Explainable Intrusion Detection for Internet of Medical Things

Shafique Ahmed Memon[1], Uffe Kock Wiil[2][a] and Mutiullah Shaikh[2]
[1]IMCS, University of Sindh, Jamshoro, Pakistan
[2]SDU Health Informatics and Technology, University of Southern, Denmark

Keywords: Intrusion Detection (ID), Explainable AI (XAI), Internet of Things (IoT), Internet of Medical Things (IoMT), Message Queuing Telemetry Transport (MQTT), Particle Swarm Optimization (PSO).

Abstract: IoMT sensors are used for continuous real-time remote monitoring of patients' health indicators. IoMT integrate several devices to capture sensitive medical data from devices such as implants and wearables that results in cost-effective and improved health. In IoT settings, the Message Queuing Telemetry Transport (MQTT) protocol is frequently used for machine-to-machine data transfer. However, secure transmission of sensitive health data is critical because these devices are resource constrained and are more vulnerable to MQTT based threats including brute force attack. This warrants a robust, effective, and reliable threat mitigation mechanism, while maintaining a fine balance between accuracy and interpretability. Based on a comprehensive overview of previous work and available datasets, we propose an explainable intrusion detection mechanism to detect MQTT-based attacks. The MQTT-IOT-IDS2020 dataset is used as a benchmark. Particle swarm optimization (PSO) is used for the selection of optimal features from the dataset. The performance of ten machine learning (ML) methods is evaluated and compared. The results demonstrate excellent classification accuracies between 97% and 99%. We applied LIME explanations to increase human interpretability for the best performing model.

1 INTRODUCTION

The Internet of Medical Things (IoMT) is a subcategory of the Internet of Things (IoT) that focuses on the application of connected devices in the healthcare domain. These devices enable continuous monitoring and real-time data transmission, leading to improved patient outcomes, reduced healthcare costs, and increased overall efficiency. The IoMT eco-system is vast, encompassing wearable devices, remote patient monitoring systems, and smart hospital infrastructure. IoMT devices are highly sensitive and critical as they transmit sensitive data which is used for diagnosis and treatment of patients. (Alsabah et al., 2021), (Asghar et al., 2022). The sensitive health data is vulnerable to unauthorized access and data breaches. Among several communication and messaging protocols that ensure machine-to-machine data transfer among devices, MQTT (Message Queuing Telemetry Transport) is commonly used and widely adopted in IoT network environments because of its capability to support resource constrained devices. (Yao et al., 2019), (Khan et al., 2021), (Hindy et al., 2021).

The legacy threats detection approaches have proved ineffective amidst the above scenario. The network traffic generated by such a huge number of devices makes the security of IoT/IoMT structure a critical factor to reckon with. The security complications necessitate and underlie the significance of a reliable and trustworthy security mechanism for IoT (Fraile et al., 2018), (Ruzafa-Alcazar et al., 2021).

There is growing interest in using machine learning (ML) and deep learning (DL) driven intrusion detection and security approaches to ensure trustworthy, resilient, and reliable cyber threat detection mechanisms. Several techniques have been used in previous studies including. Based on the overview presented in Section 2, we identified relevant areas that needs to be explored further:

- Few studies use feature selection to optimize model performance. Non-representative and

[a] https://orcid.org/0000-0001-6898-4083

Memon, S., Wiil, U. and Shaikh, M.
Explainable Intrusion Detection for Internet of Medical Things.
DOI: 10.5220/0012210300003598
In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023) - Volume 3: KMIS*, pages 40-51
ISBN: 978-989-758-671-2; ISSN: 2184-3228

irrelevant features lead to increased complexity, model convergence, and misclassification.

- Few studies focus on explainability aspects of their models to increase human interpretability. Lack of model explainability leads to lack of trust in the results.

This paper presents a novel approach towards explainable intrusion detection (XID) for IoMT. We propose an intelligent framework that combines Particle Swarm Optimization (PSO) and several ML techniques to detect impending threats to IoMT and health data. To the best of our knowledge, the proposed approach with LIME-based explanations using the MQTT-IOT-IDS2020 dataset has not been used in previous studies. The contributions of this study are:

- The use of PSO to select the optimal set of features to increase model performance.
- Evaluation and comparison of ten different ML classifiers to develop an effective model with excellent performance.
- The use of LIME on the best performing model to increase model insights and improve human interpretability.

The rest of the paper is structured as follows: Section 2 discusses the background and related work. Section 3 describes the proposed approach for XID in IoMT. Section 4 includes the evaluation of model performance, results, and discussion. Section 5 concludes the paper and describe future work.

2 BACKGROUND AND RELATED WORK

This section mainly focuses on describing major security concerns and challenges of IoMT as well as the reviewing state-of-the-art research related to the use of artificial intelligence (AI) for intrusion detection and threat mitigation strategies and techniques. This section also highlights the limitations of previous work.

2.1 IoT/IoMT Security Challenges

IoT devices are resource-constrained which make IoT based infrastructure are vulnerable to several threats including: DoS, DDoS, Mirai botnet, Spying, Scan, Fuzzers, Backdoor, Packet Flooding, Vulnerability Attack, Blackhole Attack, Sybil Attack, Sinkhole Attack, Clone Attack and hello, flood attack, man-in-the-middle (MitM), CPI Eavesdrop-ping, and system identification attack. As such, the IoMT nodes are super critical, being resource-constrained are vulnerable and can be comprised. The intruders can take advantage of these weaknesses by targeting them and can cause severe damages in terms of compromised data availability and data integrity (Jan et al., 2019), (Vargas et al., 2021), (Bagaa et al., 2020).

2.2 Intrusion Detection Systems

An intrusion detection system (IDS) is a threat or cyber-attack detection mechanism (originating from 1999) as an approach that seeks to address the security challenges by inspecting the data packets (Elrawy et al., 2018). According to Man et al. (2021), intrusion detection is of two kinds, mis-use-based and anomaly-based. The misuse-based methods detect the threat based on some already known or defined patterns in the network traffic. An anomaly is an unusual or abnormal activity in the network traffic making it suspicious. Hence, anomaly-based intrusion detection is used to segregate such patterns in the traffic for maintaining the da-ta privacy and detect malicious traffic patterns (Imrana et al., 2021), (Belavagi & Muniyal, 2016).

The enormous number of IoMT devices generates a large volume of data that requires further intelligent data analysis and processing methods such DL. Latif et al., 2020, stated that notwithstanding the fact the IoMT is improving the efficiency, they are less secure and remain always prone to varied types of attacks.

