IoTGeM: Generalizable Models for Behaviour-Based IoT Attack Detection

Kahraman Kostas, Mike Just, and Michael A. Lones, *Senior Member, IEEE*

***Abstract*— Previous research on behaviour-based attack detection on networks of IoT devices has resulted in machine learning models whose ability to adapt to unseen data is limited, and often not demonstrated. In this paper we present an approach for modeling IoT network attacks that focuses on generalizability, yet also leads to better detection and performance. First, we present an improved rolling window approach for feature extraction, and introduce a multi-step feature selection process that reduces overfitting. Second, we build and test models using isolated train and test datasets, thereby avoiding common data leaks that have limited the generalizability of previous models. Third, we rigorously evaluate our methodology using a diverse portfolio of machine learning models, evaluation metrics and datasets. Finally, we build confidence in the models by using explainable AI techniques, allowing us to identify the features that underlie accurate detection of attacks.**

***Index Terms*—Internet of Things, Intrusion detection, Machine learning, Attack detection, SHAP.**

I. Introduction

The ever-growing number of IoT (internet of things) devices creates an ever-larger attack surface. Reports suggest that a new IoT device included in a network receives its first attack within five hours and becomes the target of a specific attack within 24 hours [1]. IoT devices can be particularly challenging to secure as they can vary widely in terms of their hardware, software, and interfaces, and they typically have more limited resources compared to conventional computing devices. Thus, classical security solutions often need to be tailored for IoT devices [2].

Behaviour-based attack detection using machine learning (ML) is a common approach to securing IoT devices by detecting anomalous network attacks. However, common mistakes in ML studies can raise serious doubts about the reliability of results [3]–[5], with examples such as data leakage and feature overfitting in previous IoT attack detection studies (see Section II). In this paper we focus on automating the behaviour-based discrimination of benign and malicious network data with ML algorithms while avoiding such pitfalls. To do this, we introduce an approach (*IoTGeM*) for creating generalizable models for behaviour-based network attack detection for IoT devices, with the following contributions:

1) A rolling window method for feature extraction that achieves earlier detection and improved performance compared to conventional approaches.
2) A multi-step feature selection process based on a genetic algorithm with exogenous feedback, which is effective in eliminating features that can lead to overfitting.
3) A detailed examination of feature effectiveness by analysing the relationship between features and attacks using explainable AI techniques.

We further adopt a strict methodology for ensuring the generalizability of our models, and the ease of validating, replicating, or extending our approach [4]:

- We build and test our models using isolated train and test datasets to avoid common data leaks.
- We rigorously evaluate our methodology with a diverse portfolio of machine learning models, evaluation metrics and datasets, and we avoid using inappropriate metrics for certain situations, e.g., erroneously using only accuracy with unbalanced data sets.
- We use publicly available datasets and make our scripts available to the public[1].

This article is organized as follows. Section II gives an overview of related work. Section III describes the materials and methods we use for data selection, feature extraction and feature selection. Section IV evaluates the results of the models, compares them with other methods, and provides a detailed summary analysis relating feature effectiveness to different attacks. Limitations are discussed in Section V, and conclusions are given in Section VI.

II. Related Work

Below we review related literature from the past several years that focuses on anomaly or intrusion detection for IoT devices using supervised ML models. We categorise this research into four areas, based on the data, features, ML models and evaluation metrics used in each study. Fig. 1 summarises the changes in these areas over time and frames our discussion in each corresponding subsection. Supplementary Materials (SM)-Table V, offers a more extensive compilation of literature on ML approaches in IoT-based attack detection.

A. Data

The use of datasets in related work published from 2019 is summarised in Fig. 1a. Although a number of studies have been published in recent years, it is common practice to use old datasets. For example [6]–[13] used the KDD99[2] and NSL-KDD datasets. KDD99 dates from 1999, and NSL-KDD is an error-corrected version of KDD99. These are large, trusted datasets, and their use in a large number of studies makes them a useful benchmark. However, their age means they lack up-to-date attacks and modern devices. This seriously limits their applicability within contemporary studies of network security.

Furthermore, these datasets do not contain any data related to IoT devices. In fact, it is common practice for IoT-oriented

[1]Materials available at: github.com/kahramankostas/IoTGeM

[2]See Table IV (Appendix) for listing of open-sources datasets.

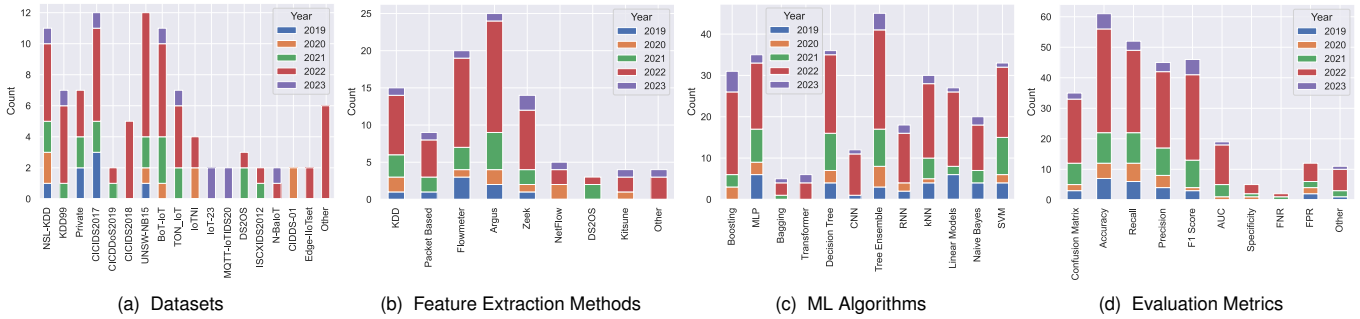


Fig. 1. Usage of datasets, machine learning (ML) algorithms, feature types, and evaluation metrics in the literature.

studies of network security to use datasets which do not contain IoT data. For example, [14]–[27] all use recent datasets, but these datasets do not contain any IoT data. Consequently, it is unclear how these studies yield results that are specific to IoT-based networks or attacks.

It is also common for studies to use private datasets. Although the datasets used in [28]–[34] do contain IoT devices, as far as we are aware they are not publicly accessible, and consequently it is difficult to confirm their data quality. Whilst it is understandable that some datasets cannot be shared, reproducibility is a serious problem in ML-based studies [3], and it is therefore important that studies which use private data are externally validated on a publicly-accessible dataset.

Another issue with datasets is that many (such as KDD99, NSL-KDD, CIDDS-01, DS2OS, MQTTset, N-BaIoT, and InSDN) contain only pre-extracted features, and do not make the raw network data (pcap -packet capture- file) available. Without raw data, it is not possible to investigate new features or extract features using different tools.

On the other hand, many datasets with IoT devices with labelled and raw data for intrusion detection, both simulation-based datasets such as BoT-IoT [35], Edge-IIoT [36], TON-IoT [37], MQTT-IoT-IDS2020 [38], and real device-based dataset such as CIC-IoT-2022 [39], IoT-ENV [40], IoT-NID [41], Kitsune [42], IoT-23 [43] are widely used in the literature. These datasets are analysed in Section III-B.

B. Features

The use of different feature extraction methods in related work published from 2019 is summarised in Fig. 1b. Most of the datasets used for intrusion detection are obtained with common methods and tools. Zeek (formerly Bro) [44], Argus [45], and CICFlowMeter (formerly ISCXFlowMeter) [46] are examples of commonly used tools. For example, the UNSW-NB15 dataset used in [22]–[27] was generated using Zeek and Argus and contains 49 features. ToN_IoT and Bot-IoT datasets were generated using Argus and contain 42 and 29 features respectively. The CICIDS2017-18-19 datasets were all generated by CICFlowMeter and contain 83 features. The common characteristics of these tools are that all of the generated features are flow-based. Unlike these tools, Kitsune, used in [47]–[50] uses a sliding window approach, converting pcap files into a dataset with 115 features.

In the use of packet-based features, individual features obtained from network packets are used. This approach has been used in various studies [31]–[34], [36], [51]. We can also list the following datasets that use individual packet features: MQTTset, Edge-IIoTset and MQTT-IoT-IDS20 (also contains flow-based). Apart from this, some studies [52], [53] detect attacks by training deep learning models on raw network data without first converting it into features. However, since the raw data (pcap file) consists of network packets, these studies face similar challenges to those which use individual packet features, such as the unintentional use of identifying features. Identifying features are those that directly or indirectly give clues about the label information, but are not generalizable. For example, in most cases, IP addresses are identifying because they uniquely identify the attacker or target. However, since these IP attributes may change in the event of another attack, this attribute is unique to that particular dataset and cannot be generalized. Therefore, using this feature will cause an overfitting problem, tying the selected features to one dataset. An analysis of the characteristics, advantages and disadvantages of the flow, window, and packet methods can be found in Section III-D.

It is also possible to create new feature sets by using different tools if the dataset includes a pcap file. For example, the IoTID20 [54] dataset was created using CICFlowMeter from IoT-NID pcap files. So, both flow [54] and sliding window [50] features were extracted from the IoT-NID dataset. On the other hand, KDD99, NSL-KDD, CIDDS-01, DS2OS, MQTTset, N-BaIoT and InSDN datasets do not include pcap files. In this respect, feature extraction and modification studies that can be done with these datasets are quite limited.

C. Machine Learning and Evaluation Metrics

The use of different ML algorithms and metrics in related work published from 2019 is summarised in Fig. 1c–1d. The most widely used algorithms in the literature are tree ensembles (especially random forest), decision trees, MLP, SVM, boosting and kNN. In addition to classical ML algorithms, the use of deep learning methods has been increasing recently. For example, the use of CNN and RNN has increased, and transformers have started to be used intensively since 2022. Researchers have often evaluated their approach using more than one ML model. In many studies with multiple comparisons, ensemble models (such as RF and XGB) gave

the best results [11]–[13], [16], [19]–[22], [25]–[29], [48]–[50], [55], [56].

Whilst deep learning (such as CNN, RNN and transformers) has become popular in recent years, classical ML algorithms remain widely used. In part, this can be attributed to the good performance of classical approaches in cases where feature extraction has already taken place [57], something that is true of most datasets and studies in the area of network security.

Accuracy is the most commonly used evaluation metric. In the majority of studies, accuracy is supported by other metrics such as recall, precision, and F1 score, but there are also studies [7], [58]–[60] where only accuracy is reported. However, in an area such as attack detection, which suffers from unbalanced data distribution, accuracy alone is not a reliable metric [61]. Although the accuracies reported vary from 0.77-1, many studies report a success rate of greater than 0.99 (see Table V). In cases of unbalanced data distribution and where only accuracy is reported, these scores may not accurately reflect model success. More generally, the reliability of results reported in the literature is questionable due to common errors made in network security and ML-based studies (e.g., data leakage, overfitting) [5]. In this context, it is essential that studies are not only focused on high metric scores, but are also free from common errors, transparent and reproducible.

III. Materials and Methods

A. System Model and Design Goals

To build a generalizable ML model that detects attacks we first analyse and decide on the most suitable datasets for our study. Then, we introduce our features by analysing the advantages and disadvantages of feature extraction methods. After feature extraction, we perform feature selection by analysing and eliminating some extracted features by analysing these features under different conditions. From the remaining features, we create the appropriate feature combination using a genetic algorithm (GA) and exogenous feedback. In the next stage, we discover the most suitable hyperparameters for each ML model using hyperparameter optimisation. Finally, we train our models with the selected features and hyperparameters, and then test these models with previously unseen data and obtain our final results. This process is visualised in Fig. 2

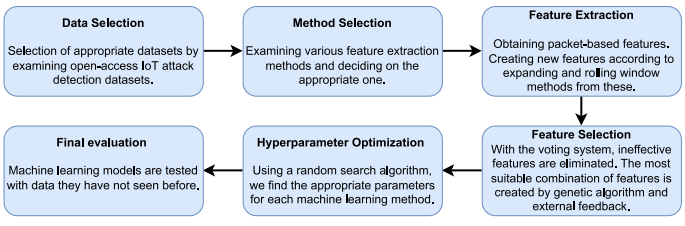


Fig. 2. The visualisation of steps in system implementation.

B. Data Selection

We examined the datasets used in IoT attack detection studies. We searched the ones that are publicly available, contain raw data (pcap files) and are labelled. The list of these is given in Table I, where we also measure them in terms of the number of devices and attack types they contain and classify them in terms of whether they contain real IoT data and multiple sessions. IoT devices differ from non-IoT devices in particular by their heterogeneous nature, in that IoT devices often contain unique software and hardware. This diversity is very difficult to simulate, and consequently it is preferable that datasets include real rather than simulated IoT devices. Repeated attacks in different sessions are also desirable since it allows comparison of each attack across sessions and a deeper analysis of the nature of the attacks. It can also be used to prevent overfitting of session-specific features.

TABLE I
Datasets that can be used for IoT attack detection

Dataset	Year	Real IoT	Session	Devices	Attacks
BoT-IoT [35]	2018	×	×	5	10
CIC-IoT-2022 [39]	2022	✓	✓	40	4
Edge-IIoT [36]	2022	×	×	9	13
IoT-ENV [40]	2021	✓	×	5	3
IoT-NID [41]	2019	✓	✓	2	10
Kitsune [42]	2018	✓	×	9	9
TON-IoT [37]	2019	×	×	7	9
IoT-23 [43]	2020	✓	×	3	8
MQTT-IoT-IDS2020 [38]	2020	×	×	13	4

The IoT-NID and CIC-IoT-22 datasets are the only ones which contain real, multi-session IoT data. While CIC-IoT-22 has a large number of devices, it contains relatively few attack types. Furthermore, attack situations are isolated in this dataset, in that only one device is involved in each session in which an attack occurs. IoT-NID, on the other hand, contains few devices but many attacks, which makes it particularly useful for training different attack detection models. Thus, IoT-NID is used as the main dataset for training models.

It is important to completely isolate training and testing data. To ensure this, we always measure the success of a model on data that the model has never seen before. Fig. 3 depicts how data is used in this study. The same colour occurring in different columns for the same attack represents different sessions from the IoT-NID dataset, i.e. where an attack was carried out more than once, in different sessions. The data in the first column are used for training the models. Data in the Train/CV, HPO and Validation columns are used to refine and select models, while data in the Session and Dataset Test columns are used to measure model generality. More specifically, data in the HPO column is used both for feature reduction and hyperparameter optimisation, and data in the validation column are used in the feature selection step.

Data is quite limited for some of the attacks. In the case of HTTPS and ACK flood attacks, there was insufficient data to use a separate IoT-NID session for validation, hence we used extra data from kb.mazebolt.com for these attacks as their structure of these two attacks closely resembled their versions in IoT-NID. For BF, we addressed the paucity of data for specific attacks by using slightly different variants of the attack in the training, validation and test datasets. Specifically, the training data involves telnet BF, the validation data involves password attacks and the test data involves a RTSP BF attack. For SHD and MHD, there were no suitable equivalent data in

other datasets. So that we could still measure the generality of these attack models, we used an MHD session to test SHD models, and vice versa. This is feasible as SHD and MHD are very similar variants of the same attack (host discovery) made using different devices and tools (scan and Mirai).

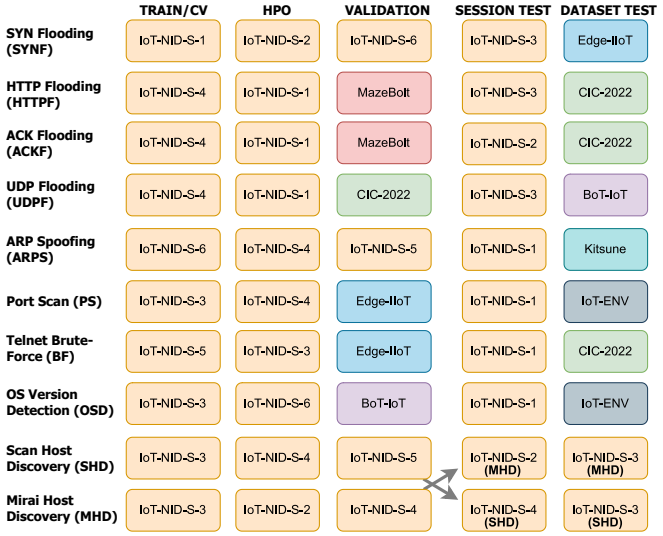


Fig. 3. Use of data in this study. Train/CV=Training and cross-validation. HPO=Feature reduction and hyperparameter selection. Validation=Data validation and feature selection. S-n= Session n.

C. ML Algorithm Selection

Since there is no one-size-fits-all approach for every problem, it is important to consider performance across a range of ML models [4]. We selected one model from each of the following model categories [62]; linear model: logistic regression (LR), tree-based: decision tree (DT), Bayesian-based: naïve Bayes (NB), kernel-based: support vector machine (SVM), ensemble-based: random forest (RF), instance-based: k-nearest neighbours (KNN), and artificial neural networks (ANN): multilayer perceptron (MLP). This set of models provides a good intersection with those used in previous studies. Additionally, we also included XGB, which, although infrequently encountered in the attack detection literature, is known to work well on tabular datasets [63].

We have made the code for our work public[1] and we have used a modular code structure so that only minor code changes are required to support other algorithms.

D. Feature Extraction Methods

In the literature, it is common to see flow-based or individual packet-based approaches to feature extraction. In this section, we describe these two approaches and propose a window-based approach as an alternative.

1) Individual Packet-Based Features: Individual packet features can be obtained from the packet headers or packet payloads of the individual network packets. As discussed in Section II-B, whilst this approach is commonly used [32]–[34], [36], [51], [64], it has a number of issues that can impair generalizability. For this reason, we do not use it in our main study. However, given its prevalence in the literature, in Appendix B we further discuss the limitations of this approach and experimentally show how even simple features such as packet size or timestamp can become identifiers and lead to models with apparently high accuracy but very poor generality. For this, individual features were extracted from raw data using Python, Scapy, and Wireshark tools, mostly from headers, but we also extracted payload-based features such as payload-entropy or payload-bytes.

2) Flow-Based Features: Unlike features from individual packets, flow features [6]–[27] are sufficient for attack detection, but they have some drawbacks. For example, in order to generate features in a flow system, the flow must end with a termination flag or timeout. In this respect, in order for a flow to be identified as an attack, it is necessary to wait for it to be terminated first. Only after the flow ends can the statistics required for attack detection be calculated. In this study, flow-based features are extracted from the raw data using the CICFlowMeter tool. We chose CICFlowMeter for this, since it extracts significantly more features than alternative tools, and also specifically focuses on features that can be used in attack detection. Note that the IoTID20 dataset also contains features extracted using CICFlowMeter from IoT-NID; however, our preliminary analysis of IoTID20 (see Appendix C) identified anomalies, prompting us to do our own feature extraction.

3) Window-Based Features: In this approach, we focus on the changes in the data flow between the source and destination. However, in contrast to the flow-based approach, instead of generating an aggregate statistic for each flow, we utilise the changes in the network data that occur as each packet arrives. While similar approaches have been used in the literature [48]–[50], we present a different approach in terms of both methodology and features. This uses both rolling (RW) and expanding windows (EW) that extract features from information carried by packets between the source and destination (MAC/IP address). In the rolling window approach, we observe the change between packets within a certain window size. The expanding window starts with a small window covering the initial data and gradually expands to include new data. Each time a new packet is added, the window becomes larger and statistics are recalculated. We use four different sources: size, time, destination-source and Transmission Control Protocol (TCP) flags, and also calculate the mean and standard deviation of the window values for some individual features. The process is visualised in Fig. 4.

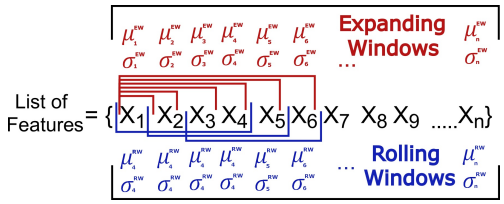


Fig. 4. Window-based feature extraction. μ is mean, σ is std. deviation.

When using the rolling window method, we also had to decide on the window size (number of packets) we would use. To avoid overfitting, we determined the window size using

two MitM (Video Injection and Active Wiretap) attacks that we did not use in our other experiments. We chose these attacks in particular because MitM attacks are sensitive to packet time changes rather than flags, packet content, and destination-source features. Therefore, statistical features such as window features play an important role in detecting them. As a preliminary study, we tried to detect these attacks using only window features, with window sizes ranging from two to 20, using a combination of EW features and EW-RW features for *packet size, time, TCP window size, payload bytes, payload entropy* features. The results are shown in Fig. 5. Considering the limited data generation of many IoT devices, it is impractical to employ large window sizes. Consequently, we limit the RW size to 10, and in future experiments only employ window sizes below this which were most effective at detecting the MitM attacks: specifically two, six, and nine packets. We apply the expanding window approach to TCP flags and source-destination elements. Our full feature list is given in the Appendix, Table VI, including an explanation of each feature type mentioned in this paper.

