weighted avg	0.94	0.93	0.93	16467
XGB	0	0.91	0.95	0.93	7384
	1	0.96	0.92	0.94	9083

	Accuracy			0.94	16467
	Macro avg	0.93	0.94	0.94	16467
	Weighted avg	0.94	0.94	0.94	16467

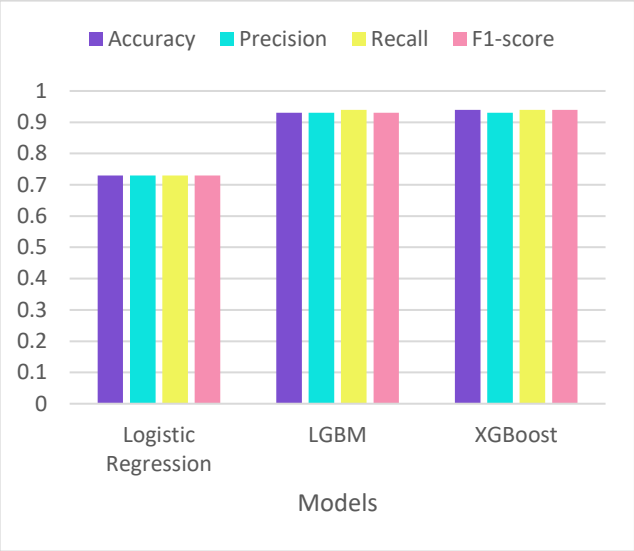

Fig.2. Comparison models and results

Figure 2 is a bar graph that shows the results of comparing all of the ML models that were employed in this research. On a scale from 0 to 1, the best outcomes are those that are closest to 1.

The LIME observations for the logistic regression, XGB, and LGBM algorithms are shown in Figures 3, 4, and 5, respectively. LIME explains the initial assignment of the probability. In order to determine the forecast, the actual class of the dependent variable is compared with the likelihood values. Additionally, there was a significant difference in the weights that the various algorithms assigned to each characteristic. Next, the feature-value table was colour-coded to show that each feature was either orange (DDOS), blue (NOT1), or not present (BENIGN). On its own, the Feature-Value table displays the real feature values for that specific record of LIME observation across all ML models:

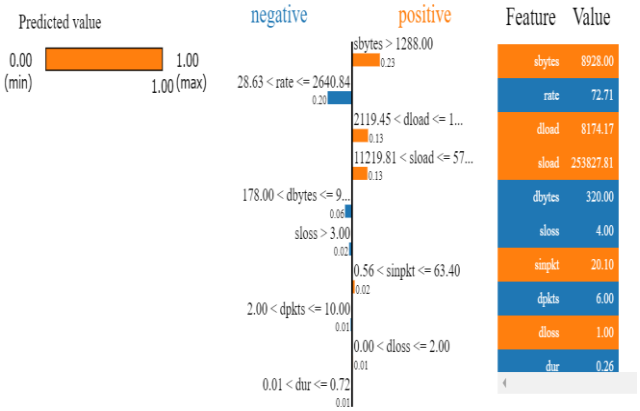

Fig.3. Logistic Regression

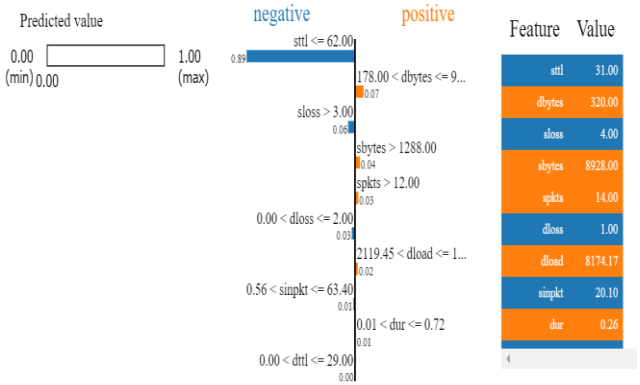

Fig.4. XGB

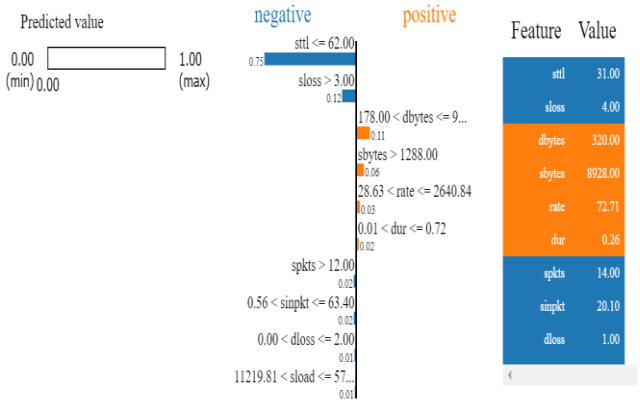

Fig.5. LGBM

5. Conclusion and Future Work

Logistic Regression, the XGB ML classifier, and the LGBM experiments are some of the ML methods suggested in this research for use in an IDS. By using a classifier, an ensemble approach, it was possible to train these models accurately. According to this research, trust is crucial for successful interactions between humans and machines. LIME provides a straightforward and easy-to-understand description of predictions using a modular and extensible methodology. Having a clear understanding of prediction is

crucial when choosing representative models. It helps with model selection, trust evaluation, fixing unreliable models, and understanding forecasts aimed at both system specialists and those without specialised training. In order to comprehend the model's forecast, this training suggests using a LIME explainable framework after training a collection of ML models. The IDS prediction accuracy was enhanced by the ML ensemble, and the prediction performance of the LGBM, XGB, and Logistic Regression methods was shown in the LIME explanation graphs.

Extension of this work to apply explainability to DL-based IDS analysis is possible; other XAI models like SHAP, AIX360 as well as Deep Lift might also be explored. An app for analysing data in real-time and evaluating the performance of predictions is also in the works. We want to use XAI in the future on a number of complicated datasets, including ToN_IoT.

Conflict of Interest

None

References

[1] Mahbooba, Basim, Mohan Timilsina, Radhya Sahal, and Martin Serrano. "Explainable artificial intelligence (XAI) to enhance trust management in Intrusion Detection systems using decision tree model." Complexity, pp. 1-11,2021.

[2] Wang, Maonan, Kangfeng Zheng, Yanqing Yang, and Xiujuan Wang. "An explainable Machine Learning framework for Intrusion Detection systems." IEEE Access 8 73127-73141, 2020.

[3] Islam, Sheikh Rabiul, William Eberle, Sheikh K. Ghafoor, Ambareen Siraj, and Mike Rogers. "Domain knowledge aided explainable artificial intelligence for Intrusion Detection and response." arXiv preprint arXiv:1911.09853 ,2019.

[4] Kuruvila, Abraham Peedikayil, Xingyu Meng, Shamik Kundu, Gaurav Pandey, and Kanad Basu. "Explainable Machine Learning for Intrusion Detection via hardware performance counters." IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 41, no. 11 pp. 4952-4964, 2022.

[5] Ahmed, Imran, Gwanggil Jeon, and Francesco Piccialli. "From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where." IEEE Transactions on Industrial Informatics 18, no. 8, pp. 5031-5042,2022.

[6] Mathews, Sherin Mary. "Explainable artificial intelligence applications in NLP, biomedical, and malware classification: a literature review." In Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 2, pp. 1269-1292. Springer International Publishing, 2019.

[7] Machlev, R., L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron. "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities." Energy and AI 100169,2022.

[8] Kuppa, Aditya, and Nhien-An Le-Khac. "Black box attacks on explainable artificial intelligence (XAI) methods in cyber security." In 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020.

[9] Abou El Houda, Zakaria, Bouziane Brik, and Lyes Khoukhi. "“why should i trust your ids?”: An explainable deep learning framework for Intrusion Detection systems in internet of things networks." IEEE Open Journal of the Communications Society 3 pp .1164-1176,2022.

[10] Barnard, Pieter, Nicola Marchetti, and Luiz A. DaSilva. "Robust Network Intrusion Detection through Explainable Artificial Intelligence (XAI)." IEEE Networking Letters 4, no. 3, pp.167-171, 2022.

[11] Zhang, Zhibo, Hussam Al Hamadi, Ernesto Damiani, Chan Yeob Yeun, and Fatma Taher. "Explainable artificial intelligence applications in cyber security: State-of-the-art in research." IEEE Access ,2022.

[12] Javed, Abdul Rehman, Waqas Ahmed, Sharnil Pandya, Praveen Kumar Reddy Maddikunta, Mamoun Alazab, and Thippa Reddy Gadekallu. "A survey of explainable artificial intelligence for smart cities." Electronics 12, no. 4, 1020, 2023.

[13] A. Reyes, Abel, Francisco D. Vaca, Gabriel A. Castro Aguayo, Quamar Niyaz, and Vijay Devabhaktuni. "A Machine Learning based two-stage Wi-Fi network Intrusion Detection system." Electronics 9, no. 10, 1689,2020.