Keeping in mind the limitations and specificity of IoMT devices, traditional intrusion detection mechanism become unsuitable and need to be improved (Jan et al., 2019). Man et al. (2021) pointed out that that ML approaches have the capability to efficiently classify the anomalous trends in the network traffic thereby detecting the threat. ML methods learn to detect such activity in the network that can be classified as anomaly or a threat. The capability of ML classification methods to efficiently classify the anomaly-based threats has inspired many researchers to use these methods extensively.

2.3 Artificial Intelligence for Cyber Security

In this section, we provide a comprehensive overview of previously used AI techniques for cyber security including ML, DL, PSO, and feature selection.

ML algorithms can analyze and identify malicious patterns from huge data and network traffic to automate the threat detection and response mechanism. ML can be trained on large datasets for a wide range of security activities including: (i) identifying and classifying different types of malwares, (ii) intrusion detection, (iii) anomalies, (iv) network traffic analysis (Xin et al., 2018). Similarly, various DL algorithms have been proposed, including Restricted Boltzmann Machine (RBM), Convolutional Neural Network (CNN), Autoencoder, and Recurrent Neural Networks (RNN) (Xin et al., 2018), (Kirsch, 2018). Recent work has focused on various AI algorithms, including:

- Decision Tree (DT) has been used for intrusion detection, vulnerability assessment, and network traffic analysis (Farnaaz and Jabbar, 2016).
- Random Forest (RF) has been used for malware detection, network intrusion detection, and classification of security events (Yang and Sinaga, 2019).
- Support Vector Machine (SVM) has been used for intrusion detection, malware detection, and identification of network attacks (Thaseen and Kumar, 2017).
- Clustering algorithms, such as k-means and hierarchical clustering, have been used for anomaly detection, network traffic analysis, and malware classification (Yang and Sinaga, 2019), (Kaluarachchi et al., 2021).
- DL algorithms such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders, have been used for malware detection, intrusion detection, and phishing detection (Peng et al., 2018), (Ahmad et al., 2018), (Farnaaz and Jabbar, 2016).
- Federated Learning (FL), which is a decentralized approach to training ML models, has recently received attention in relation to ID (Popoola et al., 2021).
- PSO has been used in various cyber security applications such as ID, malware detection, phishing detection, feature selection, and parameter optimization (Chaddad et al., 2023), (Adnan et al., 2021).

Feature selection is a critical step in developing ML models, as non-representative and irrelevant features can introduce noise into the dataset and lead to overfitting issues. To avoid these problems, datasets are often reduced to a relevant subset of features. Appropriate feature selection leads to optimized models that enhance overall performance and accuracy while substantially reducing training time. Unimportant features can also be computationally burdensome and costly. As a result, feature selection is a crucial technique for creating a viable subset of features in machine learning models. Commonly used feature selection approaches include wrapper, filter, and intrinsic/hybrid methods (Mukherjee and Sharma, 2012), (Gad et al., 2021), (Jiang et al., 2020), (Awotunde et al., 2021), (Panda et al., 2021), (Kumar et al., 2021).

Table 1 provides an overview of previous related studies that have used feature selection as well as the datasets used in these studies (see Section 2.4 for an overview of datasets).

Table 1: Overview of studies using feature selection.

Study	*Feature selection technique*	*Dataset*
Kumar et al., 2021)	Hybrid Feature Reduced Approach using Correlation coefficient, random forest, mean decrease accuracy and gain ratio.	NSL-KDD, BoT-IoT and DS2OS
Tsang and Kwong, 2005	Used four unsupervised technique for feature extraction.	KDD-Cup99
Wu et al., 2020	Feature-weighted support vector machines (WSVM)	NSL-KDD
Kasongo and Sun, 2019	Filter based feature selection algorithm.	NSL-KDD)
(Mukherjee and Sharma, 2012	Proposed Feature Vitality Based Reduction Method (FVBRM)	NSL-KDD dataset.
Chen and Hao, 2019	PCA–LDA	--
Panda et al., 2021	K-Medoid sampling and scatter search-based feature	UNSW-NB15
Ding et al., 2020	Stack Sparse Autoencoder (sSAE)	NSL-KDD, UNSW_NB15
Alqahtani et al., 2020	Fisher-score-based feature selection method along with a genetic-based extreme gradient boosting (GXGBoost) model.	N-BaIoT
Derhab et al., 2020	Feature space reduction and feature transformation technique.	Bot-IoT
Awotunde et al., 2021	Hybrid rule-based feature selection	NSL-KDD and UNSW-NB15

2.4 Publicly Available Datasets

A comprehensive review of previous work has revealed several publicly available datasets that has been used in previous studies:

- DARPA1998, KDD99, and NSL-KDD: DARPA1998 dataset is one of the most used datasets in IDS that contains four malware traffic types i.e., DoS, Probe, U2R, and R2L as well as normal traffic data (Wang et al., 2017). The KDD99 dataset was created from DARPA1998 dataset and contains 41 traffic characteristics. The KDD99 was improved and enhanced and named NSL-KDD; it contains 125,973 records. Conversely, these versions have become obsolete. The reason for their being outdated is that these versions no longer represent latest threat traffic. Moreover, these datasets cannot be used for IoT as they have non IoT representative network traffic. (Wang et al., 2017).
- WSN-DS: This dataset was created for detection of threat in wireless sensor networks (WSN). This dataset contains four types of DoS attacks i.e., Blackhole, Grayhole, Flooding, and Scheduling having 23 features (Siddiqi and Pak, 2021).
- ISCX: This ID dataset was developed in 2012 having five types of attack traffic i.e., BruteForce SSH, DDoS, HttpDoS, and Infiltrating along with normal traffic [73]. For the development of this dataset, a real time simulator was used to generate simulated attack traffic. The attack types are close to real world. This dataset contains 82 features, five classes, and 73566 samples. However, this dataset does not incorporate any new security events against IoT sensors (Moustafa, 2021), (Wang et al., 2017).
- CICIDS2017: The CICIDS (Canadian Institute for Cyber security Intrusion Detection System) was created based on diverse normal and hack scenarios. Concept of data profiling was employed. The CICIDS dataset contain 79 features, 11 attack classes (including Benign, BoT, DDoS, DoS/DDoS, PortScan, SSH, etc.), and 786,633 samples (Wang et al., 2017).
- UNSW-NB15: The UNSW-NB15 dataset was created at the University of New South Wales (UNSW) Canberra to evaluate IDS. It consists of 162724 samples with a total of 10 classes, including 9 attack classes and one normal class, and has 42 features. However, the dataset does not include security features related to IoT or operating systems (Moustafa, 2021), (Wang et al., 2017).
- Bot-IoT: The Bot-IoT dataset was created at the UNSW with the primary goal of evaluating intrusion detection and network forensics systems. This dataset comprises 14 features, including several IoT system features, and has 7 different attack classes with a total of 74680 attack samples. However, the dataset does not include hacking vectors against IoT systems (Moustafa, 2021), (Siddiqi and Pak, 2021).
- TON_IoT: This dataset is a database from the UNSW that addresses the limitations of existing datasets by incorporating four heterogeneous data sources - Telemetry datasets of IoT and IoT sensors, operating systems, and network traffic. It contains the latest attack information related to operating systems and IoT/IoMT devices. It contains instances of normal and attack data. This dataset can be used to develop machine learning models. However, the dataset suffers from an imbalance class issue as the number of normal records exceeds the number of abnormal records (Siddiqi and Pak, 2021), (Priya et al., 2020).
- Kyoto: The Kyoto dataset was created using honeypots at Kyoto University. The Kyoto dataset contains traffic generated from honeypots and not from real world environment having 24 statistical features. Out of these 24 features, 14 features are adopted from KDDCup 99. In addition to this, 10 features were derived from honeypot's network traffic system (Siddiqi & Pak, 2021). The drawback of this dataset is also non representative of IoT systems (Moustafa, 2021).
- IoT-23: This dataset is considered as one of the latest datasets containing information related to IoT traffic and network monitoring. This dataset was developed as a part of Stratosphere project and contains 20 types of attacks including Mirai and Torri for IoT devices and 3 types of benign IoT devices traffic (Tsang and Kwong, 2005).
- N-BaIoT: This dataset was developed through a simulated IoT environment to detect botnet whereby IoT devices were connected through WiFi, access points to switch, and router data was gathered using Wireshark (Siddiqi and Pak, 2021).
- MQTT-IoT-IDS2020: Message Queuing Telemetry Transport (MQTT) is one of the IoT standard protocols used for machine-to-machine communication. MQTT-IoT-IDS2020 dataset is

simulated representation of an MQTT-based network. This dataset consists of five different scenarios, which include: (1) normal operation, (2) aggressive scan, (3) UDP scan, (4) Sparta SSH brute-force, and (5) MQTT brute-force attack (Hindy et al., 2020).

Table 3 (Appendix) provides an overview of the encountered publicly available datasets.

2.5 Overview of Related Work

There are several studies using either ML, DL, or a hybrid approach for ID. Table 4 (Appendix) provides an overview of multiple studies. The overview focuses on the used AI technique(s), the used dataset(s) and the obtained accuracy of each study.

2.6 Summary

The above review of related work reveals different state-of-the-art techniques based on ML, DL, and evolutionary computing used for intrusion detection demonstrating good results in classifying threat types. The review also identified areas that need further studies, such as the use of feature selection techniques to improve model performance and the use of explainable AI techniques for increased trust in model predictions.

3 MATERIALS AND METHODS

This section contains a description of the dataset and the steps involved in developing the proposed model. Figure 1 provides an overview of the steps.

The developed model aims to achieve a binary classification of ATTACK and NO ATTACK. In doing so we have used the MQTT-IOT-IDS2020 dataset as benchmark. The dataset contains several features. Therefore, we have used the PSO algorithm to achieve an optimal feature subset. The feature subset was used as input for machine learning classifiers. The performance of ten ML methods was evaluated and compared. Finally, the best performing classifier was explained using LIME to increase human interpretability.

3.1 Dataset

The MQTT-IOT-IDS2020 generated by Hindy et al. (2020) is publicly available. This dataset contains both normal and malicious MQTT traffic. The MQTT-IOT2020 datasets contains normal, uni-flow and bi-flow. In this study, we have used bi-flow brute force scenario category dataset which has a total of 16696 instances (2152 normal and 14544 attack) and contains total 32 features inclusion the binary attack class “is attack” feature having 0 for no attack and 1 representing the attack (Hindy et al., 2020). The features of the dataset are:

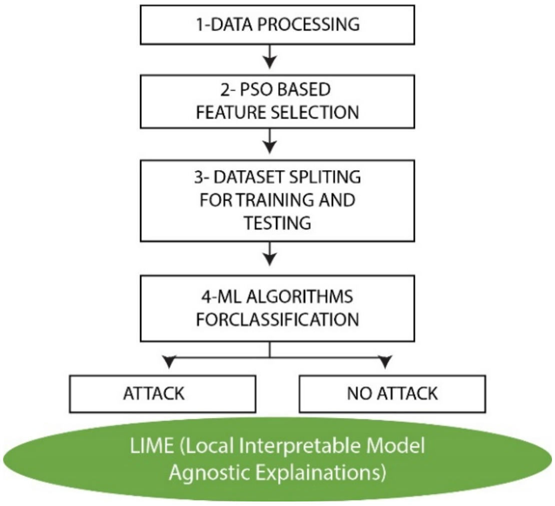
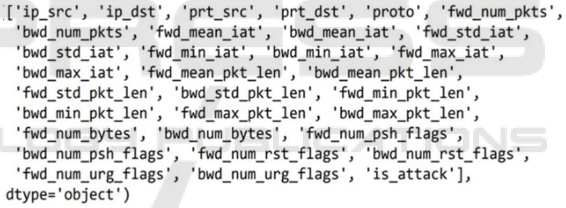

1-DATA PROCESSING
2- PSO BASED FEATURE SELECTION
3- DATASET SPLITING FOR TRAINING AND TESTING
4-ML ALGORITHMS FORCLASSIFICATION
ATTACK
NO ATTACK
LIME (Local Interpretable Model Agnostic Explainations)

Figure 1: Overview of machine learning pipeline.


```
['ip_src', 'ip_dst', 'prt_src', 'prt_dst', 'proto', 'fwd_num_pkts',
 'bwd_num_pkts', 'fwd_mean_iat', 'bwd_mean_iat', 'fwd_std_iat',
 'bwd_std_iat', 'fwd_min_iat', 'bwd_min_iat', 'fwd_max_iat',
 'bwd_max_iat', 'fwd_mean_pkt_len', 'bwd_mean_pkt_len',
 'fwd_std_pkt_len', 'bwd_std_pkt_len', 'fwd_min_pkt_len',
 'bwd_min_pkt_len', 'fwd_max_pkt_len', 'bwd_max_pkt_len',
 'fwd_num_bytes', 'bwd_num_bytes', 'fwd_num_psh_flags',
 'bwd_num_psh_flags', 'fwd_num_rst_flags', 'bwd_num_rst_flags',
 'fwd_num_urg_flags', 'bwd_num_urg_flags', 'is_attack'],
dtype='object')
```

3.2 Machine Learning Pipeline

Step 1: Data Preprocessing. Firstly, IP address features, i.e., the ip_src and ip_dst, were handled through encoding. Secondly, to interpret and easily compare that data Min-max normalization was used as preprocessing technique. (Sadaf and Sultana, 2020). The objective was to scale the numeric features to some specific minimum and maximum normalized range, i.e., between 0 and 1. The min-max normalization is mathematically presented as: X_normalized = (X - X_min)/(X_max - X_min). Some of the features in the dataset were also treated using label encoding and one-hot encoding technique (Gad et al., 2021).