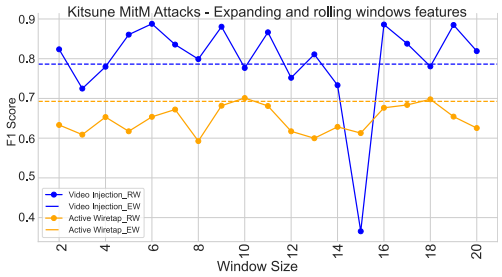


Fig. 5. Utilizing diverse window sizes for Kitsune MitM attack detection and the change of performance according to the window size.

In Section IV-A, we compare our window-based approach against a flow-based implementation in terms of its ability to support attack detection and generalizability.

E. Feature Selection

It is important to identify stable features that are capable of supporting generalizable models. We do this using a three-step method. In the first step, we remove features that are clearly inappropriate, i.e. device or session-based identifiers such as IP-checksum, ID, synchronization, and port numbers (see Appendix B). In the second step, we employ a feature elimination process that considers the value of each feature in turn, and in the third step we use a genetic algorithm (GA) to further refine the remaining features, taking into account their potential interactions.

1) Feature elimination: The aim of this stage is to identify and eliminate features that are a risk to generalizability. For three different validation scenarios, we consider each feature in turn, and measure its association with the target variable when used to train an ML model which is then evaluated using Cohen's kappa. We use an ML model (extremely randomized tree, or ExtraTree) which differs from those listed in Section III-C to avoid bias. A kappa value less than or equal to zero indicates an unreliable feature.

In the first scenario, cross-validation is used within a single session (Train/CV column in Fig. 3). In the second scenario, two independent sessions from the same dataset are used, one (Train/CV column) for training, the other (HPO column) for validation. This helps to eliminate features that do not generalize beyond a single session, a common problem when using cross-validation alone [65]. The third scenario uses a validation session from a different dataset (Validation column).

Features are assessed in each scenario, and if a feature surpasses a kappa score of 0, it is awarded one vote for that scenario. Features with at least two votes, including one from the final scenario, are included in the feature set for the GA. This approach prioritizes features selected in the last, most stringent, scenario whilst also taking into account their utility in the other scenarios to discourage spurious correlations. Fig. 6 depicts the voting process for a particular attack.

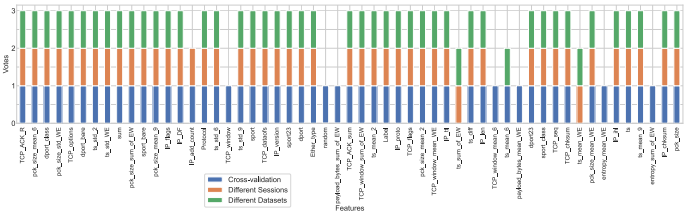


Fig. 6. Voting process during the feature elimination step for the Host Discovery attack.

2) Feature selection with GA and external feedback: We employ a GA to further optimise the feature list, taking into account the potential interactions between features. GAs are a popular approach to feature selection, since they allow the space of feature combinations to be explored in a relatively efficient manner, with a good chance of finding an optimal, or at least near-optimal, combination of features. The GA process is depicted in Fig. 7. The validation results, in terms of F1 score, are used as feedback for the GA, guiding the creation of new feature combinations. This iterative process continues for a specified number of generations (25), with the best combination of features observed during this period chosen as the final feature set. In this procedure, the GA obtains an F1 score as external performance feedback from a distinct dataset (see Fig. 3). Notably, this approach diverges from typical GA approaches to feature selection by employing disparate data for fitness evaluation. This external feedback fosters the selection of feature combinations exemplifying generalizability, thereby addressing overfitting concerns.

IV. Performance Evaluation

In this section, we examine the performance of attack detection models trained with the feature sets we have created. Table II presents the results for each attack when using window-based features. In each case, F1 scores are shown for when the model is evaluated using (a) cross-validation, (b) an independent session from the same dataset it was trained on, and (c) a dataset different to the one it was trained on. This is intended to give insight into each model's generality, with each subsequent evaluation scenario being a more stringent test of the model's ability to generalize to unseen (and therefore

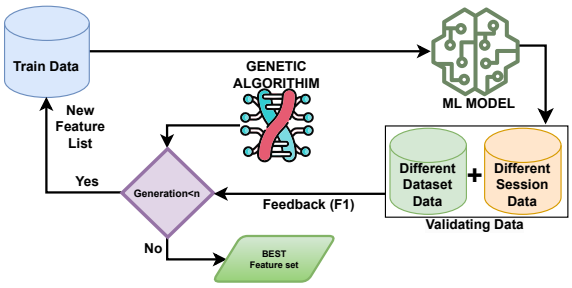


Fig. 7. Visualization of feature selection with GA. The GA outputs the best feature set from n generations with the feedback it receives from the validation.

more realistic) conditions. F1 scores are presented in order to mask the effect of unbalanced data distributions. Additional evaluation measures, including accuracy, recall, and precision, can be found in Appendix D.

For cross-validation, near-perfect discrimination is achieved for all attacks using at least one ML model. When an isolated session is used for evaluation, high levels of discrimination are observed for all but the ARPS attack. When a different dataset is used for evaluation, high levels of discrimination are achieved for all but the ARPS, BF and OSD attacks. It is promising that the majority of attacks are still detected with good levels of discrimination even in the most realistic evaluation condition. However, the success rate clearly drops as the evaluation conditions become more stringent, and this highlights the fact that cross-validation—an approach widely used in the literature—can lead to overly-confident measures of generality, even when steps are taken to remove features that cause overfitting in the feature selection stage.

Notably, the two decision tree ensemble approaches, XGB and RF, lead to the most successful models, with only a few exceptions. For BF and UDPF attacks, NB demonstrated the highest success, while LR performed best for OSD attacks.

TABLE II
F1 SCORES OF ATTACK DETECTION MODELS WHEN EVALUATED USING CROSS-VALIDATION, ISOLATED SESSIONS AND ISOLATED DATASETS.

Name	ML	CV	Session Test	Dataset Test
ACK	DT	1.000	0.999	0.912
	KNN	1.000	1.000	0.212
	LR	1.000	1.000	0.986
	MLP	1.000	0.999	0.333
	NB	1.000	1.000	0.212
	RF	1.000	0.999	0.989
	SVM	0.694	0.235	0.333
	XGB	1.000	1.000	0.989
ARP	DT	0.994	0.583	0.454
	KNN	0.925	0.607	0.341
	LR	0.808	0.631	0.567
	MLP	0.642	0.563	0.373
	NB	0.494	0.411	0.329
	RF	0.997	0.664	0.345
	SVM	0.447	0.432	0.333
	XGB	0.997	0.733	0.349
BF	DT	0.999	0.481	0.351
	KNN	0.979	0.925	0.520
	LR	0.996	0.516	0.402
	MLP	0.938	0.969	0.486
	NB	0.602	0.601	0.738
	RF	1.000	0.974	0.317
	SVM	0.973	0.971	0.313
	XGB	0.998	0.898	0.313
HTTP	DT	1.000	0.993	1.000
	KNN	0.998	0.893	0.355
	LR	0.994	0.993	0.310
	MLP	0.985	0.930	0.388
	NB	0.632	0.450	0.356
	RF	1.000	0.996	1.000
	SVM	0.490	0.486	0.333
	XGB	1.000	0.996	1.000
MHD	DT	0.994	0.808	0.795
	KNN	0.981	0.992	0.993
	LR	0.984	0.994	0.998
	MLP	0.995	0.994	0.995
	NB	0.562	0.795	0.833
	RF	0.996	0.980	0.985
	SVM	0.557	0.581	0.587
	XGB	0.996	0.793	0.811
OS	DT	0.999	0.890	0.542
	KNN	0.944	0.493	0.457
	LR	0.998	0.885	0.691
	MLP	0.485	0.490	0.332
	NB	0.590	0.247	0.453
	RF	0.975	0.784	0.618
	SVM	0.485	0.499	0.333
	XGB	0.998	0.890	0.551
SHD	DT	0.999	0.890	0.889
	KNN	0.998	0.740	0.756
	LR	0.998	0.611	0.639
	MLP	0.999	0.560	0.781
	NB	0.490	0.687	0.726
	RF	0.999	0.816	0.684
	SVM	1.000	0.497	0.495
	XGB	0.999	0.954	0.912
PS	DT	1.000	1.000	0.896
	KNN	1.000	1.000	0.997
	LR	0.999	1.000	0.997
	MLP	1.000	1.000	0.997
	NB	0.972	0.963	0.971
	RF	0.999	1.000	0.862
	SVM	0.472	0.608	0.461
	XGB	0.999	1.000	0.945
SYN	DT	1.000	0.999	0.930
	KNN	0.999	0.997	0.298
	LR	1.000	0.993	0.981
	MLP	0.999	0.997	0.333
	NB	0.993	0.993	0.998
	RF	1.000	0.999	1.000
	SVM	0.945	0.951	0.333
	XGB	1.000	0.999	1.000
UDP	DT	1.000	0.884	0.359
	KNN	1.000	1.000	0.218
	LR	1.000	1.000	0.218
	MLP	0.999	1.000	0.812
	NB	0.669	0.983	0.941
	RF	1.000	1.000	0.377
	SVM	1.000	1.000	0.556
	XGB	1.000	1.000	0.218

A. Comparison with Flow-Based Features

Next, we compare models trained using our window-based features against those trained using flow-based features, the latter being more typical in the existing attack detection literature. In order to ensure a fair comparison, we use the same raw data and apply the same feature selection steps.

Table III shows the results. For cross-validation, both approaches achieve near-perfect scores in at least one ML model for UDP, PS, and SYN attacks. However, the window-based approach outperforms the flow-based approach, especially for the ARPS, ACK and BF attacks (see Table II).

When evaluated using an independent session, both methods achieve near-perfect scores in at least one model for UDPF, PS, and SYNF attacks. The window-based approach exhibits superior performance for ACKF, BF, and HTTPF attacks, with approximately 3, 16, and 6 percentage point differences, respectively. Scores for the ARPS attack were almost identical between the two methods.

When evaluated using an independent dataset, the flow-based approach fails to detect ACKF, SYNF, and UDPF attacks, while the window-based approach achieves scores above 0.94 for at least one ML model for all three attacks. Both approaches demonstrate similar performance in PS attacks, while the window-based approach outperforms the flow-based approach for ACKF, SYNF, and UDPF attacks.

B. Feature Effectiveness

These results strongly suggest that a window-based approach is beneficial in terms of both model performance and generality. To understand this further, in this section we analyse the trained models using an explainable AI technique, SHapley Additive exPlanations (SHAP) [66], [67], and consider how features are actually used for attack detection.

1) Flood attacks: First we consider the four flood attacks: ACK, HTTP, SYN, and UDP floods.

a) ACK flood (ACKF): For this attack, we see that both types of feature support useful models for the first two evaluation scenarios, but only window-based features support models that generalize to an independent dataset. Fig. 8a shows that the most important feature in the window approach is the *sport_sum*. While this value is low in benign samples, it is usually high in attack data. In the case of ACKF attack, this value gives the idea that the source port diversity is high. We see that *payload_bytes_mean_WE* and *ts_mean_6* features also play an active role due to the relatively small size of the packet flow in a short time in the attack situation. *TCP_ACK_SR* and *TCP_ACK_ratio* are also important, because they give

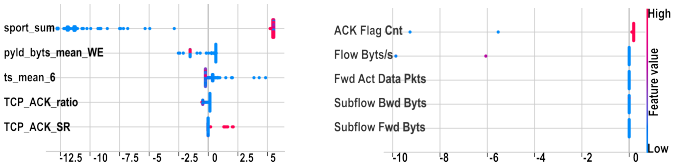
TABLE III
F1 SCORE COMPARISON OF WINDOW AND FLOW-BASED APPROACHES.

Name	ML	CV Window	CV Flow	Session Test Window	Session Test Flow	Dataset Test Window	Dataset Test Flow
ACK	DT	1.000	0.893	0.999	0.966	0.912	0.113
	KNN	1.000	0.370	1.000	0.022	0.212	0.093
	LR	1.000	0.894	1.000	0.966	0.986	0.113
	MLP	1.000	0.895	0.999	0.966	0.333	0.113
	NB	1.000	0.799	1.000	0.886	0.212	0.115
	RF	1.000	0.893	0.999	0.966	0.989	0.113
	SVM	0.694	0.894	0.235	0.966	0.333	0.113
	XGB	1.000	0.893	1.000	0.967	0.989	0.113
ARP	DT	0.994	0.659	0.583	0.698	0.454	0.401
	KNN	0.925	0.636	0.607	0.710	0.341	0.599
	LR	0.808	0.661	0.631	0.674	0.567	0.336
	MLP	0.642	0.641	0.563	0.686	0.373	0.759
	NB	0.494	0.463	0.411	0.428	0.329	0.730
	RF	0.997	0.682	0.664	0.727	0.345	0.421
	SVM	0.447	0.375	0.432	0.424	0.333	0.211
	XGB	0.997	0.681	0.733	0.731	0.349	0.389
BF	DT	0.999	0.851	0.481	0.767	0.351	0.433
	KNN	0.979	0.781	0.925	0.726	0.520	0.359
	LR	0.996	0.456	0.516	0.458	0.402	0.316
	MLP	0.938	0.640	0.969	0.564	0.486	0.443
	NB	0.602	0.486	0.601	0.506	0.738	0.686
	RF	1.000	0.883	0.974	0.810	0.317	0.343
	SVM	0.973	0.474	0.971	0.480	0.313	0.316
	XGB	0.998	0.892	0.898	0.803	0.313	0.503
HTTP	DT	1.000	0.962	0.993	0.933	1.000	0.681
	KNN	0.998	0.890	0.893	0.840	0.355	0.566
	LR	0.994	0.934	0.993	0.919	0.310	0.920
	MLP	0.985	0.761	0.930	0.791	0.388	0.359
	NB	0.632	0.804	0.450	0.747	0.356	0.971
	RF	1.000	0.971	0.996	0.962	1.000	0.596
	SVM	0.490	0.392	0.486	0.369	0.333	0.462
	XGB	1.000	0.968	0.996	0.941	1.000	0.652
PS	DT	1.000	0.995	1.000	0.994	0.896	0.949
	KNN	1.000	0.966	1.000	0.912	0.997	0.696
	LR	0.999	0.990	1.000	0.999	0.997	0.994
	MLP	1.000	0.924	1.000	0.891	0.997	0.685
	NB	0.972	0.959	0.963	0.966	0.971	0.739
	RF	0.999	1.000	1.000	0.998	0.862	0.980
	SVM	0.472	0.516	0.608	0.475	0.461	0.546
	XGB	0.999	1.000	1.000	0.997	0.945	0.980
SYN	DT	1.000	0.977	0.999	0.988	0.930	0.127
	KNN	0.999	0.932	0.997	0.913	0.298	0.367
	LR	1.000	0.960	0.993	0.901	0.981	0.422
	MLP	0.999	0.975	0.997	0.913	0.333	0.208
	NB	0.993	0.746	0.993	0.832	0.998	0.372
	RF	1.000	0.992	0.999	0.992	1.000	0.000
	SVM	0.945	0.499	0.951	0.498	0.333	0.500
	XGB	1.000	0.977	0.999	0.986	1.000	0.110
UDP	DT	1.000	1.000	0.884	1.000	0.359	0.001
	KNN	1.000	0.926	1.000	0.441	0.218	0.001
	LR	1.000	0.975	1.000	1.000	0.218	0.001
	MLP	0.999	0.945	1.000	0.437	0.812	0.001
	NB	0.669	0.628	0.983	0.687	0.941	0.001
	RF	1.000	1.000	1.000	1.000	0.377	0.001
	SVM	1.000	0.493	1.000	0.289	0.556	0.001
	XGB	1.000	0.990	1.000	1.000	0.218	0.001

statistics about the ACK flag. In the flow approach, two prominent features stand out: *Flow Byts/s*[3] and *ACK Flag Cnt*. This is expected as an ACKF attack involves an abundance of ACK flags and relatively smaller data flow in terms of size. However, relying on ACK flag packets or flow size alone may not adequately indicate an attack's presence. This limitation could be why using these features restricts the model's applicability when tested on diverse datasets.

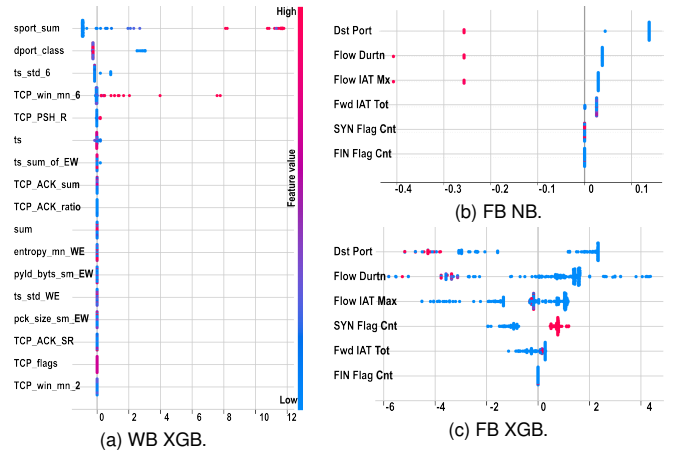
b) HTTP flood (HTTPF): Both feature approaches support the training of discriminative models, though the models build from window-based features are slightly better. Interestingly, ML model type is important here: for window features, decision tree models generalize well to an independent dataset, whereas for flow features, it is the the NB and LR models that generalize. Taking this into account, Fig. 9 shows the SHAP analysis of the XGB and NB models. For window features (Fig. 9a), the two most important features are related to port numbers. The *sport_sum* feature shows a high port diversity in the attack data, while the *dest_port_class* is varied for benign and uniform for attacks. Apart from this, another distinguishing feature is related to TCP-window size. This value appears to be higher in the attack data. For flow features (Fig. 9c), it is similarly observed that the most important characteristic is the destination port. While the normal data shows variation, the port number is stable in the attack data. Another important feature is the flow time, which can be explained by HTTPF attack-generating flows tending to be shorter. When we compare with Fig. 9b, we see that similar features are also prominent. However, while the flag-related feature creates a clear distinction in XGB, it is not noticeable in the NB model. When evaluated on an independent dataset, we believe the main reason for the failure of the flow-based XGB model is that this flag feature causes over-fitting.

[3]Feature list: github.com/ahlashkari/CICFlowMeter/blob/master/ReadMe.txt



(a) Window-Based XGB. (b) Flow-Based XGB.

Fig. 8. Comparison of window and flow-based features for the ACK attack, showing the effect of different features with their importance. Each point in this SHAP plot represents a sample in the dataset. On the y-axis, the attributes are listed in descending order of importance from top to bottom (with a maximum of 20 features shown per plot, otherwise all features used are shown). The x-axis shows the SHAP value: a high value indicates that it contributes to the attack class, while a low value contributes to the benign class. The colour transition indicates whether the value of the feature is high (red) or low (blue).

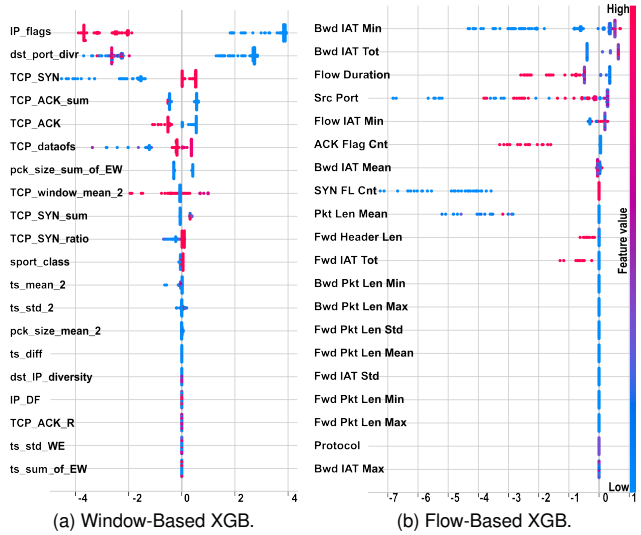


(a) WB XGB. (b) FB NB. (c) FB XGB.