[14] Liu, Hong, Chen Zhong, Awny Alnusair, and Sheikh Rabiul Islam. "FAIXID: A framework for enhancing ai explainability of Intrusion Detection results using data cleaning techniques." Journal of network and systems management 29, no. 4, 40,2021.

[15] Fernandez, Alberto, Francisco Herrera, Oscar Cordon, Maria Jose del Jesus, and Francesco Marcelloni. "Evolutionary fuzzy systems for explainable artificial intelligence: Why, when, what for, and where to?." IEEE Computational intelligence magazine 14, no. 1, pp.69-81, 2019.

[16] Thakker, Dhavalkumar, Bhupesh Kumar Mishra, Amr Abdullatif, Suvodeep Mazumdar, and Sydney Simpson. "Explainable artificial intelligence for

developing smart cities solutions." Smart Cities 3, no. 4, pp. 1353-1382,2020.

[17] Elmrabit, N., Zhou, F., Li, F. and Zhou, H., "Evaluation of Machine Learning Algorithms for Anomaly Detection." [online] IEEE Xplore,2020.

[18] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of Deep Learning-based network anomaly detection," Cluster Comput., vol. 22, no. S1, pp. 949–961, 2019.

[19] Imteaj, Ahmed, and M. Hadi Amini. "Leveraging Asynchronous Federated Learning to Predict Customers Financial Distress." Intelligent Systems with Applications ,2022.

[20] M. M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, and G. Fortino, "A hybrid Deep Learning model for efficient Intrusion Detection in big data environment," Inf. Sci. (Ny), vol. 513, pp. 386–396, 2020

[21] S. Seth, G. Singh, and K. K. Chahal, ''A novel time efficient learning-based approach for smart Intrusion Detection system,'' J. Big Data, vol. 8, no. 1, pp. 1–28, Dec. 2021.

[22] D. Jin, Y. Lu, J. Qin, Z. Cheng, and Z. Mao, ''SwiftIDS: Real-time Intrusion Detection system based on LightGBM and parallel Intrusion Detection mechanism,'' Computer. Security., vol. 97, pp. 1–17, Oct. 2020.

[23] Md. K. Islam, P. Hridi, Md. S. Hossain, and H. S. Narman, ''Network anomaly detection using LightGBM: A gradient boosting classifier,'' in Proc. 30th Int. Telecommun. Netw. Appl. Conf. (ITNAC), Melbourne, VIC, Australia, pp. 1–7, Nov. 2020.

[24] D. Rani, N. S. Gill, P. Gulia, and J. M. Chatterjee, ''An ensemble-based multiclass classifier for Intrusion Detection using Internet of Things,'' Comput. Intell. Neurosci., vol. 2022, pp. 1–16, May 2022

[25] Tjoa, E.; Guan, C. A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transact. Neural Netw. Learn. Syst.*, *32*, 4793–4813,2020.

[26] Wolf, C.T. Explainability scenarios: Towards scenario-based XAI design. In Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Ray, CA, USA, 17–20; pp. 252–257, March 2019.

[27] Das, A.; Rad, P. Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *arXiv* 2020, arXiv:2006.11371.

[28] Byrne, R.M.J. Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning. *IJCAI* 2019, 6276–6282.

[29] Booij, T.M.; Chiscop, I.; Meeuwissen, E.; Moustafa, N.; Hartog, F.T.H.d. ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets. *IEEE Internet Things J.* , *9*,pp. 485–496,2022.

[30] Patil, Shruti, Vijayakumar Varadarajan, Siddiqui Mohd Mazhar, Abdulwodood Sahibzada, Nihal Ahmed, Onkar Sinha, Satish Kumar, Kailash Shaw, and Ketan Kotecha. "Explainable artificial intelligence for Intrusion Detection system." *Electronics* 11, no. 19, 2022.

[31] Jairu, Pankaj, and Akalanka B. Mailewa. "Network anomaly uncovering on CICIDS-2017 dataset: a supervised artificial intelligence approach." In 2022 IEEE International Conference on Electro Information Technology (eIT), pp. 606-615. IEEE, 2022.

[32] Bhardwaj, Parth. "Finding IoT privacy issues through malware Detection using XGBoost Machine Learning technique." PhD diss., Dublin, National College of Ireland, 2022.

[33] Rani, Deepti, Nasib Singh Gill, Preeti Gulia, Fabio Arena, and Giovanni Pau. "Design of an Intrusion Detection Model for IoT-Enabled Smart Home." IEEE Access, 2023.