Step 2: Feature Selection. Feature selection is of vital importance in ML processes. Feature selection can help avoid dimensionality and over fitting issues and improves model accuracy and performance. Feature selection also enhance model explainability and interpretability. It is also noticeable that most ML based IDS suffer from explainability and interpretability. Explainability of AI models is essential as it unfolds insights of black box models (Ruzafa-Alcazar et al., 2021), (Kasongo, 2021), (Fatani et al., 2022), (Jabbar et al., 2017). Therefore, it has been recommended that AI models be explainable and interpretable (Duque and Omar, 2015), (Saranya et al., 2020), (Priya et al., 2020). Keeping this in mind, PSO based feature selection was performed and as a result, out of the original 32 features, the following 22 features we selected using:

{'prt_src', 'prt_dst', 'proto', 'fwd_num_pkts', 'bwd_num_pkts', 'bwd_mean_iat', 'fwd_std_iat', 'bwd_min_iat', 'fwd_max_iat', 'bwd_max_iat', 'fwd_mean_pkt_len', 'bwd_mean_pkt_len', 'fwd_std_pkt_len', 'fwd_min_pkt_len', 'bwd_min_pkt_len', 'bwd_max_pkt_len', 'fwd_num_bytes', 'bwd_num_bytes', 'fwd_num_rst_flags', 'bwd_num_rst_flags', 'bwd_num_urg_flags'}.

Step 3. Dataset Splitting for Training and Testing. The selected optimal subset of features was used as the final dataset. The final dataset was further split into two subsets, i.e., 80% for training and 20% for testing and validating purposes.

Step 4. ML Algorithms for Classification. Ten different ML techniques were used in the study: LR, NB, KNN, SVM, MLP, DT, ET, RF, XGB, and SGD. Each technique has its pros and cons. LR is a simple, easy, and powerful algorithm, Naive Bayes (NB) is a probabilistic algorithm that is also simple and fast. KNN is a non-parametric and lazy algorithm that is likely to be computationally expensive for large datasets. SVM is a powerful method to handle linear and non-linear data. SVM finds hyper plane for classifications tasks. MLP is a type of ANNs and consists of multiple layers. DT is also a supervised, simple, and interpretable technique. RF is an ensemble method which works by constructing several trees for data. ET is also an ensemble method and is an extension of RF. XGB is an optimized implementation of gradient boosting. SGD is an optimization technique.

4 RESULTS

In this section, the evaluation and comparison of the above-mentioned ML models are presented. The performance of the models is evaluated using typical performance metrics, i.e., accuracy, precision, recall, and F1_Score based on confusion matrix indicators such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These metrics suggest that all models have achieved very good results, i.e., between 97% to 99%. Table 2 provides an overview of the metrics for the ten ML techniques.

The proposed techniques achieved excellent accuracies. Hindy et al. (2020) also achieved excellent classification accuracy on the same dataset, however no feature reduction technique or any justification of selected features has been presented in the paper. Another study by Vigoya et al. (2021) validated two IoT datasets and obtained best mean accuracy of 99% for the tree-based models. Khan et al. (2021) also achieved more than 99% accuracy with a DNN-based model. However, we did not find studies using PSO based feature selection with classification techniques using MQTT-IOT-IDS2020. Also, none of these studies provide explainability methods to facilitate human interpretability.

Table 2: Model performance.

Model	Accuracy	F1_Score	Precision	Recall
LR	0.97	0.98	1.0	0.98
NB	0.97	0.98	1.0	0.98
KNN	0.97	0.98	0.98	0.93
SVM	0.97	0.98	1.0	0.98
MLP	0.97	0.98	0.97	0.89
DT	0.99	0.99	0.99	0.98
ET	**0.99**	**0.99**	**0.99**	**0.99**
RF	0.99	0.99	0.99	0.99
XGB	0.99	0.99	0.99	0.98
SGD	0.97	0.98	1.0	0.98

4.1 LIME for Human Interpretability

According to Ribeiro and Guestrin (2016), model evaluation solely based on accuracy does not suffice as the real time data may differ significantly. Due to their complex and black box nature, ML based solutions (including IDS) suffer from lack of trust and reliability (Yao et al., 2019), which has resulted in the recent focus on explainable AI (XAI) as it offers interpretation and understandability to the machine learning model predictions (Ribeiro and Guestrin, 2016), (Adadi and Berrada, 2018). In this paper, LIME has been used to provide XID (Chen and Lee, 2020), (Houda et al., 2022).

LIME stands for Local Interpretable Model-agnostic Explanations. LIME enhances the interpretability of machine learning models. LIME interprets the pre-diction of any classifier or repressor thereby making the model more trustworthy.

LIME provides local interpretability that is based on inspecting features of a specific instance. Figure 2 provides an example of the human interpretable output of a classification of a particular instance (in this case an attack instance).

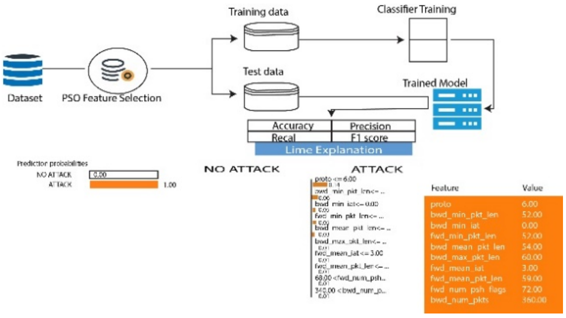


Figure 2: LIME explanations.

The visualization in Figure 2 consists of three parts. The left part (Prediction probabilities) reflects 100% confidence interval for detection of attack. The middle part shows the feature importance score on the selected sample. For instance, for proto with 14% feature importance, bwd_min_pt_len 6% and so on. This part shows the top ten features and their respective values. The features highlighted in orange (right part) show contributions toward class 1 (ATTACK), whereas features highlighted in blue contribute toward class 0 (NO ATTACK). Figure 3 visualize the LIME generated feature importance.