Fig. 9. Comparison of window (WB) and flow (FB) models for HTTP attack.

c) SYN flood (SYNF): For the SYNF attack both approaches are successful in the first two evaluation conditions,

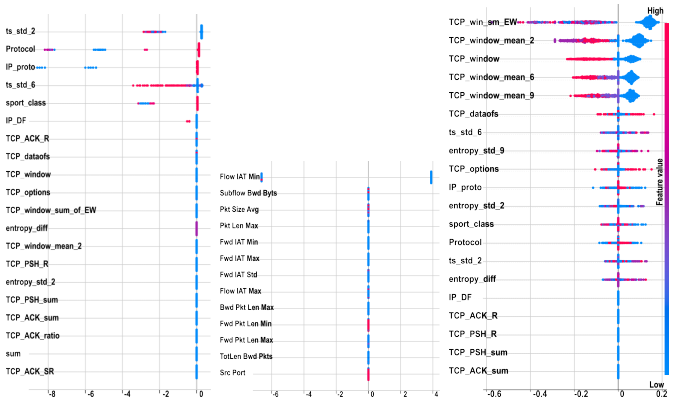
but only the window approach generalizes to an independent dataset. In Fig. 10 the XGB model for both approaches is analysed with the SHAP plot. In Fig. 10a, we can see that the variety of ports provides good discrimination. This is likely due to the fact that this attack targets a specific port. On the other hand, it can be seen that flag statistics features are concentrated. One of the biggest effects of SYN attacks on the network is the anomaly in flags. In this context, it is quite reasonable that features related to SYN and ACK flags are particularly important. While the *IP_flag* feature is not expected to make a significant contribution to the classification between the two groups, it was observed as a distinguishing feature in the model. When Fig. 10b is analysed, we can see that the inter-arrival time (IAT) features stand out and make a significant distinction. Although they have a lower importance score, flow time, port number, flag and size statistics (such as *Flow Duration, Src Port, SYN Flag Cnt, ACK Flag Cnt, Fwd Pkt Len Mean, Fwd Pkt Len Std, Bwd Pkt Len Max* etc.) are important discriminating features. Considering the attack characteristics, we expect significant success by utilizing these features alone. However, the overfitting of the model due to the prominence of IAT features overshadows the contributions of these features. The potential drawbacks associated with IAT features are outlined below. These features can lead to high-dimensional data sets, increasing model complexity and allowing patterns in the noise to cause overfitting. In addition, not all IAT-based features contain the same information, potentially introducing redundancy and irrelevance, further confounding the model. Sparse data distributions can result, with more features than meaningful data instances, leading to poor performance on unseen data. Complex temporal patterns captured by IAT-based features may struggle to generalize effectively, and insufficient data can exacerbate the risk of overfitting by preventing accurate pattern learning and signal/noise discrimination.



(a) Window-Based XGB. (b) Flow-Based XGB.

Fig. 10. Comparison of window (WB) and flow (FB) models for SYN attack.

d) UDP flood (UDPF): Only the NB and MLP models in the window approach achieved significant success when evaluated on an independent dataset. The SHAP plots for these models are presented in Fig. 11. For the flow models, Fig. 11b illustrates the prominence of a solitary feature, denoted as *Flow IAT Min* which pertains to the minimum time interval between packets within a flow. The clear dominance of this particular feature triggers concerns about overfitting which implies an excessive dependence on this attribute, hindering the model's ability to generalize effectively to other datasets. Examining Fig. 11a and Fig. 11c, it can be seen that features that focus on the time between packets cause separation in both models (*ts_std_6, ts_std_2*). However, protocol- and port-centric features (*IP_proto, Protocol, sport_class*) are also prominent. These features are quite reasonable for this attack's detection. Apart from these, the use of TCP-based features in the model is significant. The TCP header has a much more complex structure than the UDP header and therefore contributed more to the feature pool. However, since all attacks in this dataset use the UDP protocol, even the fact that a packet is TCP is a serious discriminator (if it is TCP, it is benign). In an experiment using a dataset with a predominance of UDP-based benign packets, TCP-based characteristics would not be expected to be so dominant.



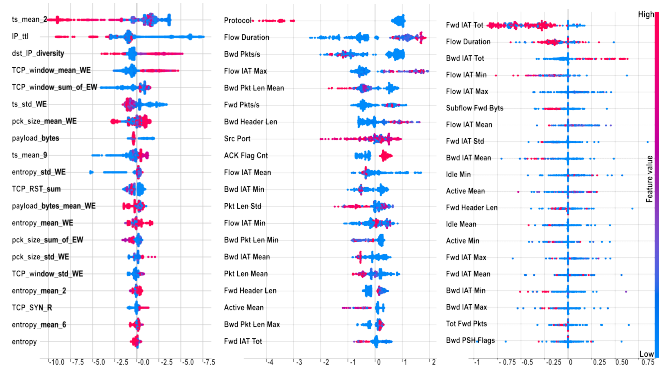
(a) WB MLP. (b) FB XGB. (c) WB NB.

Fig. 11. Comparison of window (WB) and flow (FB) models for UDP attack.

2) ARPS, BF, and PS Attacks: This section examines the feature effectiveness of three other attacks.

a) ARP Spoofing (ARPS): This attack is notable for being the only one where the flow based approach outperforms the window based approach when evaluated on an independent dataset. It is also notable in that the attack packets are very similar to benign packets. The only difference is the time fluctuations in packet transport due to the involvement of the attacker. Therefore, it is very difficult to detect. As a consequence, the detection rate of this attack is generally low. Fig. 12 shows the SHAP analysis of the XGB models, which are the most successful for the first two evaluation conditions, and the flow based MLP models, which generalize best to the independent dataset. When we examine Fig. 12a, we can see that only 3 of the 10 most important features are able to give information about the time between packets. *ts_mean_2* takes the first place, but *ts_std_WE, ts_mean_9* are far behind. *IP_ttl* seems to be a useful feature for distinguishing IP addresses that look the same, rather than being an individual feature. The

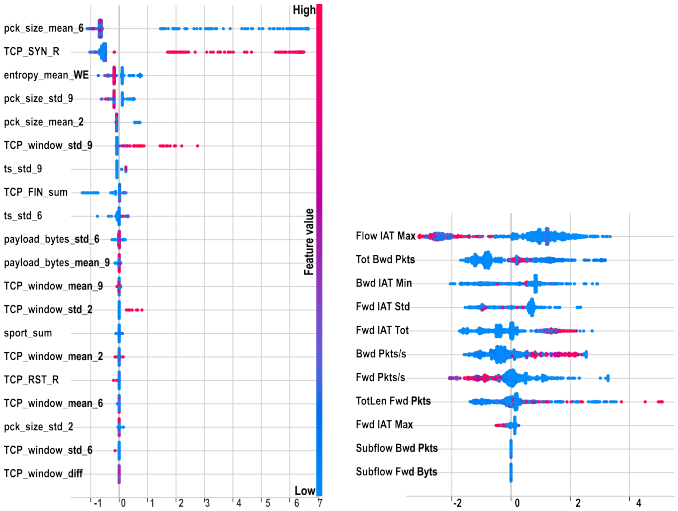
prominence of the *dst_IP_diversity* feature, which provides information about the IP-MAC relationship, is also quite understandable. Even if size, TCP-window or flags related features were highlighted, we do not think they would be useful in detecting this attack. Examination of Fig. 12b shows that protocol, size, and flag based features stand out. We do not consider these features to be of any importance for an ARPS attack. In particular, although the *Protocol* and *ACK Flag Cnt* features provide a clear distinction, we do not consider this to be specific to the dataset and therefore not generalizable. On the other hand, the inter-arrival times (IAT) features (*Fwd IAT Tot, Bwd IAT Mean, Flow IAT Min, Bwd IAT Min, Flow IAT Mean, Flow IAT Max*), which express the time statistics between packets, have the potential to provide discrimination, although they are suppressed by other features. When Fig. 12c is examined, it is seen that with a few exceptions (*Fwd Header Len, Subflow Fwd Byts, Tot Fwd Pkts*), all features are time-dependent, especially IAT based. In this respect, the overfitting problem caused by irrelevant features in the first two models is not observed in this model. Thanks to this characteristic, this model is relatively successful in comparison to the others. The distinguishing characteristic of this attack—setting it apart from benign scenarios—is the temporal fluctuations between packets. Due to the comprehensive integration of inter-packet time features in the flow method, it's clear why this approach is more effective in detecting the attack compared to our method.



(a) WB XGB. (b) FB XGB. (c) FB MLP.

Fig. 12. Comparison of window (WB) and flow (FB) models for ARPS attack.

b) Brute-Force (BF): The window based approach performs best across the three evaluation scenarios. Model type is a significant factor: RF and XGB perform well in the first two scenarios, but only NB achieves reasonable generality on the independent dataset. The telnet BF attack uses the TCP protocol to target specific ports and aims to obtain the username/password by trial and error. In this context, this attack is characterised by a large number of TCP packets targeting specific ports (23,2323 - telnet ports) in a short period of time. The fact that these packets contain passwords and usernames makes the payload-based characteristics important, and the size is expected to be in a certain range. Fig. 13 shows the SHAP analysis of the XGB models. When Fig. 13a is examined, it can be seen that size dependent properties such as *pck_size_mean6, pck_size_mean2, pck_size_std9 and entropy_mean_WE, payload_bytes_std6, payload_bytes_mean9* are used, and these provide information about the payload. In addition, TCP window size-related properties also provide good discrimination; this may be due to the fact that BF tools use a predefined window size [68]. So, it appears that the window-based models are using suitable features, suggesting that the relatively poor performance on the third evaluation scenario is due to the use of RTSP BF attack data for this. Despite both being brute force attacks, there exist notable differences in the tools and approaches employed for each [69], and this may limit the ability of telnet BF models to generalize to RTSP BF attacks. For flow-based models, although features related to the number of outgoing and incoming packets (*Tot Bwd Pkts, Fwd Pkts/s, Bwd Pkts/s*) and packet size features (*TotLen Fwd Pkts*) are used, Fig. 13b indicates that the IAT features (*Fwd IAT std, Bwd IAT Min, Tot Bwd Pkts* etc.) are more prominent, and this may explain their poorer performance.

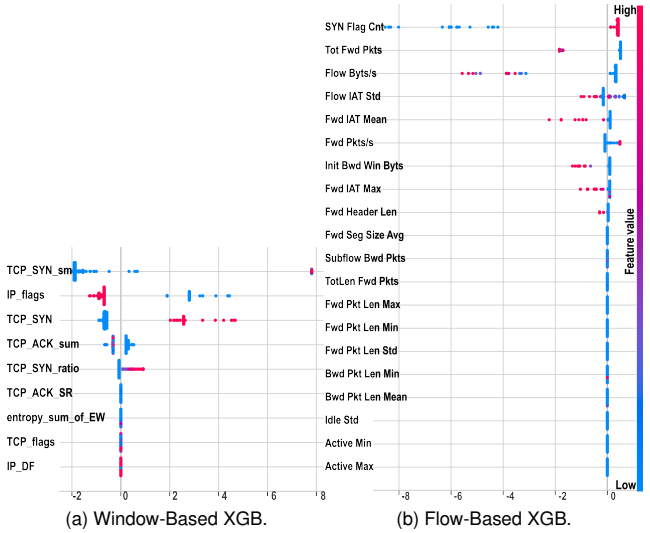


(a) Window-Based XGB. (b) Flow-Based XGB.

Fig. 13. Comparison of window and flow-based models for BF attack.

c) Port Scanning (PS): For this attack, generalizable models could be trained using both window and flow-based features. In the port scanning attack scenario, the attacker sends many TCP packets to the ports with SYN flags set. In this context, flag-based features are important. However, features that provide information about the ports can also be distinctive. When Fig. 14a is analysed, it can be seen that the features related to SYN flag statistics (*TCP_ACK_SR, TCP_SYN_ratio, TCP_ACK_sum, TCP_SYN, TCP_SYN_sum*) stand out in the window approach. In addition, the effect of payload-based properties (*entropy_sum_of_EW*) is also used, presumably since attack packets typically lack payload. Unexpected, however, is the use of an IP flag-based feature. This appears to be due to a spurious correlation in the training data, showing that even a robust approach to feature selection can not always prevent models from picking up on irrelevant features. For the flow models, Fig. 14b shows that the most important feature is the SYN packet count. Features related to packet count and packet size (*Init Bwd Win Byts,Fwd Pkts/s,Flow Byts/s,Tot Fwd Pkts* etc.) are prominent, and IAT features are also used, which seems reasonable for the analysis

of the time between packets, due to heavy packet flow during attacks. However, it is quite interesting that the port-related features are not ranked highly in either SHAP analysis.



(a) Window-Based XGB. (b) Flow-Based XGB.

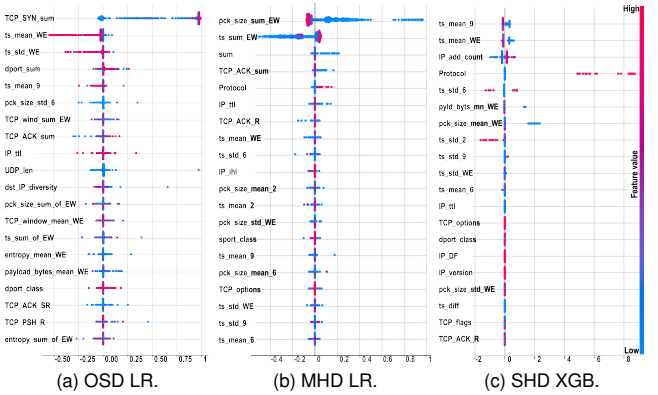
Fig. 14. Comparison of window and flow-based models for the PS attack.

3) Flow-less Attacks: Table III does not include SHD, MHD, and OSD attacks because it is not possible to extract features using CICFlowMeter for these attacks. The destination and source addresses of HD attacks are MAC addresses, and IP-based approaches like the CICFlowMeter cannot generate features for this attack. The OSD attack requires access to individual packet properties such as TCP window size and TCP flags, which is not possible in CICFlowMeter. In this context, one of the advantages of our approach is that it makes it possible to obtain a wider range of layer features and thus potentially detect many more types of attacks. Although it is not possible to compare with the flow-based approach, we have performed feature analysis for these attacks using SHAP. Below, we analyse the most successful models for OSD, MHD, and SHD attacks: LR, LR, and XGB, respectively.

Fig. 15a shows the characteristics given for the Operating System Version Discovery (OSD) attack. Since this attack consists of TCP packets with active SYN flags, TCP-based features are prominent. In addition, depending on the intensity of the attack, time-based features are also important. Even LR, which is the most successful model for detecting the OSD attack in our study, has not achieved significant success. This is likely due to the overfitting that occurred because the training dataset for this attack was very limited (see Table VIII for sample numbers per attack).

For Mirai Host Discovery (MHD) attack, Fig. 15b reveals that the Protocol, IP and TCP-related features are most significant, presumably because this attack does not have IP or TCP headers because it is composed of ARP packets. Considering that all network packets in the ARP protocol have the same size, we can say that time characteristics are effective in distinguishing between benign-attack packets.

Fig. 15c shows the XGB model analysis for the Scanning Host Discovery (SHD) attack. Since SHD is a version of MHD with a different tool, similar characteristics are evident.



(a) OSD LR. (b) MHD LR. (c) SHD XGB.

Fig. 15. SHAP graphs of models for MHD, SHD, OSD attacks.

4) Summary: With the notable exception of ARPS attacks, the window approach is generally ahead in the detection of all attacks. The success of the flow approach is particularly low in cases where data from another dataset is used to evaluate the models. Based on the analysis of these attack models, one reason for this appears to be the extensive use of network-specific statistics in the features of the flow approach. Although models trained with these features may perform well in a specific network environment, the model is unable to sustain this performance when evaluated using data collected from a different network environment. Furthermore, the limitation of flow feature extraction to the IP level restricts adaptability to certain attack types, such as MHD and SHD.

SHAP analyses offer vital insights into model decisions, enhancing our comprehension of how the models perceive attacks. This validation strengthens our confidence in the models' meaningful interpretations of attack patterns, boosting the credibility of our comparisons.

V. Limitations

Our approach is limited by the availability of publicly available datasets, restricting its applicability to attacks covered by these datasets. We focused on popular ML approaches that align with existing research; however, there are numerous other ML approaches that we did not utilize. While testing our method with alternative ML algorithms would be interesting, it is impractical to consider all of them. In terms of efficiency, our method operates on pcap files without timeouts. However, in sliding window features, the feature extraction process may exceed the necessary runtime, particularly in DDoS attacks, as it continues until the window size is complete. Incorporating a timeout in real-time applications would enhance the method's efficiency. Feature extraction was performed using Python and the Scapy library, which are functional but relatively slow. Developing faster versions of our approach using alternative languages like C++ or Go would be beneficial.

VI. Conclusions

Previous research on behaviour-based attack detection on networks of IoT devices tends to produce attack detection models whose ability to adapt to unseen data is limited, and

often not demonstrated. In this paper, we present an approach for IoT network attacks that focuses on generalizability while providing improved detection and performance.

In this context, we have analysed different feature sets, proposed a window-based approach to feature extraction that can provide packet-level attack detection, and compared it with the more conventional flow-based approach. We used a multi-step feature selection method, including a GA, to discover sets of features which are both predictive and generalizable. Since the feature selection method is based on external feedback from an independent dataset, it is very successful in eliminating features that are caused by data leakage and that lead to overfitting, which is a common problem in attack detection. The resulting feature sets were then used by eight different machine learning algorithms to detect ten attack types. The resulting attack detection models were evaluated using three different scenarios, testing their ability to generalize beyond the training distribution. Particularly notable was the superior ability of our window-based models to generalize to previously unseen datasets, when compared to flow-based models. SHAP analysis of the most important features used by models showed that highly specific features such as inter-arrival times (IAT) in the flow characteristics are prominent in many attacks. However, since these features provide information about the dynamics of the network rather than the nature of the attacks, they are far from generalizable and lead to models that suffer from overfitting. On the other hand, we find that the features of our approach are more in line with the nature of the attack, thus producing generalizable models that are successful even for previously unseen attack patterns.

Overall, our results showed success in detecting 10 different isolated attacks. In this context, it achieved $\geq$99% F1 score in ACK, HTTP, SYN, MHD, and PS attacks, 94% in UDP, and 91% in SHD. Although no significant success was achieved in ARPS, OSD and BF attacks, the reasons for the failure of the models were analysed. For ARPS, our feature set may not be well-suited; this is one case where the host of IAT features used by flow-based approaches are more effective. For OSD, the data contains few attacks on which to reliably train a model, and this appears to have led to overfitting features that reflect properties of the benign class. For BF, poor discrimination appears to be due to using a different attack sub-type for testing, rather than being due to deficiency in the learned model.

REFERENCES

[1] H. Modi, "Dawn of the terrorbit era," Feb 2019. Accessed: 2022-04-04,Available at https://www.netscout.com/blog/dawn-terrorbit-era.

[2] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017.

[3] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, 2023.

[4] M. A. Lones, "How to avoid machine learning pitfalls: a guide for academic researchers," *arXiv preprint arXiv:2108.02497*, 2021.

[5] D. Arp, E. Quiring, F. Pendlebury, A. Warnecke, F. Pierazzi, C. Wressnegger, L. Cavallaro, and K. Rieck, "Dos and don'ts of machine learning in computer security," in *USENIX Security 22*, pp. 3971–3988, 2022.

[6] M. Bagaa, T. Taleb, J. B. Bernabe, and A. Skarmeta, "A Machine Learning Security Framework for Iot Systems," *IEEE Access*, vol. 8, pp. 114066–114077, 2020.

[7] A. Raghuvanshi, U. K. Singh, G. S. Sajja, H. Pallathadka, E. Asenso, M. Kamal, A. Singh, and K. Phasinam, "Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming," *Journal of Food Quality*, vol. 2022, pp. 1–8, 2022.

[8] T. Saba, T. Sadad, A. Rehman, Z. Mehmood, and Q. Javaid, "Intrusion detection system through advance machine learning for the internet of things networks," *IT Professional*, vol. 23, no. 2, pp. 58–64, 2021.

[9] J. Manhas and S. Kotwal, *Implementation of Intrusion Detection System for Internet of Things Using Machine Learning Techniques*, pp. 217–237. Singapore: Springer Singapore, 2021.

[10] M. S. Ahmad and S. M. Shah, "Supervised machine learning approaches for attack detection in the IoT network," in *IoT and Its App.: Select Proceedings of ICIA 2020*, pp. 247–260, Springer, 2022.

[11] C. Zhang, D. Jia, L. Wang, W. Wang, F. Liu, and A. Yang, "Comparative research on network intrusion detection methods based on machine learning," *Computers & Security*, p. 102861, 2022.

[12] H. Mliki, A. H. Kaceam, and L. Chaari, "Intrusion Detection Study and Enhancement Using Machine Learning," *Lecture Notes in Computer Science*, vol. 12026 LNCS, pp. 263–278, 2020.

[13] G. El Baltaji, *Anomaly Detection at the Edge implementing Machine Learning Techniques*. PhD thesis, Politecnico di Torino, 2022.

[14] M. Roopak, G. Yun Tian, and J. Chambers, "Deep learning models for cyber security in IoT networks," *IEEE 9th CCWC*, pp. 452–457, 2019.

[15] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, "Toward a Lightweight Intrusion Detection System for the IoT," *IEEE Access*, vol. 7, 2019.

[16] V. Gaur and R. Kumar, "Analysis of machine learning classifiers for early detection of DDoS attacks on IoT devices," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 1353–1374, 2022.

[17] P. Amangele, M. J. Reed, M. Al-Naday, N. Thomos, and M. Nowak, "Hierarchical Machine Learning for IoT Anomaly Detection in SDN," *2019 ICIT, InfoTech 2019 - Proceedings*, no. 780139, 2019.

[18] K. Ren, Y. Zeng, Z. Cao, and Y. Zhang, "ID-RDRL: a deep reinforcement learning-based feature selection intrusion detection model," *Scientific Reports*, vol. 12, no. 1, pp. 1–18, 2022.

[19] Q.-V. Dang, "Using machine learning for intrusion detection systems," *Computing and Informatics*, vol. 41, no. 1, pp. 12–33, 2022.

[20] O. D. Okey, S. S. Maidin, P. Adasme, R. Lopes Rosa, M. Saadi, D. Carrillo Melgarejo, and D. Zegarra Rodríguez, "BoostedEnML: Efficient technique for detecting cyberattacks in IoT systems using boosted ensemble ML," *Sensors*, vol. 22, no. 19, p. 7409, 2022.

[21] A. P. Singh, S. Kumar, A. Kumar, and M. Usama, "Machine learning based intrusion detection system for minority attacks classification," in *2022 CISES*, pp. 256–261, IEEE, 2022.

[22] M. Ahmad, Q. Riaz, M. Zeeshan, H. Tahir, S. A. Haider, and M. S. Khan, "Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set," *EURASIP JWCN*, vol. 2021, no. 1, 2021.

[23] S. Hanif, T. Ilyas, and M. Zeeshan, "Intrusion Detection in IoT Using Artificial Neural Networks on UNSW-15 Dataset," *HONET-ICT 2019 - IEEE 16th International Conference on Smart Cities: Improving Quality of Life using ICT, IoT and AI*, pp. 152–156, 2019.

[24] Y. K. Saheed, A. I. Abiodun, S. Misra, M. K. Holone, and R. Colomo-Palacios, "A machine learning-based intrusion detection for detecting internet of things network attacks," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9395–9409, 2022.

[25] M. I. Mohmand, H. Hussain, A. A. Khan, U. Ullah, M. Zakarya, A. Ahmed, M. Raza, I. U. Rahman, M. Haleem, *et al.*, "A machine learning-based classification and prediction technique for DDoS attacks," *IEEE Access*, vol. 10, pp. 21443–21454, 2022.

[26] A. Basharat, M. M. B. Mohamad, and A. Khan, "Machine learning techniques for intrusion detection in smart healthcare systems: A comparative analysis," in *2022 4th ICSSA*, pp. 29–33, IEEE, 2022.

[27] G. Tiwari and R. Jain, "Detecting and classifying incoming traffic in a secure cloud computing environment using machine learning and deep learning system," in *2022 IEEE 7th International Conference on Smart Cloud (SmartCloud)*, pp. 16–21, IEEE, 2022.

[28] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things," *IEEE IoT Journal*, vol. 6, no. 4, 2019.

[29] B. Gupta, P. Chaudhary, X. Chang, and N. Nedjah, "Smart defense against distributed denial of service attack in iot networks using supervised learning classifiers," *Comput. Electr. Eng.*, vol. 98, 2022.

[30] S. Adeleye, C. Obruche, R. Idama, B. Iyamah, K. Oladele, and E. Aworonye, "Security concept in internet of things using supervised & unsupervised machine learning algorithm," *GSJ*, vol. 10, no. 2, 2022.

[31] E. Anthi, L. Williams, M. Słowi, G. Theodorakopoulos, and P. Burnap, "A Supervised Intrusion Detection System for Smart Home IoT Devices," *IEEE IoT Journal*, vol. 6, no. 5, pp. 9042–9053, 2019.

[32] G. De Carvalho Bertoli, L. A. Pereira Junior, O. Saotome, A. L. Dos Santos, F. A. N. Verri, C. A. C. Marcondes, S. Barbieri, M. S. Rodrigues, and J. M. Parente De Oliveira, "An End-to-End Framework for Machine Learning-Based Network Intrusion Detection System," *IEEE Access*, vol. 9, pp. 106790–106805, 2021.

[33] E. Anthi, L. Williams, A. Javed, and P. Burnap, "Hardening machine learning denial of service (DoS) defences against adversarial attacks in IoT smart home networks," *Computers and Security*, vol. 108, 2021.

[34] B. Mondal and S. K. Singh, "A comparative analysis of network intrusion detection system for IoT using machine learning," in *IoT and Its App.: Select Proc. of ICIA 2020*, pp. 211–221, Springer, 2022.

[35] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset," *FGCS*, vol. 100, 2019.

[36] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022.

[37] N. Moustafa, "A new distributed architecture for evaluating ai-based security systems at the edge: Network TON_IoT datasets," *Sustainable Cities and Society*, vol. 72, p. 102994, 2021.

[38] H. Hindy, C. Tachtatzis, R. Atkinson, E. Bayne, and X. Bellekens, "MQTT-IoT-IDS2020: MQTT IoT intrusion detection dataset," 2020.

[39] S. Dadkhah, H. Mahdikhani, P. K. Danso, A. Zohourian, K. A. Truong, and A. A. Ghorbani, "Towards the Development of a Realistic Multidimensional IoT Profiling Dataset," in *Anual Int. PST*, pp. 1–11, 2022.

[40] K. K. Huy, "IoT environment dataset." Accessed: 2023-09-15, https://ocslab.hksecurity.net/Datasets/iot-environment-dataset.

[41] K. Hyunjae, A. D. Hyun, L. G. Min, Y. J. Do, P. K. Ho, and K. H. Kang, "IoT network intrusion dataset." Accessed: 2023-09-15, https://dx.doi.org/10.21227/q70p-q449.

[42] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: an ensemble of autoencoders for online network intrusion detection," *arXiv preprint arXiv:1802.09089*, 2018.

[43] S. Garcia, A. Parmisano, and M. Erquiaga, Jose, "IoT-23: A labeled dataset with malicious and benign IoT network traffic," 2020.

[44] V. Paxson, "Bro: a System for Detecting Network Intruders in Real-Time," *Computer Networks*, vol. 31, no. 23-24, pp. 2435–2463, 1999.

[45] Q. LLC, "Argus," *URL https://openargus. org*, 2020.

[46] A. H. Lashkari, G. Draper-Gil, M. S. I. Mamun, A. A. Ghorbani, *et al.*, "Characterization of Tor traffic using time based features," in *ICISSp*, pp. 253–262, 2017.

[47] S. Han, Q. Wu, and Y. Yang, "Machine learning for internet of things anomaly detection under low-quality data," *International Journal of Distributed Sensor Networks*, vol. 18, no. 10, 2022.

[48] Q. Abu Al-Haija and M. Al-Dala'ien, "ELBA-IoT: an ensemble learning model for botnet attack detection in IoT networks," *Journal of Sensor and Actuator Networks*, vol. 11, no. 1, p. 18, 2022.

[49] J. A. Faysal, S. T. Mostafa, J. S. Tamanna, K. M. Mumenin, M. M. Arifin, M. A. Awal, A. Shome, and S. S. Mostafa, "XGB-RF: A hybrid machine learning approach for IoT intrusion detection," in *Telecom*, vol. 3, pp. 52–69, MDPI, 2022.

[50] M. G. Desai, Y. Shi, and K. Suo, "IoT Bonet and Network Intrusion Detection using Dimensionality Reduction and Supervised Machine Learning," *2020 11th IEEE UEMCON 2020*, pp. 0316–0322, 2020.

[51] N. Saran and N. Kesswani, "A comparative study of supervised machine learning classifiers for intrusion detection in internet of things," *Procedia Computer Science*, vol. 218, pp. 2049–2057, 2023.

[52] G. Marín, P. Casas, and G. Capdehourat, "Rawpower: Deep learning based anomaly detection from raw network traffic measurements," in *Proceedings of the ACM SIGCOMM*, pp. 75–77, 2018.

[53] M. J. De Lucia, P. E. Maxwell, N. D. Bastian, A. Swami, B. Jalaian, and N. Leslie, "Machine learning raw network traffic detection," in *AI & ML for Multi-Domain Op. App.*, vol. 11746, pp. 185–194, SPIE, 2021.

[54] K. Albulayhi, Q. Abu Al-Haija, S. A. Alsuhibany, A. A. Jillepalli, M. Ashrafuzzaman, and F. T. Sheldon, "IoT intrusion detection using machine learning with a novel high performing feature selection method," *Applied Sciences*, vol. 12, no. 10, p. 5015, 2022.

[55] R. Yao, N. Wang, P. Chen, D. Ma, and X. Sheng, "A CNN-transformer hybrid approach for an intrusion detection system in advanced metering infrastructure," *Multimedia Tools and Applications*, pp. 1–24, 2022.

[56] B. A. Bhuvaneswari and S. S., "Anomaly detection framework for Internet of things traffic using vector convolutional deep learning approach in fog environment," *FGCS*, vol. 113, pp. 255–265, 2020.

[57] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 56–67, 2020.

[58] M. Leon, T. Markovic, and S. Punnekkat, "Comparative evaluation of machine learning algorithms for network intrusion detection and attack classification," in *2022 IJCNN*, pp. 01–08, IEEE, 2022.

[59] J. Vitorino, I. Praça, and E. Maia, "Towards adversarial realism and robust learning for IoT intrusion detection and classification," *arXiv preprint arXiv:2301.13122*, 2023.

[60] R. Manzano, N. Goel, M. Zaman, R. Joshi, and K. Naik, "Design of a machine learning based intrusion detection framework and methodology for IoT networks," in *12th CCWC*, pp. 0191–0198, IEEE, 2022.

[61] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans Knowl Data Eng*, vol. 21, no. 9, pp. 1263–1284, 2009.

[62] S. Gollapudi, *Practical machine learning*. Packt Publishing Ltd, 2016.

[63] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022.

[64] N. Prazeres, R. L. C. Costa, L. Santos, and C. Rabadão, "Engineering the application of machine learning in an IDS based on IoT traffic flow," *Intelligent Systems with Applications*, p. 200189, 2023.

[65] K. Kostas, M. Just, and M. A. Lones, "IoTDevID: A Behavior-Based Device Identification Method for the IoT," *IEEE IoT Journal*, vol. 9, no. 23, pp. 23741–23749, 2022.

[66] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Adv.in neural info. process. systems*, vol. 30, 2017.

[67] C. Molnar, *Interpretable machine learning*. Lulu. com, 2020.

[68] Regit, "Investigation on an attack tool used in China," 2014. Accessed: 2023-03-16, URL: https://home.regit.org/2014/02/chinese-scanner.

[69] Y. Jin, "IoT remote access attack." Accessed: 2023-03-16, Available at http://cyberforensic.net/labs/remote.html.

[70] P. Maniriho, E. Niyigaba, Z. Bizimana, V. Twiringiyimana, L. J. Mahoro, and T. Ahmad, "Anomaly-based Intrusion Detection Approach for IoT Networks Using Machine Learning," *CENIM*, pp. 303–308, 2020.

[71] V. Bonandrini, J. F. Bercher, and N. Zangar, "Machine Learning Methods for Anomaly Detection in IoT Networks, with Illustrations," *Lecture Notes in Computer Science*, vol. 12081 LNCS, 2020.

[72] A. Verma and V. Ranga, "Machine Learning Based Intrusion Detection Systems for IoT Applications," *Wireless Personal Communications*, vol. 111, no. 4, pp. 2287–2310, 2020.

[73] A. Churcher, R. Ullah, J. Ahmad, S. Ur Rehman, F. Masood, M. Gogate, F. Alqahtani, B. Nour, and W. J. Buchanan, "An experimental analysis of attack classification using machine learning in IoT networks," *Sensors (Switzerland)*, vol. 21, no. 2, pp. 1–32, 2021.

[74] G. Abdelmoumin, D. B. Rawat, and A. Rahman, "On the Performance of Machine Learning Models for Anomaly-Based Intelligent Intrusion Detection Systems for the Internet of Things," *IEEE IoT Journal*, vol. 4662, pp. 1–1, 2021.

[75] Q. Abu Al-Haija, A. Al Badawi, and G. R. Bojja, "Boost-Defence for resilient IoT networks: A head-to-toe approach," *Expert Systems*, vol. 39, no. 10, p. e12934, 2022.

[76] D. K. K. Reddy, H. S. Behera, G. M. Pratyusha, and R. Karri, "Ensemble Bagging Approach for IoT Sensor Based Anomaly Detection," *Lecture Notes in Electrical Engineering*, vol. 702, pp. 647–665, 2021.

[77] R. V. Mendonça, J. C. Silva, R. L. Rosa, M. Saadi, D. Z. Rodriguez, and A. Farouk, "A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithms," *Expert Systems*, vol. 39, no. 5, p. e12917, 2022.

[78] M. A. Siddiqi and W. Pak, "An Agile Approach to Identify Single and Hybrid Normalization for Enhancing Machine Learning-Based Network Intrusion Detection," *IEEE Access*, vol. 9, 2021.

[79] S. Baniasadi, O. Rostami, D. Martín, and M. Kaveh, "A novel deep supervised learning-based approach for intrusion detection in IoT systems," *Sensors*, vol. 22, no. 12, p. 4459, 2022.

[80] A. Mihoub, O. B. Fredj, O. Cheikhrouhou, A. Derhab, and M. Krichen, "Denial of service attack detection and mitigation for internet of things using looking-back-enabled machine learning techniques," *Computers & Electrical Engineering*, vol. 98, p. 107716, 2022.

[81] N. Karmous, M. O.-E. Aoueileyine, M. Abdelkader, and N. Youssef, "A proposed intrusion detection method based on machine learning used for IoT systems," in *the 36th AINA, Volume 3*, Springer, 2022.

[82] K. Ibrahimi and H. Benaddi, "Improving the IDS for BoT-IoT dataset-based machine learning classifiers," in *5th CommNet*, IEEE, 2022.

[83] U. Garg, H. Sivaraman, A. Bamola, and P. Kumari, "To evaluate and analyze the performance of anomaly detection in cloud of things," in *ICCCNT*, pp. 1–7, IEEE, 2022.

[84] T. Saba, A. Rehman, T. Sadad, H. Kolivand, and S. A. Bahaj, "Anomaly-based intrusion detection system for IoT networks through deep learning model," *Computers and Electrical Engineering*, vol. 99, 2022.

[85] M. Sarhan, S. Layeghy, N. Moustafa, M. Gallagher, and M. Portmann, "Feature extraction for machine learning-based intrusion detection in IoT networks," *Digital Communications and Networks*, 2022.

[86] A. A. Alsulami, Q. Abu Al-Haija, A. Tayeb, and A. Alqahtani, "An intrusion detection and classification system for IoT traffic with improved data engineering," *Applied Sciences*, vol. 12, no. 23, 2022.

[87] Y. Chen, Q. Lin, W. Wei, J. Ji, K.-C. Wong, and C. A. C. Coello, "Intrusion detection using multi-objective evolutionary convolutional neural network for internet of things in fog computing," *Knowledge-Based Systems*, vol. 244, p. 108505, 2022.

[88] D. K. K. Reddy and H. Behera, "CatBoosting Approach for Anomaly Detection in IoT-Based Smart Home Environment," in *Computational Intelligence in Data Mining: ICCIDM 2021*, Springer, 2022.

[89] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, K.-K. R. Choo, and M. Nafaa, "Felids: Federated learning-based intrusion detection system for agricultural internet of things," *Journal of Parallel and Distributed Computing*, vol. 165, pp. 17–31, 2022.

[90] P. S. Molcer, A. Pejić, K. Gulači, and R. Szalma, "Machine learning based network intrusion detection system for internet of things cybersecurity," in *Security-Related Adv. Technol. in Critical Infra. Protection: Theoretical and Practical Apch.*, pp. 95–110, Springer, 2022.

[91] A. Ghourabi, "A security model based on LightGBM and transformer to protect healthcare systems from cyberattacks," *IEEE Access*, vol. 10, pp. 48890–48903, 2022.

[92] A. I. Alzahrani, A. Al-Rasheed, A. Ksibi, M. Ayadi, M. M. Asiri, and M. Zakariah, "Anomaly detection in fog computing architectures using custom tab transformer for internet of things," *Electronics*, vol. 11, no. 23, p. 4017, 2022.

[93] X. Han, S. Cui, S. Liu, C. Zhang, B. Jiang, and Z. Lu, "Network intrusion detection based on n-gram frequency and time-aware transformer," *Computers & Security*, vol. 128, p. 103171, 2023.

[94] Z. Wu, H. Zhang, P. Wang, and Z. Sun, "RTIDS: A robust transformer-based approach for intrusion detection system," *IEEE Access*, vol. 10, pp. 64375–64387, 2022.

[95] M. Wang, N. Yang, and N. Weng, "Securing a smart home with a transformer-based IoT intrusion detection system," *Electronics*, vol. 12, no. 9, p. 2100, 2023.

[96] P. Gulihar and B. B. Gupta, "Cooperative mechanisms for defending distributed denial of service (DDoS) attacks," in *Handbook of Computer Networks and Cyber Security*, pp. 421–443, Springer, 2020.

[97] Hewlett-Packard-Enterprise, "Single-packet attacks," in *HPE FlexNetwork 7500 Switch Series Security Configuration Guide*, Hewlett Packard Enterprise Development LP, 2017. Accessed: 2023-09-17, Available at https://techhub.hpe.com/eginfolib/networking/docs/switches/7500/5200-1952a_security_cg/content/495505992.htm.

[98] A. C. Lorena, L. P. Garcia, J. Lehmann, M. C. Souto, and T. K. Ho, "How complex is your classification problem? a survey on measuring classification complexity," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1–34, 2019.

Kahraman Kostas received the MSc degree in Computer Networks and Security from the University of Essex, Colchester, U.K., in 2018. He is a PhD candidate in Computer Science at Heriot-Watt University, Edinburgh, U.K. His research focuses on the security of computer networks and Internet of Things. You can find more information at https://kahramankostas.github.io/.

Mike Just earned his Ph.D. in Computer Science from Carleton University in 1998 and is currently an Associate Professor at Heriot-Watt University. He is primarily interested in computer security, and in applying human-computer interaction and machine learning techniques to solve computer security problems. You can find more information at https://justmikejust.wordpress.com/.

Michael A. Lones (M'01—SM'10) is an Associate Professor of Computer Science at Heriot-Watt University. He received both MEng and PhD degrees from the University of York. He carries out research in the areas of machine learning and optimisation, where he has a particular interest in biologically-inspired approaches and issues of replicability. Application areas of his work include medicine, robotics and security. You can find more information at http://www.macs.hw.ac.uk/∼ml355.