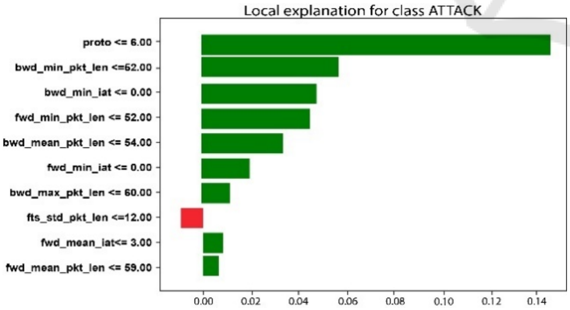


Figure 3: LIME generated feature importance.

The bar graph in Figure 3 reflects local feature explanations for attack class with features and their respective values generated by LIME showing contribution to classifications, such as ('proto <= 6.00', 0.144), ('bwd_min_pkt_len <= 52.00', 0.056), ('bwd_min_iat <= 0.00', 0.048), ('fwd_min_pkt_len <= 52.00', 0.045), ('bwd_mean_pkt_len <= 54.00', 0.033), ('fwd_min_iat <= 0.00', 0.019), ('bwd_max_pkt_len <= 60.00', .0113), ('fwd_std_pkt_len <= 12.00', -0.011), ('fwd_mean_iat <= 3.00', 0.008), and ('fwd_mean_pkt_len <= 59.00', 0.007), respectively.

5 DISCUSSION

There is growing interest in using ML and DL driven ID and security approaches. However, the issue of trustworthiness of ML and DL models has raised focus on the interpretability of these approaches.

This study demonstrates that PSO can be used to optimize the selection of relevant features for IDS. Using PSO, the features were reduced from 32 to 22. PSO helped identify an optimal subset of relevant and informative features, which lead to significantly improved model accuracy and increased model interpretability. Furthermore, given the overhead and complexity of ID techniques, the reduced feature size also leads to efficient utilization of resources given the context of resource constrained devices.

LIME explanations demonstrate how each selected feature contributes to classification both at the class level (ATTACK or NO ATTACK) and at the instance level. See Figures 3 and 2, respectively.

The study showcases that the chosen hybrid approach using PSO for feature selection, ML models for classification, and LIME based explanations for model interpretability is a viable solution for accurate XID.

The study also emphasizes the need to reconsider the complex traffic patterns and the key security concerns and challenges that data driven healthcare faces.

The key lessons learnt include that evolutionary computing techniques can be employed for an efficient and optimal feature selection. This improves model performance and interpretability and enables more informed decision making. The study has highlighted the significance of explainability, interpretability, and gaining insights into the logic or reason behind specific predictions, which help increase the trust of domain experts and stakeholders in the decision support. The development of trust can further result in acceptance and adoption of ML based ID models for resource constrained devices.

6 CONCLUSION

The paper provided a comprehensive overview of state-of-the-art ML, DL, and hybrid techniques with promising results. We also discussed datasets that are frequently used in ID research. Based on this, the present study proposed a novel explainable intrusion detection (XID) approach based on PSO based feature selection and ten different ML classifiers trained and tested on a new dataset in the domain of IoMT. LIME has been applied to support human model prediction interpretability using the classifier with the best performance (ExtraTreeClasssifier). The approach used typical performance metrics to test the validity of the proposed XID methodology. Our proposed hybrid methodology using PSO and ML has resulted in an excellent attack classification accuracy between 97-99%.

LIME can be applied to wide range of models for explainability. However, LIME can be computationally expensive when applied to large datasets. Hence as part of our future work, we plan to test and use other explainable and interpretable AI techniques such as SHAP to provide insights into how the developed models work. We also plan to conduct a systematic literature review to investigate and compare various techniques for AI explainability to provide insights into the strengths and weaknesses of each technique in relation to XID.

Furthermore, we plan to use evolutionary computing techniques such as Genetic Algorithm, Ant Colony Optimization, Grey Wolf Optimization to evaluate the performance of these as compared to the ten developed ML models.

Finally, we plan perform a qualitative reinforcement study (i.e., a Delphi study) involving domain experts and other relevant stakeholders to increase our understanding of how to best provide future XID decision support.

REFERENCES

Aboelwafa, M. M. N., Seddik, K. G., Eldefrawy, M. H., Gadallah, Y., & Gidlund, M. (2020). A Machine-Learning-Based Technique for False Data Injection Attacks Detection in Industrial IoT. *IEEE Internet of Things Journal*, *7*(9). https://doi.org/10.1109/JIOT.2020.2991693

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*. https://doi.org/10.1109/ACCESS.2018.2870052

Adnan, A., Muhammed, A., Ghani, A. A. A., Abdullah, A., & Hakim, F. (2021). An intrusion detection system for the internet of things based on machine learning: Review and challenges. *Symmetry*, *13*(6). https://doi.org/10.3390/sym13061011

Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection. *IEEE Access*, *6*. https://doi.org/10.1109/ACCESS.2018.2841987

Al-Haija, Q. A., & Zein-Sabatto, S. (2020). An efficient deep-learning-based detection and classification system for cyber-attacks in iot communication networks. *Electronics*, *9*(12). https://doi.org/10.3390/ELECTRONICS9122152

Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks. *Journal of Sensors*. https://doi.org/10.1155/2016/4731953

Alqahtani, M., Mathkour, H., & Ben Ismail, M. M. (2020). IoT Botnet Attack Detection Based on Optimized Extreme Gradient Boosting and Feature Selection. *Sensors*, *20*(21). https://doi.org/10.3390/S20216336

Alsabah, M., Naser, M. A., Mahmmod, B. M., Abdulhussain, S. H., Eissa, M. R., Al-Baidhani, A., Noordin, N. K., Sait, S. M., Al-Utaibi, K. A., & Hashim, F. (2021). 6G Wireless Communications Networks: A Comprehensive Survey. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3124812

Anand, A., Rani, S., Anand, D., Aljahdali, H. M., & Kerr, D. (2021). An efficient CNN-based deep learning model to detect malware attacks (CNN-DMA) in 5G-IoT healthcare applications. *Sensors*, *21*(19). https://doi.org/10.3390/s21196346

Asghar, M. Z., Memon, S. A., & Hämäläinen, J. (2022). Evolution of Wireless Communication to 6G: Potential Applications and Research Directions. *Sustainability*, *14*(10). https://doi.org/10.3390/su14106356

Awotunde, J. B., Chakraborty, C., & Adeniyi, A. E. (2021). Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection. *Wireless Comm and Mobile Comp*. https://doi.org/10.1155/2021/ 7154587