APPENDIX

TABLE IV
PUBLICLY AVAILABLE DATASETS RELEVANT TO ATTACK DETECTION

Datasets	Web address
AWID2	icsdweb.aegean.gr/awid/awid2
BoT-IoT	research.unsw.edu.au/projects/bot-iot-dataset
CICDDoS2019	www.unb.ca/cic/datasets/ddos-2019.html
CICIDS2017	www.unb.ca/cic/datasets/ids-2017.html
CICIDS2018	www.unb.ca/cic/datasets/ids-2018.html
CIDDS-01	https://www.hs-coburg.de/forschung/forschungsprojekte-oeffentlich/informationstechnologie/cidds-coburg-intrusion-detection-data-sets.html
DS2OS	www.kaggle.com/datasets/francoisxa/ds2ostraffictraces?resource=download
ECU-IoHT	github.com/iMohi/ECU-IoFT
Edge-IIoTset	ieee-dataport.org/documents/edge-iiotset-new-comprehensive-realistic-cyber-security-dataset-iot-and-iiot-applications
EMBER	github.com/elastic/ember
InSDN	aseados.ucd.ie/datasets/SDN
IoT-23	www.stratosphereips.org/datasets-iot23
IoTNID	ocslab.hksecurity.net/Datasets/iot-network-intrusion-dataset
ISCXIDS2012	www.unb.ca/cic/datasets/ids.html
KDD99	kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
Kitsune	archive.ics.uci.edu/dataset/516/kitsune+network+attack+dataset
MQTT-IoT-IDS20	ieee-dataport.org/open-access/mqtt-iot-ids2020-mqtt-internet-things-intrusion-detection-dataset
MQTTset	www.kaggle.com/datasets/cnrieiit/mqttset
N-BaIoT	archive.ics.uci.edu/dataset/442/detection+of+iot+botnet+attacks+n+baiot
NSL-KDD	www.kaggle.com/datasets/hassan06/nslkdd
TON_IoT	research.unsw.edu.au/projects/toniot-datasets
UNSW-NB15	research.unsw.edu.au/projects/unsw-nb15-dataset

A. Literature Review Criteria

We applied the following criteria while searching the literature:

- Published in 2019 and later,
- IoT focused,
- Including anomaly detection or intrusion detection,
- Used supervised machine learning methods.

We did not include Wireless Sensor Networks (WSN), which is included under IoT in many studies. This is because WSN studies generally focus on radio frequency-based attacks at the physical layer, which is beyond our scope of research. We searched Google Scholar using the conditions mentioned above and the following queries: "(IoT OR internet of things) AND ("anomaly detection" OR "Intrusion detection" OR IDS) AND (machine learning OR deep learning) AND (Supervised)". This identified about 350 articles. After reviewing these, we eliminated unrelated ones (articles not including IoT device, intrusion/attack detection, or ML methods). The remaining articles are listed in Table V with summary information such as dataset, ML algorithm(s) and publication year.

B. Data Leakage in the Use of Individual Features

It is not possible to detect an attack using individual packet features alone. This situation can be explained with a simple example: SYN Flood [96], which is a DoS attack based on the exploit of 3-handshake in the TCP protocol. During the 3-way handshake process, the client that wants to establish a connection sends a synchronization (SYN) packet to the server. The server receiving this packet sends a synchronization and acknowledgment (SYN-ACK) packet to the client. Finally, the client responds with the acknowledgment (ACK) packet, and the TCP connection is established (see Fig. 16).

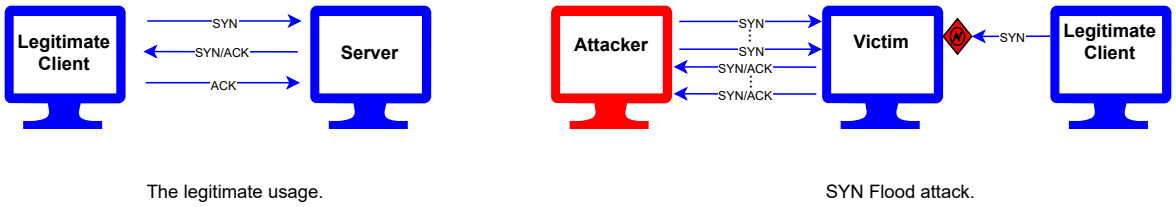


Fig. 16. Legitimate and malicious use of 3-way handshake.

In the SYN Flood attack, the attacker sends a large number of SYN packets to the server. The server sends the SYN-ACK packet in response and waits for the ACK packet from the other side to establish the connection. The attacker does not

TABLE V
COMPARISON OF RELATED WORK. ONLY THE BEST SCORES ARE INCLUDED IN THE EVALUATION SECTION.

		Dataset																			Machine Learning														Feature Type										Evaluation					
Year	Study	KDD99	NSL-KDD	ISCXIDS2012	CICIDS2017	CICIDS2018	CICDDoS2019	UNSW-NB15	BoT-IoT	TON_IoT	IoTNI	IoT-23	DS2OS	N-BaIoT	Private	CIDDS-01	Edge-IIoTset	MQTT-IoTIDS20	Other	Bagging	Boosting	CNN	Decision Tree	Tree Ensemble	kNN	Linear Models	MLP	Naive Bayes	RNN	SVM	Transformer	Argus	DS2OS	Flowmeter	KDD	Kitsune	NetFlow	Other	Packet Based	Zeek	Confusion Matrix	Accuracy	Recall	Precision	F1 Score	AUC
2019	[12]		●																					●					●					●								93	92	100	96	
	[14]				●																	●		●			●	●	●				●									97	99	98		
	[15]				●																		●		●		●			●			●								●	98				
	[17]				●																		●		●	●		●					●									100	99	99	99	
	[23]							●																			●					●								●		84					
	[28]														●								●	●	●	●	●	●		●		●								●	100	97				
	[31]														●								●	●	●		●		●									●			●			98	99	99	
2020	[6]		●							●																			●					●								100	99			
	[56]								●																		●		●			●								●	100	100	100		
	[70]										●													●								●										100	100		
	[50]										●													●											●						●	100	99	99		
	[71]															●								●	●		●									●					●	100	100	100	100	
	[72]		●					●								●					●			●	●		●					●			●		●			●	95	100			99
2021	[8]		●																	●	●		●											●						●	100	99			
	[9]	●																					●		●		●	●		●				●								95	91	96	93	
	[16]						●														●		●	●									●								98	98	98	98	
	[22]							●																●			●		●		●								●	●	99				
	[73]								●														●	●	●	●	●	●		●		●								●	97	97	97	97	
	[74]								●	●																	●			●		●									100	100	100	100	100
	[75]										●										●		●									●								●	100	96	99		100
	[76]												●										●	●	●							●								●	100	100	100	100	100
	[77]												●										●	●			●		●			●	●								99	98	98	98	
					●																		●	●			●		●			●	●								99	98	99	99	
	[78]								●															●			●		●			●	●						●	●	99	99	98	98	
					●																			●			●		●			●	●						●	●	100	100	99	100	
				●																				●			●		●			●	●						●	●	96	94	94	94	
			●																					●			●		●			●	●						●	●	100	100	100	100	
								●																●			●		●			●	●						●	●	77	68	55	57	
	[32]																		●		●		●	●	●	●	●		●									●						100		
	[33]														●								●	●			●		●									●			●				100	100	100	
2022	[7]		●																					●		●			●					●								98				
	[10]		●																				●	●	●	●	●		●					●							●	79	79	84	78	96
	[11]	●																			●	●	●	●				●	●	●				●								93	92	100	95	98
			●																		●	●	●	●				●	●	●				●								81	81	97	82	92
	[34]														●						●		●	●	●	●		●											●		100	100	100	100	
	[13]	●																					●	●	●	●								●							●	99	99	100	100	
	[18]					●																●				●	●		●				●								●	96			95	98
	[19]					●															●			●		●		●		●			●								●					100
	[20]				●																●		●	●									●								●	100	100	100	100	100
						●															●		●	●									●								●	100	100	100	100	100
	[21]				●																		●	●	●								●								●	100	100	100	100	100
	[24]							●													●				●	●		●		●		●								●		100		100	100	
	[25]							●													●						●					●								●	●	90	90	90	90	
	[26]							●												●	●						●			●		●								●	●	96	96			
	[27]							●													●		●	●	●	●	●		●			●								●	●	98	98	98	98	
	[79]							●	●														●									●								●		99	99			
	[80]								●															●	●	●		●		●		●								●	●	100				
	[60]								●															●								●										98				
	[81]									●														●	●	●	●		●		●													100	100
	[82]								●													●		●	●				●											●	●	99	99	99	99	
	[83]								●														●			●			●											●	●	99	99	100	100	100
	[84]	●																					●									●			●					●	●	100				

TABLE V
COMPARISON OF RELATED WORK. ONLY THE BEST SCORES ARE INCLUDED IN THE EVALUATION SECTION (CONTINUED).

Year	Study	Dataset																			Machine Learning												Feature Type									Evaluation					
		KDD99	NSL-KDD	ISCXIDS2012	CICIDS2017	CICIDS2018	CICDDoS2019	UNSW-NB15	BoT-IoT	TON_IoT	IoTNI	IoT-23	DS2OS	N-BaIoT	Private	CIDDS-01	Edge-IIoTset	MQTT-IoTIDS20	Other	Bagging	Boosting	CNN	Decision Tree	Ensemble Trees	kNN	Linear Models	MLP	Naive Bayes	RNN	SVM	Transformer	Argus	DS2OS	Flowmeter	KDD	Kitsune	NetFlow	Other	Packet Based	Zeek	Confusion Matrix	Accuracy	Recall	Precision	F1 Score	AUC
2022	[85]					●		●														●	●			●	●	●	●			●		●					●		98	95		94	97
																						●	●			●	●	●	●			●		●					●		99	92		92	96
										●												●	●			●	●	●	●			●		●					●		98	100		99	72
	[86]										●									●			●		●		●		●				●							●	100				
	[87]																		●			●										●				●				●	100	100	100	100	
					●																	●										●				●				●	100	100	100	100	
	[88]												●								●											●								●	99	96	100	98	100
	[29]														●								●	●	●	●		●		●							●				99	99	97	98	
	[54]										●									●			●		●		●						●	●						●	100	100	100	100	100
			●																	●			●		●		●						●	●						●	100	100	100	100	100
	[89]					●																●					●		●				●					●			94	92	82	84	
																			●			●					●		●				●					●			91	91	77	80	
																			●			●					●		●				●					●			99	85	92	86	
	[47]	●																	●				●	●		●	●	●		●				●	●							100	100	100	
	[58]	●																						●	●	●	●			●		●			●	●				●		100				
			●																					●	●	●	●			●		●			●	●				●		100				
								●																●	●	●	●			●		●			●	●				●		97				
					●																			●	●	●	●			●		●			●	●				●		100				
	[49]													●							●			●											●					●	100	100	100	100	
	[30]														●									●	●		●			●							●				91	94	88	88	
	[90]									●											●				●		●									●					100	100			100
	[36]																●						●	●	●		●			●								●		●	100	95	89	87	
	[91]									●							●		●		●								●		●	●					●		●		100	100	100	100	100
	[92]							●														●	●	●	●	●		●	●	●	●						●			●	●	98	98	96	97	
	[93]			●																	●	●	●		●	●		●	●	●								●			●	99	99	99	99	
					●																●	●	●		●	●		●	●	●								●			●	100	100	100	100	
	[94]				●																						●		●	●	●											99	99	99	99	
							●																				●		●	●	●											99	99	99	98	
2023	[59]								●			●									●			●							●								●		99				
	[48]													●						●	●														●					●		100	100	100	100	
	[51]																	●					●	●	●	●		●		●								●			100	100	100	100	
	[64]											●															●											●		●			100	100	100	
																		●									●											●		●			100	100	100	
	[55]	●																					●		●			●	●				●	●						●	98			73	
			●																				●		●			●	●				●	●						●	92			82	
					●																		●		●			●	●				●	●						●	91			76	
	[95]									●											●			●			●									●					●	98			98	100

send these packets, resulting in a large number of unfinished handshakes. The server whose resources are exhausted cannot then serve legitimate users. If we look at these SYN packets individually, we cannot see any features that distinguish them from each other (except for identifying features such as IP or MAC addresses. This information is useless in attack detection because it is impossible to know beforehand). It is impossible to tell whether SYN packets are viewed individually as benign or malicious. However, if we look at the whole, we see that few or no ACK packets are received despite thousands of SYN packets from the same source. So while looking at the individual packet is useless for detection, it is possible to detect this attack by looking at the whole. This is roughly the same as other types of attacks. However, like with any generalizations, there are exceptions. One example is single-packet attacks that exploit malformed packets. These attacks typically involve sending malformed network packets that the device cannot handle, leading to malfunctions or crashes in the receiving device [97]. However, addressing these attacks usually involves packet filtering or firewalls, not machine learning behavioural analysis, since they can be effectively managed using signature/rule-based approaches.

The high scores obtained by using individual packet features can be explained in two different ways. Identifying features: they can be specific features that identify a target or an attacker such as IP address, MAC address, or in some cases port number. They can be features that identify or give hints about a session. The use of these features will lead to over-fitting of the model, resulting in systems that are not useful in real-life data.

On the other hand, apart from some specific attacks, most attacks follow a pattern and produce uniform outputs. For example,

TABLE VI
SUMMARY FEATURE LIST, AVOIDING REPETITION. THE COMPLETE LIST OF FEATURES AND DESCRIPTIONS ARE AVAILABLE AT: GITHUB.COM/KAHRAMANKOSTAS/HERIOT/BLOB/MAIN/FEATURELIST.MD

Feature	Explanation	Feature	Explanation
ts	Time Stamp	ID	Source + Destination
Ether_dst	Destination MAC Address	pck_size_diff	the size difference of consecutive packets
Ether_src	Source MAC Address	pck_size_mean_WE	EW (expanding windows) packet size mean
IP_src	Source IP Address	pck_size_std_WE	EW packet size Std.
IP_dst	Destination IP Address	pck_size_sum_of_EW	EW packet size total.
pck_size	Packet (Frame) Size	ts_diff	the time difference of consecutive packets
Ether_type	Ethernet Type	ts_mean_WE	EW packet time mean
IP_ihl	IP Internet Header Length	ts_std_WE	EW packet time Std.
IP_tos	IP type of service	ts_sum_of_EW	EW packet time total.
IP_len	IP Length	TCP_window_diff	TCP Windows size difference of consecutive packets
IP_MF	IP More Fragments	TCP_window_mean_WE	EW TCP Windows size mean
IP_id	IP identifier	TCP_window_std_WE	EW TCP Windows size Std.
IP_chksum	IP Checksum	TCP_window_sum_of_EW	EW TCP Windows size total.
IP_DF	IP Don't Fragment	payload_bytes_diff	Payload bytes difference of consecutive packets
IP_frag	IP fragmentation	payload_bytes_mean_WE	EW Payload bytes mean
IP_ttl	IP Time To Live	payload_bytes_std_WE	EW Payload bytes Std.
IP_proto	IP Protocols	payload_bytes_sum_of_EW	EW Payload bytes total.
TCP_seq	TCP Sequence Number	entropy_diff	Entropy difference of consecutive packets
TCP_ack	TCP Acknowledgment Number	entropy_mean_WE	EW Entropy mean
TCP_dataofs	TCP data ofset	entropy_std_WE	EW Entropy Std.
TCP_flags	TCP Flags	entropy_sum_of_EW	EW Entropy total.
TCP_FIN	FINished Flag	pck_size_mean_2	RW (Rolling windows) packet size mean
TCP_SYN	Sync Flag	pck_size_std_2	RW packet size Std. - Window size =2
TCP_RST	Reset Flag	ts_mean_2	RW packet time mean - Window size =2
TCP_PSH	Push Flag	ts_std_2	RW packet time Std.- Window size =2
TCP_ACK	Acknowledgment Flag	TCP_window_mean_2	RW TCP Windows size mean - Window size =2
TCP_URG	Urgent Flag	TCP_window_std_2	RW TCP Windows size Std. - Window size =2
TCP_ECE	ECE Flag	payload_bytes_mean_2	RW Payload bytes mean - Window size =2
TCP_CWR	CWR Flag	payload_bytes_std_2	RW Payload bytes Std. - Window size =2
TCP_window	TCP Window Size	entropy_mean_2	RW Entropy mean - Window size =2
TCP_chksum	TCP Checksum	entropy_std_2	RW Entropy Std. - Window size =2
payload_bytes	Payload size in Bytes	dport_sum	EW Unique destination port number
entropy	Payload Entropy	sport_sum	EW Unique source port number
Protocol	WireShark Protocol	TCP_FIN_sum	EW FINished Flag
dst_IP_diversity	Number of Destination IP (by Source IP)	TCP_FIN_ratio	TCP_FIN/TCP_FIN_sum
dst_port_diversity	Number of Destination Port (by Source IP)	sum	EW all TCP Flags
src_IP_diversity	Number of Source IP (by destination IP)	TCP_FIN_SR	TCP_FIN/sum
dport	Destination Port Number	IP_add_count	Number of Source IP (by destination MAC)
		sport	Source Port Number

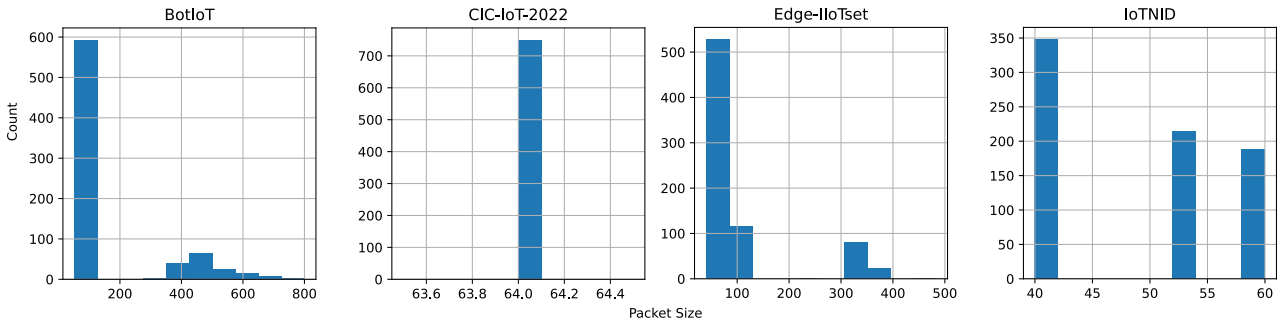
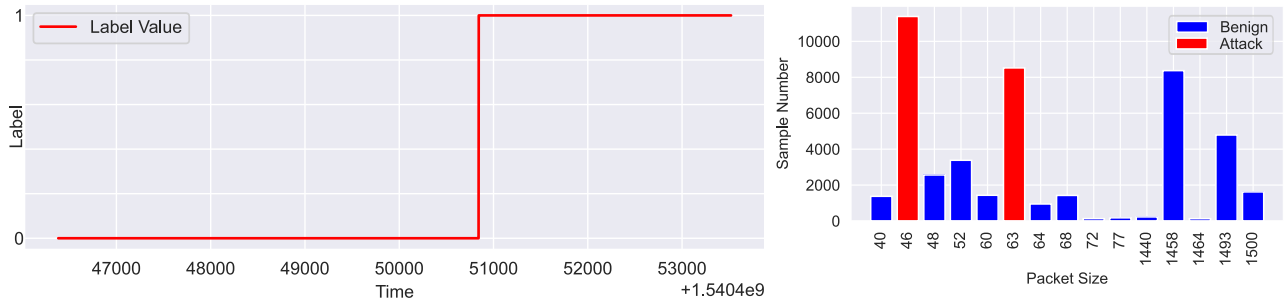


Fig. 17. The distribution of packet sizes of malicious data (HTTP Flood attack) in four different datasets.



(a) In Kitsune dataset - Mirai attack, the labels follow a certain time pattern.

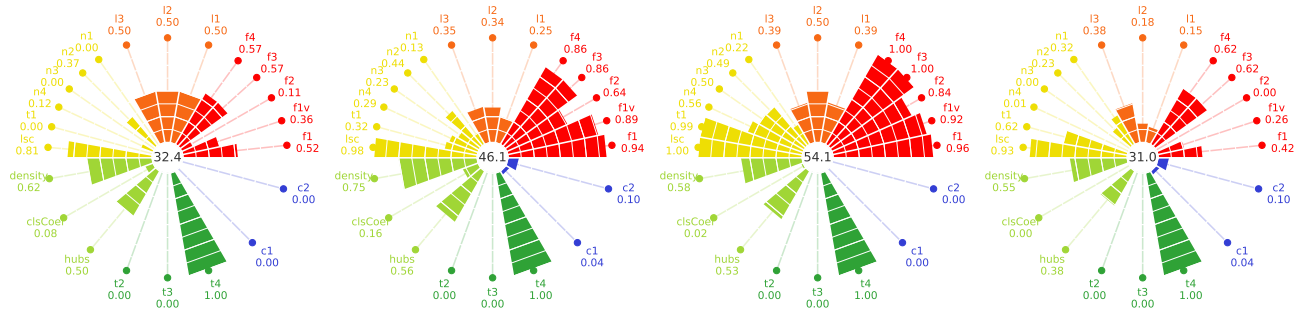
(b) In the IoT-NID dataset - UDP flood attack, attack packets have a specific size.