Bagaa, M., Taleb, T., Bernabe, J. B., & Skarmeta, A. (2020). A Machine Learning Security Framework for Iot Systems. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.2996214

Belavagi, M. C., & Muniyal, B. (2016). Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection. *Procedia Computer Science*, *89*. https://doi.org/10.1016/j.procs.2016.06.016

Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. *Sensors*, *23*(2). https://doi.org/10.3390/s23020634

Chen, H. Y., & Lee, C. H. (2020). Vibration Signals Analysis by Explainable Artificial Intelligence (XAI) Approach: Application on Bearing Faults Diagnosis. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3006491

Chen, X., & Hao, X. (2019). Feature reduction method for cognition and classification of IoT devices based on artificial intelligence. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2929311

CIC-IDS (2017). https://www.unb.ca/cic/datasets/nsl.html

Derhab, A., Aldweesh, A., Emam, A. Z., & Khan, F. A. (2020). Intrusion Detection System for Internet of Things Based on Temporal Convolution Neural

Network and Efficient Feature Engineering. *Wireless Comm and Mobile Comp*. https://doi.org/10.1155/2020/6689134

Ding, P., Li, J., Wen, M., Wang, L., & Li, H. (2020). Efficient BiSRU combined with feature dimensionality reduction for abnormal traffic detection. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3022355

Duque, S., & Omar, M. N. Bin. (2015). Using Data Mining Algorithms for Developing a Model for Intrusion Detection System (IDS). *Procedia Computer Science*, *61*. https://doi.org/10.1016/j.procs.2015.09.145

Elrawy, M. F., Awad, A. I., & Hamed, H. F. A. (2018). Intrusion detection systems for IoT-based smart environments: a survey. *Journal of Cloud Computing*, *7*(1). https://doi.org/10.1186/s13677-018-0123-6

Farnaaz, N., & Jabbar, M. A. (2016). Random Forest Modeling for Network Intrusion Detection System. *Procedia Computer Science*, *89*. https://doi.org/10.1016/j.procs.2016.06.047

Fatani, A., Dahou, A., Al-qaness, M. A. A., Lu, S., & Abd Elaziz, M. A. (2022). Advanced Feature Extraction and Selection Approach Using Deep Learning and Aquila Optimizer for IoT Intrusion Detection System. *Sensors*, *22*(1). https://doi.org/10.3390/S22010140

Fraile, F., Tagawa, T., Poler, R., & Ortiz, A. (2018). Trustworthy Industrial IoT Gateways for Interoperability Platforms and Ecosystems. *IEEE Internet of Things Journal*, *5*(6). https://doi.org/10.1109/JIOT.2018.2832041

Gad, A. R., Nashat, A. A., & Barkat, T. M. (2021). Intrusion Detection System Using Machine Learning for Vehicular Ad Hoc Networks Based on ToN-IoT Dataset. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3120626

Gao, X., Hu, C., Shan, C., Liu, B., Niu, Z., & Xie, H. (2020). Malware classification for the cloud via semi-supervised transfer learning. *Journal of Information Security and Applications*, *55*. https://doi.org/10.1016/j.jisa.2020.102661

Ghanem, W. A. H. M., Jantan, A., Ghaleb, S. A. A., & Nasser, A. B. (2020). An Efficient Intrusion Detection Model Based on Hybridization of Artificial Bee Colony and Dragonfly Algorithms for Training Multilayer Perceptrons. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3009533

Ghubaish, A., Salman, T., Zolanvari, M., Unal, D., Al-Ali, A., & Jain, R. (2021). Recent Advances in the Internet-of-Medical-Things (IoMT) Systems Security. *IEEE Internet of Things Journal*, *8*(11). https://doi.org/10.1109/JIOT.2020.3045653

Hindy, H., Bayne, E., Bures, M., Atkinson, R., Tachtatzis, C., & Bellekens, X. (2021). Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study (MQTT-IoT-IDS2020 Dataset). *Lecture Notes in Networks and Systems*, *180*. https://doi.org/10.1007/978-3-030-64758-2_6

Houda, Z. A. El, Brik, B., & Khoukhi, L. (2022). “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*. https://doi.org/10.1109/OJCOMS.2022.3188750

Huma, Z. E., Latif, S., Ahmad, J., Idrees, Z., Ibrar, A., Zou, Z., Alqahtani, F., & Baothman, F. (2021). A Hybrid Deep Random Neural Network for Cyberattack Detection in the Industrial Internet of Things. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3071766

IoT-23 (2020). https://www.kaggle.com/datasets/ engraqeel/iot23preprocesseddata

Imrana, Y., Xiang, Y., Ali, L., & Abdul-Rauf, Z. (2021). A bidirectional LSTM deep learning approach for intrusion detection. *Expert Systems with Applications*, *185*. https://doi.org/10.1016/j.eswa.2021.115524

ISCXIDS (2012). https://www.unb.ca/cic/datasets/nsl.html

Jabbar, M. A., Aluvalu, R., & Reddy, S. S. (2017). RFAODE: A Novel Ensemble Intrusion Detection System. *Procedia Computer Science*, *115*, 226–234. https://doi.org/10.1016/j.procs.2017.09.129

Jan, S. U., Ahmed, S., Shakhov, V., & Koo, I. (2019). Toward a Lightweight Intrusion Detection System for the Internet of Things. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2907965

Jiang, H., He, Z., Ye, G., & Zhang, H. (2020). Network Intrusion Detection Based on PSO-Xgboost Model. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.2982418

Kaluarachchi, T., Reis, A., & Nanayakkara, S. (2021). A review of recent deep learning approaches in human-centered machine learning. *Sensors*, *21*(7). https://doi.org/10.3390/s21072514

Kasongo, S. M. (2021). An advanced intrusion detection system for IIoT Based on GA and tree based algorithms. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3104113

Kasongo, S. M., & Sun, Y. (2019). A deep learning method with filter based feature engineering for wireless intrusion detection system. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2905633

Khan, M. A. M. A., Khan, M. A. M. A., Jan, S. U., Ahmad, J., Jamal, S. S., Shah, A. A., Pitropakis, N., & Buchanan, W. J. (2021). A deep learning-based intrusion detection system for mqtt enabled iot. *Sensors*, *21*(21). https://doi.org/10.3390/s21217016

Kirsch, J. H. and D. (2018). Machine Learning For Dummies®, IBM Limited Edition Published. In *Journal of the American Society for Information Science*, 35, 5. https://doi.org/10.1002/asi.4630350509

Kumar, P., Gupta, G. P., & Tripathi, R. (2021). Toward Design of an Intelligent Cyber Attack Detection System using Hybrid Feature Reduced Approach for IoT Networks. *Arab Journal for Science and Eng*, *46*(4). https://doi.org/10.1007/S13369-020-05181-3

Kyoto (2006), https://www.impactcybertrust.org/dataset_view?idDataset=918

Latif, S., Huma, Z. e, Jamal, S. S., Ahmed, F., Ahmad, J., Zahid, A., Dashtipour, K., Umar Aftab, M., Ahmad, M., & Abbasi, Q. H. (2021). Intrusion Detection Framework for the Internet of Things using a Dense

Random Neural Network. *IEEE Trans on Industrial Informatics*. https://doi.org/10.1109/tii.2021.3130248

Man, D., Zeng, F., Yang, W., Yu, M., Lv, J., & Wang, Y. (2021). Intelligent Intrusion Detection Based on Federated Learning for Edge-Assisted Internet of Things. *Security and Communication Networks*, *2021*. https://doi.org/10.1155/2021/9361348

Mendonca, R. V., Teodoro, A. A. M., Rosa, R. L., Saadi, M., Melgarejo, D. C., Nardelli, P. H. J., & Rodriguez, D. Z. (2021). Intrusion Detection System Based on Fast Hierarchical Deep Convolutional Neural Network. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3074664

Moustafa, N. (2021). A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets. *Sustainable Cities and Society*, *72*. https://doi.org/10.1016/j.scs.2021.102994

MQTT-IOT (2020). https://ieee-dataport.org/open-access/mqtt-iot-ids2020-mqtt-internet-things-intrusion-detection-dataset

Mukherjee, S., & Sharma, N. (2012). Intrusion Detection using Naive Bayes Classifier with Feature Reduction. *Procedia Technology*, *4*. https://doi.org/10.1016/j.protcy.2012.05.017

Naseer, S., Saleem, Y., Khalid, S., Bashir, M. K., Han, J., Iqbal, M. M., & Han, K. (2018). Enhanced network anomaly detection based on deep neural networks. *IEEE Access*, *6*. https://doi.org/10.1109/ACCESS.2018.2863036

N-BaIoT (2018). https://www.kaggle.com/datasets/mkashifn/nbaiot-dataset

Panda, M., Mousa, A. A. A., & Hassanien, A. E. (2021). Developing an Efficient Feature Engineering and Machine Learning Model for Detecting IoT-Botnet Cyber Attacks. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3092054

Peng, K., Leung, V. C. M., Zheng, L., Wang, S., Huang, C., & Lin, T. (2018). Intrusion detection system based on decision tree over big data in fog environment. *Wireless Comm and Mobile Comp*. https://doi.org/10.1155/2018/4680867

Popoola, S. I., Adebisi, B., Ande, R., Hammoudeh, M., Anoh, K., & Atayero, A. A. (2021). Smote-drnn: A deep learning algorithm for botnet detection in the internet-of-things networks. *Sensors*, *21*(9). https://doi.org/10.3390/s21092985

Priya, V., Thaseen, I. S., Gadekallu, T. R., Aboudaif, M. K., & Nasr, E. A. (2020). Robust attack detection approach for iiot using ensemble classifier. *Comp, Mat and Cont*, *66*(3). https://doi.org/10.32604/cmc.2021.013852

Qureshi, A. U. H., Larijani, H., Mtetwa, N., Javed, A., & Ahmad, J. (2019). RNN-ABC: A new swarm optimization based technique for anomaly detection. *Computers*, *8*(3). https://doi.org/10.3390/computers8030059

Ribeiro, M. T., & Guestrin, C. (2016). *" Why Should I Trust You ?" Explaining the Predictions of Any Classifier*.

Ruzafa-Alcazar, P., Fernandez-Saura, P., Marmol-Campos, E., Gonzalez-Vidal, A., Hernandez Ramos, J. L., Bernal, J., & Skarmeta, A. F. (2021). Intrusion Detection based on Privacy-preserving Federated Learning for the Industrial IoT. *IEEE Trans on Industrial Informatics*, 1–10. https://doi.org/10.1109/TII.2021.3126728

Sadaf, K., & Sultana, J. (2020). Intrusion detection based on autoencoder and isolation forest in fog computing. *IEEE Access*, *8*. https://doi.org/10.1109/ ACCESS. 2020.3022855

Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. K. A. A. (2020). Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review. *Procedia Computer Science*, *171*. https://doi.org/10.1016/j.procs.2020.04.133

Siddiqi, M. A., & Pak, W. (2021). An Agile Approach to Identify Single and Hybrid Normalization for Enhancing Machine Learning-Based Network Intrusion Detection. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3118361

Sumaiya Thaseen, I., & Aswani Kumar, C. (2017). Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *Journal of King Saud University - Computer and Information Sciences*, *29*(4). https://doi.org/10.1016/j.jksuci.2015.12.004

Tavallaee, M., Bagheri, E., Lu, W. & Ghorbani A. A., A detailed analysis of the KDD CUP 99 data set, 2009 IEEE Symp on Comp Intel for Security and Defense Applications. doi: 10.1109/CISDA.2009.5356528.

Tian, Q., Li, J., & Liu, H. (2019). A Method for Guaranteeing Wireless Communication Based on a Combination of Deep and Shallow Learning. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS. 2019. 2905754

TS, P., & Shrinivasacharya, P. (2021). Evaluating neural networks using Bi-Directional LSTM for network IDS in cyber security. *Global Transition Proc*, *2*(2). https://doi.org/10.1016/ j.gltp.2021.08.017

Tsang, C. H., & Kwong, S. (2005). Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction. *Proceedings of the IEEE Inter Conf on Ind Tech*. https://doi.org/10.1109/ICIT.2005.1600609

UNSW (2015). https://research.unsw.edu.au/projects/unsw-nb15-dataset

UNSW (2019). https://research.unsw.edu.au/projects/bot-iot-dataset

UNSW (2021). https://research.unsw.edu.au/projects/toniot-datasets

Vargas, H., Lozano-Garzon, C., Montoya, G. A., & Donoso, Y. (2021). Detection of security attacks in industrial iot networks: A blockchain and machine learning approach. *Electronics*, *10*(21). https://doi.org/10.3390/electronics10212662

Vigoya, L., Fernandez, D., Carneiro, V., & Nóvoa, F. J. (2021). IoT dataset validation using machine learning techniques for traffic anomaly detection. *Electronics*, *10*(22). https://doi.org/10.3390/electronics10222857