Fig. 18. Distribution of labels in Kitsune and IoT-NID datasets.

the distribution of the size of the attack packets in the HTTP flood attack on four different datasets is shown in Fig. 17. Because of this uniform nature of the attacks, if the complexity level of the dataset is low in studies using individual packets, basic characteristics, such as size or time, can become identifying features. For example, by using only packet size in the IoT-NID UDP dataset or only time in the Kitsune Mirai dataset, a near-perfect classification can be made (see Table VII). This is only because the complexity of these datasets in terms of these features is quite low (see the complexity analyses in Fig. 19). However, these classification successes will be specific to this dataset, and so the model will not be successful on other datasets, since they have different time and size patterns. Fig. 20 shows the results of another experiment on complexity. In this experiment, a basic feature (size) was used to separate the attack data. Especially if the complexity is below the critical level (45), a high rate of discrimination is achieved. As complexity increases, success decreases.

TABLE VII
NEAR-PERFECT CLASSIFICATION OF 2 DATASETS AFFLICTED BY LOW DATA COMPLEXITY. IN KITSUNE DATA, THE LABELS FOLLOW A CERTAIN TIME PATTERN. IN THE IOT-NID DATASET, ATTACK PACKETS HAVE A SPECIFIC SIZE.

Dataset	Attack	Feature	ML	Acc	Prec	Rec	F1	Kappa
Kitsune	Mirai	Time Stamp	ExtraTree	0.952	0.885	0.972	0.920	0.841
IoT-NID	UDP	Packet Size	ExtraTree	1.000	1.000	0.999	1.000	0.999



(a) Mirai Attack Time Complexity (b) UPD Flood Attack Time Complexity (c) Mirai Attack Size Complexity (d) UPD Flood Attack Size Complexity

Fig. 19. Complexity analysis of UDP and Mirai attacks according to time and size characteristics. The central score depicted in the figures represents the comprehensive complexity score, which is derived as the average of 22 distinct analytical methods (The complexity score ranges from 0 to 100, where higher values indicate increasing complexity). These 22 methods stem from six different approaches to assessing complexity, each distinctly colour-coded for clarity of categorization [98]: red for feature-based, orange for linearity-based, yellow for neighbourhood-based, light green for network-based, dark green for dimensionality-based, and blue for class imbalance-based. When these figures are analysed, it can be seen that the UDP flood attack has low size complexity and high time complexity. On the other hand, Mirai attack show low size complexity and high time complexity.

C. Issues with the IoTID20 Dataset

We performed the feature extraction process by applying CICFlowMeter to all the pcap files we used. Although a dataset (IoTID20) has already been produced using CICFlowMeter from IoT-NID raw data, we preferred to redo the feature extraction process instead of using it. We used the CICFlowMeter[4] tool to extract features from the dataset and labeled them in parallel

[4]https://github.com/ahlashkari/CICFlowMeter

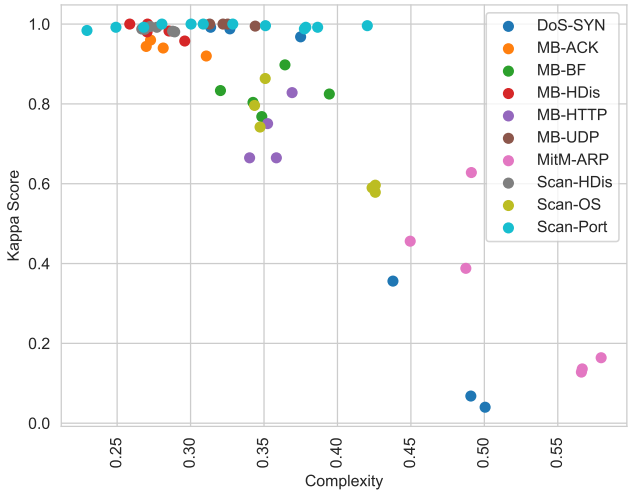


Fig. 20. Reflection of how dataset complexity affects the predictive value of individual features.

with the defined rules in IoT-NID dataset [41] using our flow-labeller script[5]. We believe that there may have been issues with the feature extraction process of the IoTID20 dataset. Some clues leading to this conclusion are as follows. Although the raw data contains 10 different attacks, there are 8 attacks in the dataset. It is conceivable that some pcap files are labelled as if they contain a single attack, ignoring the fact that they contain more than one type of attack.

When examining the content generator directives in the IoT-NID dataset [41], distinct pcap files are apparent, each corresponding to a specific attack and session. However, these files encompass not only malicious packets but also benign ones within each attack file. This becomes most apparent when considering that the benign file contains 137,000 packets, whereas the entire dataset comprises over 15,000,000 benign packets. Notably, this dataset lacks pre-defined labels; solely raw data and filtering rules for packet-level labelling are provided. Employing the rules from content creators facilitates the separation of benign and malicious segments within all files. Yet, in the context of IoTID20, it appears that these rules were disregarded, resulting in the mislabelling of all pcap files as attacks. Substantiating this assertion, we observe the following: a flow comprises either a single packet or a combination of several. In this regard, the number of flows derived from a pcap file must consistently be less than or equal to the number of packets. However, a scrutiny of IoTID20 reveals numerous instances where this principle is violated, with the count of attacks often surpassing the expected count (refer to Table VIII). For many attacks such as HTTP, BF, OSD, PS, MHD, and SHD the number of samples provided in IoTID20 is much higher than the number of samples in the raw data.

TABLE VIII
THE COMPARISON OF THE NUMBER OF PACKETS IN THE IOT-NID RAW DATA WITH THE NUMBER OF FLOWS IN THE IOTID20 DATASET

Type	Label	Packets	Flows
Normal	Normal	1,756,276	40,073
Mirai Botnet	Host Discovery (MHD)	673	22,192
Scanning	Host Discovery (SHD)	2,454	
Scanning	Port Scanning (PS)	20,939	53,073
Scanning	OS/Version Detection (OSD)	1,817	
Man in the Middle (MITM)	ARP Spoofing (ARPS)	101,885	35,377
Denial of Service (DoS)	SYN Flooding (SYN)	64,646	59,391
Mirai Botnet	Telnet Bruteforce (BF)	1,924	121,181
Mirai Botnet	UDP Flooding (UDP)	949,284	183,554
Mirai Botnet	ACK Flooding (ACK)	75,632	55,124
Mirai Botnet	HTTP Flooding(HTTP)	10,464	55,818

D. Full ML Results

[5]https://github.com/kahramankostas/IoTGeM/blob/main/0001-Feature-Extraction-PCAP2CSV/000-FLOW-LABELLER.ipynb

TABLE IX
WINDOW APPROACH CV RESULTS. T IS TIME

Window approach CV results.

Attack	ML	Accuracy	BalancedAcc	Precision	Recall	F1-Score	kappa	ROC	Train-T	Test-T
ACK	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.106	0.000
	KNN	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.130	0.128
	LR	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.437	0.000
	MLP	1.000±0.000	1.000±0.001	1.000±0.000	1.000±0.001	1.000±0.001	1.000±0.001	1.000±0.001	16.252	0.132
	NB	1.000±0.000	0.999±0.001	1.000±0.000	0.999±0.001	1.000±0.000	0.999±0.001	0.999±0.001	0.031	0.006
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	5.534	0.103
	SVM	0.948±0.005	0.631±0.011	0.974±0.002	0.631±0.011	0.694±0.014	0.398±0.027	0.631±0.011	16.365	1.069
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	14.246	0.064
ARP	DT	0.995±0.001	0.994±0.002	0.993±0.002	0.994±0.002	0.994±0.002	0.987±0.004	0.994±0.002	0.440	0.003
	KNN	0.944±0.002	0.928±0.003	0.922±0.002	0.928±0.003	0.925±0.002	0.850±0.005	0.928±0.003	0.202	0.116
	LR	0.869±0.008	0.784±0.015	0.846±0.011	0.784±0.015	0.808±0.014	0.618±0.027	0.784±0.015	39.338	0.001
	MLP	0.760±0.059	0.651±0.063	0.758±0.103	0.651±0.063	0.642±0.053	0.315±0.089	0.651±0.063	16.574	0.015
	NB	0.759±0.007	0.529±0.004	0.705±0.050	0.529±0.004	0.494±0.009	0.082±0.012	0.529±0.004	0.027	0.013
	RF	0.998±0.001	0.997±0.001	0.997±0.001	0.997±0.001	0.997±0.001	0.994±0.002	0.997±0.001	11.375	0.091
	SVM	0.755±0.008	0.509±0.002	0.877±0.004	0.509±0.002	0.447±0.004	0.027±0.005	0.509±0.002	68.318	4.711
	XGB	0.998±0.001	0.997±0.001	0.997±0.001	0.997±0.001	0.997±0.001	0.994±0.002	0.997±0.001	14.762	0.042
BF	DT	1.000±0.000	0.999±0.002	0.999±0.002	0.999±0.002	0.999±0.001	0.999±0.002	0.999±0.002	0.159	0.003
	KNN	0.999±0.000	0.982±0.008	0.976±0.010	0.982±0.008	0.979±0.007	0.957±0.013	0.982±0.008	0.004	2.133
	LR	1.000±0.000	0.999±0.002	0.992±0.006	0.999±0.002	0.996±0.003	0.991±0.006	0.999±0.002	2.328	0.001
	MLP	0.989±0.005	0.929±0.060	0.951±0.022	0.929±0.060	0.938±0.039	0.875±0.078	0.929±0.060	53.638	0.033
	NB	0.972±0.001	0.659±0.031	0.577±0.019	0.659±0.031	0.602±0.023	0.205±0.046	0.659±0.031	0.025	0.008
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	8.303	0.108
	SVM	0.995±0.001	0.951±0.014	0.998±0.001	0.951±0.014	0.973±0.008	0.946±0.016	0.951±0.014	2.791	0.520
	XGB	1.000±0.000	0.996±0.004	1.000±0.000	0.996±0.004	0.998±0.002	0.995±0.004	0.996±0.004	9.904	0.039
HTTP	DT	1.000±0.000	0.999±0.001	1.000±0.000	0.999±0.001	1.000±0.000	0.999±0.001	0.999±0.001	0.227	0.001
	KNN	1.000±0.000	0.998±0.001	0.998±0.001	0.998±0.001	0.998±0.001	0.996±0.001	0.998±0.001	0.257	0.550
	LR	0.999±0.001	0.997±0.002	0.999±0.008	0.997±0.002	0.994±0.005	0.988±0.009	0.997±0.002	1.618	0.002
	MLP	0.998±0.003	0.990±0.012	0.982±0.030	0.999±0.012	0.985±0.016	0.971±0.032	0.999±0.012	4.185	0.006
	NB	0.849±0.003	0.922±0.001	0.606±0.003	0.922±0.001	0.632±0.004	0.303±0.007	0.922±0.001	0.043	0.014
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	9.788	0.129
	SVM	0.960±0.002	0.500±0.000	0.480±0.001	0.500±0.000	0.490±0.000	0.000±0.000	0.500±0.000	38.264	2.876
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	14.758	0.052
MHD	DT	1.000±0.000	0.998±0.007	0.999±0.014	0.998±0.007	0.994±0.007	0.987±0.015	0.998±0.007	0.130	0.000
	KNN	1.000±0.000	0.999±0.004	0.965±0.019	0.999±0.004	0.981±0.011	0.963±0.021	0.999±0.004	0.207	1.006
	LR	1.000±0.000	1.000±0.000	0.969±0.019	1.000±0.000	0.984±0.010	0.967±0.021	1.000±0.000	1.259	0.000
	MLP	1.000±0.001	0.996±0.013	0.995±0.008	0.996±0.013	0.995±0.008	0.990±0.015	0.996±0.013	321.473	0.157
	NB	0.972±0.001	0.986±0.000	0.537±0.006	0.986±0.000	0.562±0.010	0.134±0.020	0.986±0.000	0.044	0.017
	RF	1.000±0.000	1.000±0.000	0.992±0.009	1.000±0.000	0.996±0.004	0.992±0.009	1.000±0.000	2.198	0.107
	SVM	0.981±0.002	0.533±0.018	0.990±0.001	0.533±0.018	0.557±0.032	0.121±0.062	0.533±0.018	5.476	0.788
	XGB	1.000±0.000	1.000±0.000	0.993±0.012	1.000±0.000	0.996±0.006	0.992±0.013	1.000±0.000	21.24	0.063
OS	DT	1.000±0.000	0.999±0.002	0.998±0.002	0.999±0.002	0.999±0.001	0.997±0.002	0.999±0.002	0.321	0.003
	KNN	0.998±0.000	0.952±0.015	0.937±0.015	0.952±0.015	0.944±0.010	0.888±0.021	0.952±0.015	0.013	5.711
	LR	1.000±0.000	0.998±0.004	0.998±0.003	0.998±0.004	0.998±0.003	0.996±0.005	0.998±0.004	0.326	0.002
	MLP	0.943±0.006	0.500±0.000	0.471±0.003	0.500±0.000	0.485±0.001	0.000±0.001	0.500±0.000	151.776	0.053
	NB	0.931±0.001	0.965±0.001	0.560±0.006	0.965±0.001	0.590±0.010	0.202±0.018	0.965±0.001	0.061	0.018
	RF	0.999±0.002	0.966±0.093	0.997±0.004	0.966±0.093	0.975±0.067	0.951±0.134	0.966±0.093	0.090	0.008
	SVM	0.943±0.005	0.500±0.000	0.471±0.003	0.500±0.000	0.485±0.001	0.000±0.000	0.500±0.000	10.279	1.134
	XGB	1.000±0.000	0.998±0.002	0.998±0.002	0.998±0.002	0.998±0.001	0.997±0.002	0.998±0.002	3.201	0.035
SHD	DT	1.000±0.000	0.999±0.003	0.998±0.003	0.999±0.003	0.999±0.002	0.997±0.004	0.999±0.003	0.040	0.000
	KNN	1.000±0.000	1.000±0.000	0.997±0.004	1.000±0.000	0.998±0.002	0.997±0.004	1.000±0.000	0.054	0.046
	LR	1.000±0.000	0.999±0.001	0.997±0.003	0.999±0.001	0.998±0.002	0.996±0.004	0.999±0.001	0.165	0.000
	MLP	1.000±0.000	1.000±0.000	0.999±0.002	1.000±0.000	0.999±0.001	0.999±0.002	1.000±0.000	129.221	0.053
	NB	0.960±0.004	0.500±0.000	0.480±0.002	0.500±0.000	0.490±0.001	0.000±0.000	0.500±0.000	0.006	0.001
	RF	1.000±0.000	1.000±0.000	0.999±0.002	1.000±0.000	0.999±0.001	0.999±0.002	1.000±0.000	0.790	0.027
	SVM	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	3.350	0.814
	XGB	1.000±0.000	1.000±0.000	0.998±0.004	1.000±0.000	0.999±0.002	0.998±0.004	1.000±0.000	0.735	0.008
PS	DT	1.000±0.000	0.999±0.001	1.000±0.000	0.999±0.001	1.000±0.001	0.999±0.001	0.999±0.001	0.008	0.000
	KNN	1.000±0.000	1.000±0.000	0.999±0.002	1.000±0.000	1.000±0.001	0.999±0.002	1.000±0.000	0.016	0.017
	LR	1.000±0.000	0.999±0.001	0.999±0.002	0.999±0.001	0.999±0.001	0.999±0.002	0.999±0.001	0.073	0.003
	MLP	1.000±0.000	1.000±0.001	1.000±0.000	1.000±0.001	1.000±0.000	1.000±0.001	1.000±0.001	7.311	0.016
	NB	0.999±0.001	0.994±0.001	0.953±0.006	0.994±0.001	0.972±0.004	0.945±0.008	0.994±0.001	0.003	0.002
	RF	1.000±0.001	0.999±0.003	0.999±0.002	0.999±0.003	0.999±0.002	0.998±0.003	0.999±0.003	0.013	0.000
	SVM	0.890±0.007	0.500±0.001	0.545±0.212	0.500±0.001	0.472±0.003	0.002±0.003	0.500±0.001	6.131	0.900
	XGB	1.000±0.000	0.999±0.001	1.000±0.001	0.999±0.001	0.999±0.001	0.999±0.001	0.999±0.001	2.479	0.012
SYN	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.033	0.001
	KNN	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.998±0.001	0.999±0.000	0.180	3.301
	LR	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.001	1.000±0.000	0.079	0.001
	MLP	0.999±0.001	0.999±0.001	0.999±0.001	0.999±0.001	0.999±0.001	0.998±0.002	0.999±0.001	168.516	0.168
	NB	0.994±0.001	0.994±0.001	0.993±0.001	0.994±0.001	0.993±0.001	0.987±0.001	0.994±0.001	0.022	0.010
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	2.135	0.083
	SVM	0.946±0.004	0.941±0.005	0.954±0.004	0.941±0.005	0.945±0.004	0.890±0.009	0.941±0.005	27.904	2.189
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	14.174	0.041
UDP	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.278	0.006
	KNN	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.538	56.370
	LR	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	2.819	0.007
	MLP	0.999±0.001	0.999±0.000	0.999±0.001	0.999±0.000	0.999±0.001	0.999±0.001	0.999±0.000	236.567	0.176
	NB	0.674±0.004	0.730±0.002	0.775±0.002	0.730±0.002	0.669±0.004	0.403±0.005	0.730±0.002	0.091	0.028
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	3.728	0.127
	SVM	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.076	0.003
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	38.689	0.121

TABLE X
WINDOW APPROACH SESSION RESULTS. T IS TIME

Window approach Session results.