Wang, W., Sheng, Y., Wang, J., Zeng, X., Ye, X., Huang, Y., & Zhu, M. (2017). HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection. *IEEE*

Access, *6*. https://doi.org/10.1109/ACCESS.2017.2780250

Wu, Y., Lee, W. W., Xu, Z., & Ni, M. (2020). Large-scale and robust intrusion detection model combining improved deep belief network with feature-weighted svm. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.2994947

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. *IEEE Access*, *6*. https://doi.org/10.1109/ACCESS.2018.2836950

Yang, M. S., & Sinaga, K. P. (2019). A feature-reduction multi-view k-means clustering algorithm. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2934179

Yao, H., Gao, P., Zhang, P., Wang, J., Jiang, C., & Lu, L. (2019). Hybrid intrusion detection system for edge-based IIoT relying on machine-learning-aided detection. *IEEE Network*, *33*(5). https://doi.org/10.1109/MNET.001.1800479

Yoo, J., Min, B., Kim, S., Shin, D., & Shin, D. (2021). Study on Network Intrusion Detection Method Using Discrete Pre-Processing Method and Convolution Neural Network. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3120839

Yu, Y., Long, J., & Cai, Z. (2017). Network Intrusion Detection through Stacking Dilated Convolutional Autoencoders. *Security and Communication Networks*. https://doi.org/10.1155/2017/4184196

Zavrak, S., & Iskefiyeli, M. (2020). Anomaly-Based Intrusion Detection from Network Flow Features Using Variational Autoencoder. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3001350.

APPENDIX

Table 3: Overview of publicly available datasets.

Dataset	Threat type	Features	Records	Limitations
NSL-KDD (Tavallaee et al., 2009)	Four malware traffic types i.e., DoS, Probe, U2R, and R2L	41	125,973	Non IoT
WSN-DS (Almomani et al., 2016)	DoS attacks (Blackhole, Grayhole, Flooding, and Scheduling)	23	NA	WSN
ISCX (ISCXIDS, 2012)	BruteForce SSH, DDoS, HttpDoS	82	73566	Non IoT
CICIDS2017 (CIC-IDS, 2017)	Benign, BoT,DDoS, DoS/DDoS, PortScan, SSH	79	786,633	NA
UNSW-NB15 (UNSW, 2015)	9 attack classes	42	162724	Non IoT
Bot-IoT (UNSW, 2019)	7 attack classes	14	74680	Non IoT
TON_IoT (UNSW, 2021)	DoS, DDoS, R2L, U2R, probe attacks	NA	NA	IoT/Imbalance
Kyoto (Kyoto, 2006)	10 Honeypots generated features and 14 adopted from KDDcup99.	24	NA	Non IoT
IoT-23 (IoT-23, 2020)	20 types of attacks including Mirai and Torri etc. 3 types of benign IoT devices traffic	23	NA	IoT traffic
N-BaIoT (N-BaIoT, 2018)	Public botnet IoT dataset.	115	7062606	IoT
MQTT-IoT-IDS2020 (MQTT-IOT, 2020)	Contain simulated five MQTT-based network attack scenarios including the Brute Force attack.	32	16696	IoT

Table 4: Overview of recent intrusion detection studies.

Study	**AI technique(s)**	**Type**	**Dataset(s)**	**Accuracy**
Yao et al., 2019	Light GBM	DL	NSL-KDD and CSE-CIC-IDS 2018	NA
Khan et al., 2021	DNN	DL	MQTT-IoT-IDS2020	NA
Imrana et al., 2021	Bidirectional Long-Short-Term-Memory (BiDLSTM)	DL	NSL-KDD	94.26%
Belavagi and Muniyal 2016)	LR, Gaussian NB, SVM, RF	ML	NSL-KDD	NA
Mukherjee and Sharma, 2012	Naive Bayes	ML	NSL-KDD	NA
Gad et al., 2021	XGBoost	ML	Ton_IoT	NA
Jiang et al., 2020	PSO, XGBoost	ML	NSL-KDD	NA
Ghanem et al., 2020	Hybridization of ABC and Dragonfly algorithm (DA) for training ANN	Hybrid	KDD CUP 1999, NSL-KDD, UNSW- NB15, and ISCX2012	NA
Awotunde et al., 2021	DFFNN	DL	NSL-KDD and UNSW-NB15	NA
TS & Shrinivasacharya, 2021	Bi-directional LSTM	DL	KDDCUP-99 and UNSW-NB15	99%
Sadaf and Sultana, 2020	Autoencoder and IF	DL	NLS-KDD	95.40%
Yoo et al., 2021	CNN	DL	NSL-KDD and CSE-CIC-IDS 2018	NA
Qureshi et al., 2019	RNN-ABC	Hybrid	NLS-KDD	91.65%
Huma et al., 2021	DRaNN and MLP	DL	DS2OS and UNSW-NB15	98%
Anand et al., 2021	CNN	DL	Malimg	99%
Adnan et al., 2021	Neural Network (NN)	DL	KDD99, NSLand Kyoto	NA
Naseer et al., 2018	Autoencoder and CNN	DL	NSLKDD	NA
Latif et al., 2021	DnRaNN	DL	ToN_IoT	99.14%
Popoola et al., 2021	SMOTE-DRNN	DL	Bot-IoT	NA
Aboelwafa et al., 2020	Autoencoder	DL	NA	NA
Ruzafa-Alcazar et al., 2021	FL	FL	ToN IoT	NA
Kasongo, 2021	GA-RF	Hybrid	UNSW-NB15	87.61%
Fatani et al., 2022	CNN, SI	Hybrid	CIC2017, NSL-KDD, BoT-IoT, and KDD99	NA
Jabbar et al., 2017	RF and AODE. Average One-Dependence Estimator (AODE) based	ML	Kyoto	90.51%
Duque and Omar, 2015	K-means clustering	ML	NSL-KDD	NA
Saranya et al., 2020	Linear Discriminant Analysis (LDA), Classification and Regression Trees compared along with the latest research. (CART) and Random Forest.	ML	KDD-CUP	NA
Priya et al., 2020	Multi class SVM	ML	NSL-KDD	NA
Tian et al., 2019	Auto encoder and SVM	ML/DL	NA	NA
Mendonca et al., 2021	Tree-CNN	ML/DL	CICIDS2017	98%
Ribeiro and Guestrin, 2016	Lime explanation technique	XAI	NA	NA
Adadi and Berrada, 2018	DNN with Lime and Shap	XAI	NSL-KDD, UNSW-NB15	NA
Chen and Lee, 2020	Random Forest with Lime interpretations	XAI	NA	100%