Attack	ML	Accuracy	BalancedAcc	Precision	Recall	F1-Score	kappa	ROC	Train-T	Test-T
ACK	DT	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.999±0.000	0.999±0.001	1.000±0.000	0.021	0.002
	KNN	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.032	0.190
	LR	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.098	0.003
	MLP	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.999±0.000	0.999±0.000	1.000±0.000	26.160	0.814
	NB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.003	0.012
	RF	0.999±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.999±0.000	0.998±0.000	1.000±0.000	1.113	0.169
	SVM	0.306±0.000	0.501±0.000	0.653±0.000	0.501±0.000	0.235±0.000	0.001±0.000	0.501±0.000	29.912	6.743
	XGB	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	2.940	0.065
ARP	DT	0.653±0.024	0.583±0.034	0.597±0.036	0.583±0.034	0.583±0.037	0.175±0.069	0.583±0.034	0.551	0.005
	KNN	0.688±0.000	0.604±0.000	0.641±0.000	0.604±0.000	0.607±0.000	0.23±0.000	0.604±0.000	0.271	0.664
	LR	0.661±0.011	0.636±0.011	0.629±0.011	0.636±0.011	0.631±0.011	0.264±0.022	0.636±0.011	57.483	0.003
	MLP	0.655±0.064	0.571±0.051	0.615±0.068	0.571±0.051	0.563±0.006	0.157±0.108	0.571±0.051	29.770	0.125
	NB	0.662±0.000	0.502±0.000	0.545±0.000	0.502±0.000	0.411±0.000	0.006±0.000	0.502±0.000	0.024	0.039
	RF	0.731±0.019	0.655±0.023	0.704±0.026	0.655±0.023	0.664±0.026	0.340±0.049	0.655±0.023	13.298	0.331
	SVM	0.674±0.000	0.516±0.000	0.835±0.000	0.516±0.000	0.432±0.000	0.041±0.000	0.516±0.000	184.752	36.351
	XGB	0.785±0.000	0.718±0.000	0.779±0.000	0.718±0.000	0.733±0.000	0.475±0.000	0.718±0.000	15.000	0.106
BF	DT	0.831±0.006	0.876±0.002	0.514±0.000	0.876±0.002	0.481±0.002	0.045±0.002	0.876±0.002	0.041	0.016
	KNN	0.998±0.000	0.987±0.000	0.877±0.000	0.987±0.000	0.925±0.000	0.850±0.000	0.987±0.000	0.002	2.146
	LR	0.896±0.002	0.911±0.001	0.523±0.000	0.911±0.001	0.516±0.001	0.078±0.001	0.911±0.001	0.270	0.005
	MLP	0.994±0.001	0.969±0.016	0.968±0.006	0.969±0.016	0.969±0.008	0.937±0.016	0.969±0.016	137.646	0.283
	NB	0.980±0.000	0.747±0.000	0.566±0.000	0.747±0.000	0.601±0.000	0.206±0.000	0.747±0.000	0.009	0.086
	RF	0.999±0.000	0.963±0.002	0.984±0.001	0.963±0.002	0.974±0.001	0.947±0.002	0.963±0.002	2.456	1.030
	SVM	0.995±0.000	0.948±0.000	0.997±0.000	0.948±0.000	0.971±0.000	0.943±0.000	0.948±0.000	7.031	3.603
	XGB	0.997±0.000	0.960±0.000	0.851±0.000	0.960±0.000	0.898±0.000	0.796±0.000	0.960±0.000	2.895	0.159
HTTP	DT	0.999±0.002	0.988±0.014	0.999±0.001	0.988±0.014	0.993±0.008	0.987±0.015	0.988±0.014	0.035	0.003
	KNN	0.981±0.000	0.834±0.000	0.985±0.000	0.834±0.000	0.893±0.000	0.787±0.000	0.834±0.000	0.053	0.463
	LR	0.999±0.000	0.996±0.000	0.991±0.000	0.996±0.000	0.993±0.000	0.986±0.000	0.996±0.000	0.388	0.002
	MLP	0.987±0.006	0.901±0.061	0.974±0.026	0.901±0.061	0.930±0.038	0.860±0.076	0.901±0.061	5.933	0.035
	NB	0.562±0.000	0.768±0.000	0.556±0.000	0.768±0.000	0.450±0.000	0.114±0.000	0.768±0.000	0.012	0.008
	RF	0.999±0.000	0.992±0.000	1.000±0.000	0.992±0.000	0.996±0.000	0.992±0.000	0.992±0.000	1.955	0.078
	SVM	0.944±0.000	0.500±0.000	0.472±0.000	0.500±0.000	0.486±0.000	0.000±0.000	0.500±0.000	77.398	14.51
	XGB	0.999±0.000	0.992±0.000	1.000±0.000	0.992±0.000	0.996±0.000	0.992±0.000	0.992±0.000	3.645	0.029
MHD	DT	0.970±0.032	0.794±0.221	0.975±0.029	0.794±0.221	0.808±0.225	0.626±0.436	0.794±0.221	0.017	0.000
	KNN	0.998±0.000	0.999±0.000	0.986±0.000	0.999±0.000	0.992±0.000	0.984±0.000	0.999±0.000	0.020	0.355
	LR	0.998±0.000	0.999±0.000	0.989±0.000	0.999±0.000	0.994±0.000	0.988±0.000	0.999±0.000	0.141	0.003
	MLP	0.998±0.000	0.999±0.000	0.988±0.000	0.999±0.000	0.994±0.000	0.987±0.001	0.999±0.000	311.831	0.520
	NB	0.917±0.000	0.955±0.000	0.734±0.000	0.955±0.000	0.795±0.000	0.597±0.000	0.955±0.000	0.006	0.005
	RF	0.995±0.000	0.964±0.001	0.996±0.000	0.964±0.001	0.980±0.001	0.959±0.001	0.964±0.001	0.365	0.043
	SVM	0.934±0.000	0.555±0.000	0.888±0.000	0.555±0.000	0.581±0.000	0.182±0.000	0.555±0.000	14.811	3.674
	XGB	0.959±0.000	0.719±0.000	0.977±0.000	0.719±0.000	0.793±0.000	0.59±0.000	0.719±0.000	3.092	0.024
OS	DT	0.997±0.000	0.988±0.000	0.824±0.000	0.988±0.000	0.890±0.000	0.779±0.000	0.988±0.000	0.06	0.007
	KNN	0.863±0.000	0.840±0.000	0.515±0.000	0.840±0.000	0.493±0.000	0.050±0.000	0.840±0.000	0.002	1.207
	LR	0.997±0.000	0.996±0.004	0.815±0.007	0.996±0.004	0.885±0.006	0.770±0.011	0.996±0.004	0.050	0.003
	MLP	0.966±0.086	0.486±0.043	0.497±0.000	0.486±0.043	0.490±0.025	0.003±0.003	0.486±0.043	211.455	0.197
	NB	0.325±0.000	0.326±0.000	0.496±0.000	0.326±0.000	0.247±0.000	0.005±0.000	0.326±0.000	0.016	0.020
	RF	0.994±0.004	0.867±0.202	0.747±0.150	0.867±0.202	0.784±0.157	0.568±0.314	0.867±0.202	0.020	0.002
	SVM	0.995±0.000	0.500±0.000	0.497±0.000	0.500±0.000	0.499±0.000	0.000±0.000	0.500±0.000	39.534	8.556
	XGB	0.997±0.000	0.988±0.000	0.824±0.000	0.988±0.000	0.890±0.000	0.779±0.000	0.988±0.000	0.712	0.025
SHD	DT	1.000±0.000	0.877±0.153	0.908±0.147	0.877±0.153	0.890±0.150	0.780±0.299	0.877±0.153	0.044	0.016
	KNN	0.999±0.000	0.658±0.000	1.000±0.000	0.658±0.000	0.740±0.000	0.481±0.000	0.658±0.000	0.055	2.050
	LR	0.997±0.000	0.677±0.000	0.581±0.000	0.677±0.000	0.611±0.000	0.222±0.000	0.677±0.000	0.180	0.005
	MLP	0.991±0.003	0.556±0.150	0.592±0.205	0.556±0.150	0.560±0.152	0.124±0.302	0.556±0.150	149.257	0.230
	NB	0.996±0.000	0.998±0.000	0.616±0.000	0.998±0.000	0.687±0.000	0.375±0.000	0.998±0.000	0.009	0.086
	RF	0.999±0.000	0.735±0.052	1.000±0.000	0.735±0.052	0.816±0.047	0.633±0.094	0.735±0.052	0.868	0.440
	SVM	0.990±0.000	0.500±0.000	0.495±0.000	0.500±0.000	0.497±0.000	0.000±0.000	0.500±0.000	9.767	6.008
	XGB	1.000±0.000	0.946±0.000	0.964±0.000	0.946±0.000	0.954±0.000	0.909±0.000	0.946±0.000	0.595	0.104
PS	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.012	0.002
	KNN	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.025	0.053
	LR	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.066	0.000
	MLP	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	13.383	0.095
	NB	0.978±0.000	0.987±0.000	0.942±0.000	0.987±0.000	0.963±0.000	0.925±0.000	0.987±0.000	0.003	0.002
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.013	0.003
	SVM	0.861±0.000	0.586±0.000	0.928±0.000	0.586±0.000	0.608±0.000	0.256±0.000	0.586±0.000	13.930	4.091
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.000	1.000±0.000	1.850	0.016
SYN	DT	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.998±0.000	0.999±0.000	0.018	0.002
	KNN	0.997±0.000	0.997±0.000	0.996±0.000	0.997±0.000	0.997±0.000	0.993±0.000	0.997±0.000	0.113	9.999
	LR	0.993±0.008	0.994±0.007	0.993±0.009	0.994±0.007	0.993±0.009	0.987±0.017	0.994±0.007	0.053	0.004
	MLP	0.997±0.001	0.997±0.001	0.996±0.001	0.997±0.001	0.997±0.001	0.993±0.001	0.997±0.001	369.283	1.311
	NB	0.993±0.000	0.994±0.000	0.992±0.000	0.994±0.000	0.993±0.000	0.986±0.000	0.994±0.000	0.017	0.045
	RF	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.998±0.000	0.999±0.000	1.196	0.333
	SVM	0.953±0.000	0.945±0.000	0.962±0.000	0.945±0.000	0.951±0.000	0.902±0.000	0.945±0.000	66.502	15.793
	XGB	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	0.998±0.000	0.999±0.000	9.864	0.146
UDP	DT	0.913±0.183	0.955±0.094	0.907±0.197	0.955±0.094	0.884±0.245	0.814±0.392	0.955±0.094	0.036	0.056
	KNN	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.062	196.800
	LR	1.000±0.000	0.999±0.000	1.000±0.000	0.999±0.000	1.000±0.000	0.999±0.000	0.999±0.000	0.399	0.022
	MLP	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	478.953	1.236
	NB	0.998±0.000	0.969±0.000	0.999±0.000	0.969±0.000	0.983±0.000	0.967±0.000	0.969±0.000	0.013	0.340
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.454	1.509
	SVM	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.149	0.025
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	5.069	0.547

TABLE XI
WINDOW APPROACH TEST RESULTS. T IS TIME

Window approach Test results.

Attack	ML	Accuracy	BalancedAcc	Precision	Recall	F1-Score	kappa	ROC	Train-T	Test-T
ACK	DT	0.920±0.229	0.945±0.156	0.900±0.269	0.945±0.156	0.912±0.246	0.882±0.310	0.945±0.156	0.018	0.006
	KNN	0.269±0.000	0.500±0.000	0.134±0.000	0.500±0.000	0.212±0.000	0.000±0.000	0.500±0.000	0.029	1.555
	LR	0.989±0.000	0.992±0.000	0.980±0.000	0.992±0.000	0.986±0.000	0.972±0.000	0.992±0.000	0.084	0.001
	MLP	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	20.468	0.613
	NB	0.269±0.000	0.500±0.000	0.134±0.000	0.500±0.000	0.212±0.000	0.000±0.000	0.500±0.000	0.006	0.011
	RF	0.991±0.000	0.994±0.000	0.984±0.000	0.994±0.000	0.989±0.000	0.978±0.000	0.994±0.000	0.941	0.296
	SVM	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	30.356	6.798
	XGB	0.991±0.000	0.994±0.000	0.985±0.000	0.994±0.000	0.989±0.000	0.978±0.000	0.994±0.000	1.745	0.082
ARP	DT	0.521±0.021	0.503±0.034	0.444±0.127	0.503±0.034	0.454±0.084	0.004±0.071	0.503±0.034	0.680	1.131
	KNN	0.519±0.000	0.478±0.000	0.266±0.000	0.478±0.000	0.341±0.000	0.048±0.000	0.478±0.000	0.306	98.193
	LR	0.615±0.065	0.611±0.092	0.653±0.116	0.611±0.092	0.567±0.103	0.214±0.174	0.611±0.092	77.085	0.397
	MLP	0.513±0.019	0.513±0.019	0.708±0.065	0.513±0.019	0.373±0.063	0.026±0.038	0.513±0.019	24.572	0.091
	NB	0.490±0.000	0.451±0.000	0.259±0.000	0.451±0.000	0.329±0.000	0.105±0.000	0.451±0.000	0.027	4.659
	RF	0.527±0.002	0.485±0.002	0.268±0.000	0.485±0.002	0.345±0.001	0.032±0.004	0.485±0.002	18.588	45.121
	SVM	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	194.755	41.650
	XGB	0.536±0.000	0.494±0.000	0.270±0.000	0.494±0.000	0.349±0.000	0.013±0.000	0.494±0.000	11.910	4.855
BF	DT	0.476±0.039	0.519±0.036	0.633±0.214	0.519±0.036	0.351±0.070	0.036±0.067	0.519±0.036	0.038	0.000
	KNN	0.574±0.000	0.609±0.000	0.758±0.000	0.609±0.000	0.520±0.000	0.202±0.000	0.609±0.000	0.005	0.042
	LR	0.502±0.004	0.543±0.004	0.739±0.001	0.543±0.004	0.402±0.007	0.079±0.007	0.543±0.004	0.242	0.000
	MLP	0.545±0.047	0.578±0.042	0.677±0.062	0.578±0.042	0.486±0.072	0.146±0.080	0.578±0.042	82.584	0.009
	NB	0.745±0.000	0.766±0.000	0.821±0.000	0.766±0.000	0.738±0.000	0.509±0.000	0.766±0.000	0.006	0.000
	RF	0.457±0.000	0.502±0.000	0.728±0.000	0.502±0.000	0.317±0.000	0.003±0.000	0.502±0.000	2.283	0.024
	SVM	0.455±0.000	0.500±0.000	0.228±0.000	0.500±0.000	0.313±0.000	0.000±0.000	0.500±0.000	7.066	0.193
	XGB	0.455±0.000	0.500±0.000	0.228±0.000	0.500±0.000	0.313±0.000	0.000±0.000	0.500±0.000	2.904	0.005
HTTP	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.032	0.030
	KNN	0.473±0.000	0.520±0.000	0.730±0.000	0.520±0.000	0.355±0.000	0.037±0.000	0.520±0.000	0.046	6.081
	LR	0.450±0.000	0.500±0.000	0.225±0.000	0.500±0.000	0.310±0.000	0.000±0.000	0.500±0.000	0.335	0.013
	MLP	0.527±0.022	0.527±0.022	0.757±0.006	0.527±0.022	0.388±0.044	0.053±0.043	0.527±0.022	4.813	0.039
	NB	0.473±0.000	0.521±0.000	0.714±0.000	0.521±0.000	0.356±0.000	0.037±0.000	0.521±0.000	0.011	0.121
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.803	1.537
	SVM	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	78.257	21.362
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	2.135	0.219
MHD	DT	0.982±0.017	0.770±0.218	0.986±0.013	0.770±0.218	0.795±0.204	0.595±0.403	0.770±0.218	0.020	0.003
	KNN	0.999±0.000	0.999±0.000	0.987±0.000	0.999±0.000	0.993±0.000	0.986±0.000	0.999±0.000	0.025	0.475
	LR	1.000±0.000	1.000±0.000	0.997±0.000	1.000±0.000	0.998±0.000	0.997±0.000	1.000±0.000	0.230	0.002
	MLP	0.999±0.000	1.000±0.000	0.990±0.004	1.000±0.000	0.995±0.002	0.990±0.005	1.000±0.000	316.916	0.821
	NB	0.964±0.000	0.981±0.000	0.760±0.000	0.981±0.000	0.833±0.000	0.668±0.000	0.981±0.000	0.007	0.008
	RF	0.998±0.001	0.973±0.007	0.998±0.000	0.973±0.007	0.985±0.004	0.970±0.008	0.973±0.007	0.480	0.075
	SVM	0.959±0.000	0.554±0.000	0.979±0.000	0.554±0.000	0.587±0.000	0.189±0.000	0.554±0.000	14.813	5.792
	XGB	0.979±0.000	0.732±0.000	0.987±0.000	0.732±0.000	0.811±0.000	0.623±0.000	0.732±0.000	4.638	0.067
OS	DT	0.597±0.012	0.757±0.001	0.616±0.000	0.757±0.001	0.542±0.008	0.223±0.006	0.757±0.001	0.087	0.031
	KNN	0.785±0.000	0.464±0.000	0.452±0.000	0.464±0.000	0.457±0.000	0.082±0.000	0.464±0.000	0.006	7.621
	LR	0.774±0.044	0.862±0.025	0.681±0.023	0.862±0.025	0.691±0.041	0.423±0.065	0.862±0.025	0.073	0.011
	MLP	0.486±0.055	0.486±0.055	0.348±0.208	0.486±0.055	0.332±0.033	0.028±0.111	0.486±0.055	183.464	0.379
	NB	0.724±0.000	0.449±0.000	0.459±0.000	0.449±0.000	0.453±0.000	0.090±0.000	0.449±0.000	0.017	0.185
	RF	0.826±0.111	0.639±0.146	0.698±0.127	0.639±0.146	0.618±0.119	0.266±0.227	0.639±0.146	0.031	0.066
	SVM	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	37.010	17.52
	XGB	0.610±0.000	0.758±0.000	0.616±0.000	0.758±0.000	0.551±0.000	0.230±0.000	0.758±0.000	1.039	0.259
SHD	DT	0.999±0.000	0.897±0.098	0.897±0.002	0.897±0.098	0.889±0.072	0.778±0.144	0.897±0.098	0.059	0.022
	KNN	0.998±0.000	0.676±0.000	0.968±0.000	0.676±0.000	0.756±0.000	0.512±0.000	0.676±0.000	0.078	1.901
	LR	0.989±0.000	0.971±0.000	0.583±0.000	0.971±0.000	0.639±0.000	0.281±0.000	0.971±0.000	0.288	0.007
	MLP	0.989±0.007	0.733±0.163	0.972±0.027	0.733±0.163	0.781±0.158	0.564±0.314	0.733±0.163	150.392	0.233
	NB	0.995±0.000	0.997±0.000	0.647±0.000	0.997±0.000	0.726±0.000	0.453±0.000	0.997±0.000	0.010	0.113
	RF	0.998±0.000	0.621±0.030	0.910±0.071	0.621±0.030	0.684±0.031	0.368±0.061	0.621±0.030	1.147	0.543
	SVM	0.979±0.000	0.500±0.000	0.490±0.000	0.500±0.000	0.495±0.000	0.000±0.000	0.500±0.000	8.656	5.830
	XGB	0.999±0.000	0.928±0.000	0.896±0.000	0.928±0.000	0.912±0.000	0.823±0.000	0.928±0.000	0.976	0.176
PS	DT	0.898±0.068	0.903±0.067	0.916±0.054	0.903±0.067	0.896±0.070	0.797±0.136	0.903±0.067	0.014	0.000
	KNN	0.997±0.000	0.997±0.000	0.997±0.000	0.997±0.000	0.997±0.000	0.994±0.000	0.997±0.000	0.020	0.042
	LR	0.997±0.000	0.997±0.000	0.997±0.000	0.997±0.000	0.997±0.000	0.994±0.000	0.997±0.000	0.069	0.000
	MLP	0.997±0.000	0.997±0.000	0.997±0.000	0.997±0.000	0.997±0.000	0.994±0.000	0.997±0.000	8.699	0.074
	NB	0.972±0.000	0.974±0.000	0.971±0.000	0.974±0.000	0.971±0.000	0.943±0.000	0.974±0.000	0.000	0.002
	RF	0.863±0.000	0.874±0.000	0.884±0.000	0.874±0.000	0.862±0.000	0.730±0.000	0.874±0.000	0.010	0.003
	SVM	0.592±0.000	0.554±0.000	0.783±0.000	0.554±0.000	0.461±0.000	0.116±0.000	0.554±0.000	13.858	4.933
	XGB	0.945±0.000	0.949±0.000	0.946±0.000	0.949±0.000	0.945±0.000	0.890±0.000	0.949±0.000	1.438	0.014
SYN	DT	0.943±0.181	0.950±0.158	0.971±0.091	0.950±0.158	0.930±0.221	0.900±0.315	0.950±0.158	0.016	0.180
	KNN	0.425±0.000	0.500±0.000	0.212±0.000	0.500±0.000	0.298±0.000	0.000±0.000	0.500±0.000	0.104	66.594
	LR	0.982±0.053	0.980±0.062	0.988±0.036	0.980±0.062	0.981±0.059	0.962±0.114	0.980±0.062	0.056	0.069
	MLP	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	172.875	0.613
	NB	0.998±0.000	0.998±0.000	0.998±0.000	0.998±0.000	0.998±0.000	0.996±0.000	0.998±0.000	0.012	0.838
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.170	8.297
	SVM	0.500±0.000	0.500±0.000	0.250±0.000	0.500±0.000	0.333±0.000	0.000±0.000	0.500±0.000	71.319	16.116
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	5.706	1.489
UDP	DT	0.454±0.323	0.695±0.125	0.560±0.121	0.695±0.125	0.359±0.246	0.128±0.262	0.695±0.125	0.032	0.288
	KNN	0.255±0.000	0.620±0.000	0.512±0.000	0.620±0.000	0.218±0.000	0.012±0.000	0.620±0.000	0.057	944.879
	LR	0.255±0.000	0.620±0.000	0.512±0.000	0.620±0.000	0.218±0.000	0.012±0.000	0.620±0.000	0.380	0.142
	MLP	0.838±0.188	0.838±0.188	0.904±0.104	0.838±0.188	0.812±0.221	0.676±0.376	0.838±0.188	239.447	0.882
	NB	0.996±0.000	0.897±0.000	0.998±0.000	0.897±0.000	0.941±0.000	0.883±0.000	0.897±0.000	0.016	1.911
	RF	0.514±0.335	0.752±0.171	0.540±0.036	0.752±0.171	0.377±0.206	0.111±0.128	0.752±0.171	0.443	7.048
	SVM	0.620±0.000	0.620±0.000	0.784±0.000	0.620±0.000	0.556±0.000	0.240±0.000	0.620±0.000	0.140	0.019
	XGB	0.255±0.000	0.620±0.000	0.512±0.000	0.620±0.000	0.218±0.000	0.012±0.000	0.620±0.000	16.237	4.475

TABLE XII
FLOW APPROACH CV RESULTS. T IS TIME

Attack	ML	Accuracy	BalancedAcc	Precision	Recall	F1-Score	kappa	ROC	Train-T	Test-T
ACK	DT	0.894±0.015	0.910±0.013	0.897±0.015	0.910±0.013	0.893±0.016	0.789±0.030	0.910±0.013	0.000	0.000
	KNN	0.588±0.020	0.498±0.002	0.294±0.010	0.498±0.002	0.370±0.008	0.004±0.004	0.498±0.002	0.002	0.010
	LR	0.896±0.017	0.911±0.015	0.898±0.015	0.911±0.015	0.894±0.017	0.791±0.033	0.911±0.015	0.043	0.002
	MLP	0.896±0.012	0.912±0.009	0.898±0.012	0.912±0.009	0.895±0.012	0.792±0.024	0.912±0.009	1.808	0.003
	NB	0.799±0.028	0.830±0.021	0.836±0.020	0.830±0.021	0.799±0.028	0.613±0.050	0.830±0.021	0.001	0.002
	RF	0.894±0.011	0.910±0.007	0.897±0.014	0.910±0.007	0.893±0.011	0.788±0.022	0.910±0.007	0.330	0.025
	SVM	0.895±0.019	0.911±0.017	0.898±0.016	0.911±0.017	0.894±0.019	0.790±0.037	0.911±0.017	0.031	0.019
	XGB	0.894±0.027	0.910±0.023	0.898±0.023	0.910±0.023	0.893±0.027	0.789±0.053	0.910±0.023	0.982	0.001
ARP	DT	0.677±0.064	0.662±0.066	0.673±0.074	0.662±0.066	0.659±0.066	0.325±0.132	0.662±0.066	0.007	0.000
	KNN	0.673±0.047	0.639±0.058	0.669±0.061	0.639±0.058	0.636±0.062	0.290±0.116	0.639±0.058	0.002	0.002
	LR	0.678±0.059	0.660±0.060	0.667±0.065	0.660±0.060	0.661±0.062	0.325±0.124	0.660±0.060	1.347	0.000
	MLP	0.672±0.040	0.646±0.035	0.666±0.034	0.646±0.035	0.641±0.040	0.298±0.069	0.646±0.035	12.560	0.012
	NB	0.583±0.029	0.517±0.023	0.546±0.054	0.517±0.023	0.463±0.035	0.038±0.051	0.517±0.023	0.001	0.001
	RF	0.705±0.037	0.679±0.035	0.695±0.035	0.679±0.035	0.682±0.035	0.368±0.068	0.679±0.035	0.081	0.007
	SVM	0.584±0.057	0.498±0.009	0.369±0.127	0.498±0.009	0.375±0.027	0.006±0.023	0.498±0.009	0.014	0.006
	XGB	0.705±0.048	0.679±0.059	0.700±0.064	0.679±0.059	0.681±0.059	0.368±0.117	0.679±0.059	1.387	0.004
BF	DT	0.926±0.018	0.853±0.044	0.854±0.039	0.853±0.044	0.851±0.036	0.703±0.073	0.853±0.044	0.004	0.000
	KNN	0.890±0.030	0.782±0.050	0.788±0.074	0.782±0.050	0.781±0.062	0.564±0.122	0.782±0.050	0.002	0.002
	LR	0.838±0.037	0.494±0.007	0.423±0.016	0.494±0.007	0.456±0.011	-0.02±0.020	0.494±0.007	0.006	0.000
	MLP	0.858±0.026	0.621±0.052	0.732±0.114	0.621±0.052	0.640±0.062	0.295±0.117	0.621±0.052	57.888	0.066
	NB	0.817±0.040	0.501±0.029	0.555±0.124	0.501±0.029	0.486±0.034	0.008±0.080	0.501±0.029	0.001	0.000
	RF	0.943±0.019	0.863±0.035	0.908±0.034	0.863±0.035	0.883±0.032	0.766±0.064	0.863±0.035	0.299	0.021
	SVM	0.851±0.027	0.507±0.010	0.625±0.265	0.507±0.010	0.474±0.023	0.025±0.032	0.507±0.010	0.063	0.027
	XGB	0.948±0.013	0.878±0.044	0.915±0.041	0.878±0.044	0.892±0.030	0.785±0.059	0.878±0.044	1.580	0.006
HTTP	DT	0.964±0.010	0.961±0.009	0.962±0.012	0.961±0.009	0.962±0.010	0.923±0.021	0.961±0.009	0.004	0.000
	KNN	0.897±0.016	0.885±0.019	0.898±0.018	0.885±0.019	0.890±0.018	0.781±0.036	0.885±0.019	0.002	0.034
	LR	0.939±0.010	0.922±0.013	0.953±0.008	0.922±0.013	0.934±0.011	0.868±0.022	0.922±0.013	0.018	0.002
	MLP	0.796±0.017	0.749±0.015	0.838±0.017	0.749±0.015	0.761±0.017	0.538±0.031	0.749±0.015	2.353	0.007
	NB	0.831±0.017	0.788±0.019	0.879±0.014	0.788±0.019	0.804±0.020	0.619±0.038	0.788±0.019	0.002	0.000
	RF	0.973±0.008	0.970±0.009	0.973±0.008	0.970±0.009	0.971±0.009	0.942±0.017	0.970±0.009	0.048	0.003
	SVM	0.611±0.012	0.506±0.004	0.662±0.160	0.506±0.004	0.392±0.009	0.013±0.011	0.506±0.004	0.227	0.096
	XGB	0.970±0.007	0.966±0.009	0.971±0.006	0.966±0.009	0.968±0.007	0.937±0.014	0.966±0.009	1.859	0.009
PS	DT	0.999±0.002	0.990±0.021	0.999±0.001	0.990±0.021	0.995±0.012	0.989±0.024	0.990±0.021	0.001	0.000
	KNN	0.993±0.006	0.941±0.043	0.996±0.003	0.941±0.043	0.966±0.026	0.932±0.052	0.941±0.043	0.003	0.011
	LR	0.998±0.005	0.983±0.039	0.999±0.002	0.983±0.039	0.990±0.023	0.980±0.046	0.983±0.039	0.003	0.000
	MLP	0.981±0.026	0.932±0.053	0.924±0.083	0.932±0.053	0.924±0.064	0.849±0.128	0.932±0.053	3.001	0.008
	NB	0.993±0.004	0.940±0.054	0.986±0.035	0.940±0.054	0.959±0.039	0.918±0.079	0.940±0.054	0.001	0.000
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.053	0.006
	SVM	0.944±0.015	0.517±0.022	0.672±0.258	0.517±0.022	0.516±0.041	0.058±0.077	0.517±0.022	0.013	0.005
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.854	0.005
SYN	DT	1.000±0.000	0.967±0.081	0.995±0.014	0.967±0.081	0.977±0.054	0.954±0.109	0.967±0.081	0.047	0.000
	KNN	0.999±0.000	0.885±0.068	1.000±0.000	0.885±0.068	0.932±0.044	0.863±0.088	0.885±0.068	0.001	0.469
	LR	1.000±0.000	0.954±0.063	0.973±0.058	0.954±0.063	0.960±0.050	0.920±0.100	0.954±0.063	0.446	0.000
	MLP	1.000±0.000	0.980±0.037	0.975±0.060	0.980±0.037	0.975±0.042	0.950±0.084	0.980±0.037	21.141	0.043
	NB	0.996±0.001	0.919±0.079	0.679±0.044	0.919±0.079	0.746±0.047	0.493±0.093	0.919±0.079	0.007	0.009
	RF	1.000±0.000	0.985±0.032	1.000±0.000	0.985±0.032	0.992±0.017	0.984±0.035	0.985±0.032	0.907	0.028
	SVM	0.998±0.001	0.500±0.000	0.499±0.001	0.500±0.000	0.499±0.000	0.000±0.000	0.500±0.000	42.601	10.340
	XGB	1.000±0.000	0.960±0.053	1.000±0.000	0.960±0.053	0.977±0.031	0.955±0.062	0.960±0.053	0.678	0.013
UDP	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.005	0.000
	KNN	0.993±0.006	0.996±0.003	0.883±0.088	0.996±0.003	0.926±0.063	0.853±0.125	0.996±0.003	0.002	0.002
	LR	0.997±0.007	0.998±0.003	0.959±0.074	0.998±0.003	0.975±0.048	0.949±0.095	0.998±0.003	0.037	0.000
	MLP	0.995±0.008	0.966±0.071	0.949±0.121	0.966±0.071	0.945±0.104	0.890±0.207	0.966±0.071	25.329	0.011
	NB	0.889±0.026	0.943±0.013	0.598±0.053	0.943±0.013	0.628±0.075	0.288±0.134	0.943±0.013	0.001	0.000
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.012	0.002
	SVM	0.973±0.010	0.500±0.000	0.487±0.005	0.500±0.000	0.493±0.002	0.000±0.000	0.500±0.000	0.038	0.017
	XGB	0.999±0.003	0.999±0.001	0.983±0.037	0.999±0.001	0.990±0.021	0.981±0.042	0.999±0.001	0.414	0.004

Flow approach CV results.

TABLE XIII
FLOW APPROACH SESSION RESULTS. T IS TIME

Attack	ML	Accuracy	BalancedAcc	Precision	Recall	F1-Score	kappa	ROC	Train-T	Test-T
ACK	DT	0.997±0.000	0.939±0.001	0.995±0.001	0.939±0.001	0.966±0.001	0.931±0.002	0.939±0.001	0.001	0.000
	KNN	0.022±0.000	0.500±0.000	0.511±0.000	0.500±0.000	0.022±0.000	0.000±0.000	0.500±0.000	0.002	0.295
	LR	0.997±0.000	0.940±0.000	0.995±0.000	0.940±0.000	0.966±0.000	0.931±0.000	0.940±0.000	0.060	0.001
	MLP	0.997±0.000	0.940±0.000	0.995±0.000	0.940±0.000	0.966±0.000	0.931±0.000	0.940±0.000	2.264	0.014
	NB	0.992±0.000	0.816±0.000	0.996±0.000	0.816±0.000	0.886±0.000	0.771±0.000	0.816±0.000	0.001	0.004
	RF	0.997±0.000	0.940±0.000	0.995±0.000	0.940±0.000	0.966±0.000	0.931±0.000	0.940±0.000	0.340	0.119
	SVM	0.997±0.000	0.940±0.000	0.995±0.000	0.940±0.000	0.966±0.000	0.931±0.000	0.940±0.000	0.035	0.407
	XGB	0.997±0.000	0.940±0.000	0.999±0.000	0.940±0.000	0.967±0.000	0.935±0.000	0.940±0.000	1.000	0.020
ARP	DT	0.702±0.018	0.721±0.017	0.707±0.016	0.721±0.017	0.698±0.017	0.409±0.032	0.721±0.017	0.011	0.000
	KNN	0.721±0.000	0.719±0.000	0.708±0.000	0.719±0.000	0.710±0.000	0.423±0.000	0.719±0.000	0.002	0.004
	LR	0.675±0.000	0.704±0.000	0.693±0.000	0.704±0.000	0.674±0.000	0.370±0.000	0.704±0.000	1.297	0.000
	MLP	0.694±0.040	0.698±0.033	0.688±0.034	0.698±0.033	0.686±0.037	0.379±0.070	0.698±0.033	16.508	0.035
	NB	0.554±0.000	0.464±0.000	0.437±0.000	0.464±0.000	0.428±0.000	-0.082±0.00	0.464±0.000	0.001	0.000
	RF	0.730±0.009	0.753±0.007	0.736±0.006	0.753±0.007	0.727±0.008	0.466±0.015	0.753±0.007	0.091	0.010
	SVM	0.643±0.000	0.518±0.000	0.819±0.000	0.518±0.000	0.424±0.000	0.044±0.000	0.518±0.000	0.022	0.025
	XGB	0.734±0.000	0.754±0.000	0.738±0.000	0.754±0.000	0.731±0.000	0.472±0.000	0.754±0.000	1.282	0.006
BF	DT	0.886±0.007	0.752±0.009	0.787±0.017	0.752±0.009	0.767±0.012	0.535±0.024	0.752±0.009	0.005	0.000
	KNN	0.869±0.000	0.709±0.000	0.749±0.000	0.709±0.000	0.726±0.000	0.452±0.000	0.709±0.000	0.002	0.008
	LR	0.844±0.000	0.498±0.000	0.423±0.000	0.498±0.000	0.458±0.000	-0.005±0.00	0.498±0.000	0.009	0.001
	MLP	0.846±0.015	0.558±0.035	0.682±0.073	0.558±0.035	0.564±0.052	0.159±0.092	0.558±0.035	58.708	0.267
	NB	0.829±0.000	0.517±0.000	0.562±0.000	0.517±0.000	0.506±0.000	0.050±0.000	0.517±0.000	0.001	0.000
	RF	0.916±0.001	0.767±0.003	0.884±0.004	0.767±0.003	0.810±0.003	0.622±0.006	0.767±0.003	0.340	0.038
	SVM	0.848±0.000	0.510±0.000	0.758±0.000	0.510±0.000	0.480±0.000	0.033±0.000	0.510±0.000	0.104	0.205
	XGB	0.912±0.000	0.762±0.000	0.873±0.000	0.762±0.000	0.803±0.000	0.608±0.000	0.762±0.000	4.756	0.024
HTTP	DT	0.934±0.075	0.936±0.068	0.936±0.063	0.936±0.068	0.933±0.076	0.868±0.144	0.936±0.068	0.005	0.000
	KNN	0.843±0.000	0.842±0.000	0.839±0.000	0.842±0.000	0.840±0.000	0.681±0.000	0.842±0.000	0.002	0.047
	LR	0.923±0.000	0.911±0.000	0.940±0.000	0.911±0.000	0.919±0.000	0.839±0.000	0.911±0.000	0.023	0.000
	MLP	0.810±0.007	0.784±0.008	0.859±0.010	0.784±0.008	0.791±0.009	0.594±0.016	0.784±0.008	2.555	0.011
	NB	0.779±0.000	0.745±0.000	0.860±0.000	0.745±0.000	0.747±0.000	0.522±0.000	0.745±0.000	0.000	0.000
	RF	0.963±0.001	0.964±0.001	0.961±0.002	0.964±0.001	0.962±0.001	0.924±0.003	0.964±0.001	0.062	0.001
	SVM	0.569±0.000	0.503±0.000	0.784±0.000	0.503±0.000	0.369±0.000	0.007±0.000	0.503±0.000	0.367	0.120
	XGB	0.942±0.000	0.942±0.000	0.940±0.000	0.942±0.000	0.941±0.000	0.882±0.000	0.942±0.000	2.052	0.010
PS	DT	0.998±0.003	0.992±0.010	0.996±0.010	0.992±0.010	0.994±0.010	0.988±0.020	0.992±0.010	0.002	0.000
	KNN	0.971±0.000	0.901±0.000	0.923±0.000	0.901±0.000	0.912±0.000	0.824±0.000	0.901±0.000	0.003	0.046
	LR	1.000±0.001	0.998±0.003	1.000±0.000	0.998±0.003	0.999±0.002	0.997±0.004	0.998±0.003	0.003	0.001
	MLP	0.949±0.050	0.961±0.030	0.863±0.121	0.961±0.030	0.891±0.095	0.787±0.184	0.961±0.030	3.235	0.020
	NB	0.989±0.000	0.942±0.000	0.994±0.000	0.942±0.000	0.966±0.000	0.932±0.000	0.942±0.000	0.002	0.001
	RF	0.999±0.000	0.996±0.002	1.000±0.000	0.996±0.002	0.998±0.001	0.995±0.002	0.996±0.002	0.043	0.008
	SVM	0.906±0.000	0.500±0.000	0.453±0.000	0.500±0.000	0.475±0.000	0.000±0.000	0.500±0.000	0.023	0.039
	XGB	0.999±0.000	0.995±0.000	0.999±0.000	0.995±0.000	0.997±0.000	0.995±0.000	0.995±0.000	1.632	0.009
SYN	DT	1.000±0.000	0.983±0.011	0.994±0.006	0.983±0.011	0.988±0.006	0.977±0.012	0.983±0.011	0.059	0.003
	KNN	0.998±0.000	0.879±0.000	0.953±0.000	0.879±0.000	0.913±0.000	0.825±0.000	0.879±0.000	0.003	1.996
	LR	0.998±0.000	0.879±0.000	0.925±0.000	0.879±0.000	0.901±0.000	0.801±0.000	0.879±0.000	0.530	0.002
	MLP	0.998±0.001	0.906±0.015	0.923±0.046	0.906±0.015	0.913±0.019	0.825±0.039	0.906±0.015	21.131	0.177
	NB	0.995±0.000	0.899±0.000	0.785±0.000	0.899±0.000	0.832±0.000	0.664±0.000	0.899±0.000	0.014	0.017
	RF	1.000±0.000	0.989±0.006	0.994±0.004	0.989±0.006	0.992±0.002	0.983±0.005	0.989±0.006	1.040	0.099
	SVM	0.994±0.000	0.500±0.000	0.497±0.000	0.500±0.000	0.498±0.000	0.000±0.000	0.500±0.000	65.403	50.148
	XGB	1.000±0.000	0.973±0.000	1.000±0.000	0.973±0.000	0.986±0.000	0.972±0.000	0.973±0.000	0.728	0.032
UDP	DT	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.001	0.000
	KNN	0.496±0.000	0.576±0.000	0.723±0.000	0.576±0.000	0.441±0.000	0.127±0.000	0.576±0.000	0.001	0.003
	LR	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.165	0.000
	MLP	0.497±0.058	0.577±0.049	0.673±0.166	0.577±0.049	0.437±0.084	0.130±0.086	0.577±0.049	36.424	0.027
	NB	0.687±0.000	0.718±0.000	0.723±0.000	0.718±0.000	0.687±0.000	0.402±0.000	0.718±0.000	0.000	0.002
	RF	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.012	0.002
	SVM	0.406±0.000	0.500±0.000	0.203±0.000	0.500±0.000	0.289±0.000	0.000±0.000	0.500±0.000	0.057	0.054
	XGB	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.397	0.003

Flow approach Session results.

TABLE XIV
FLOW APPROACH TEST RESULTS. T IS TIME

Attack	ML	Accuracy	BalancedAcc	Precision	Recall	F1-Score	kappa	ROC	Train-T	Test-T
ACK	DT	0.116±0.000	0.443±0.000	0.375±0.000	0.443±0.000	0.113±0.000	-0.024±0.00	0.443±0.000	0.000	0.000
	KNN	0.102±0.000	0.500±0.000	0.051±0.000	0.500±0.000	0.093±0.000	0.000±0.000	0.500±0.000	0.003	0.021
	LR	0.116±0.000	0.443±0.000	0.375±0.000	0.443±0.000	0.113±0.000	-0.024±0.00	0.443±0.000	0.044	0.000
	MLP	0.116±0.000	0.443±0.000	0.375±0.000	0.443±0.000	0.113±0.000	-0.024±0.00	0.443±0.000	2.851	0.002
	NB	0.118±0.000	0.444±0.000	0.383±0.000	0.444±0.000	0.115±0.000	-0.024±0.00	0.444±0.000	0.001	0.001
	RF	0.116±0.000	0.443±0.000	0.375±0.000	0.443±0.000	0.113±0.000	-0.024±0.00	0.443±0.000	0.357	0.036
	SVM	0.116±0.000	0.443±0.000	0.375±0.000	0.443±0.000	0.113±0.000	-0.024±0.00	0.443±0.000	0.024	0.022
	XGB	0.116±0.000	0.443±0.000	0.375±0.000	0.443±0.000	0.113±0.000	-0.024±0.00	0.443±0.000	0.670	0.005
ARP	DT	0.670±0.047	0.457±0.032	0.357±0.007	0.457±0.032	0.401±0.017	-0.108±0.08	0.457±0.032	0.009	0.000
	KNN	0.782±0.000	0.597±0.000	0.848±0.000	0.597±0.000	0.599±0.000	0.258±0.000	0.597±0.000	0.002	0.002
	LR	0.462±0.000	0.331±0.000	0.342±0.000	0.331±0.000	0.336±0.000	-0.326±0.00	0.331±0.000	1.491	0.000
	MLP	0.832±0.044	0.739±0.065	0.808±0.068	0.739±0.065	0.759±0.065	0.524±0.128	0.739±0.065	7.357	0.009
	NB	0.787±0.000	0.733±0.000	0.728±0.000	0.733±0.000	0.730±0.000	0.460±0.000	0.733±0.000	0.002	0.000
	RF	0.728±0.012	0.496±0.008	0.366±0.002	0.496±0.008	0.421±0.004	-0.01±0.023	0.496±0.008	0.087	0.010
	SVM	0.267±0.000	0.500±0.000	0.133±0.000	0.500±0.000	0.211±0.000	0.000±0.000	0.500±0.000	0.013	0.007
	XGB	0.636±0.000	0.433±0.000	0.352±0.000	0.433±0.000	0.389±0.000	-0.167±0.00	0.433±0.000	1.302	0.002
BF	DT	0.521±0.042	0.553±0.039	0.686±0.084	0.553±0.039	0.433±0.077	0.099±0.073	0.553±0.039	0.005	0.000
	KNN	0.463±0.000	0.496±0.000	0.480±0.000	0.496±0.000	0.359±0.000	-0.008±0.00	0.496±0.000	0.002	0.001
	LR	0.463±0.000	0.500±0.000	0.231±0.000	0.500±0.000	0.316±0.000	0.000±0.000	0.500±0.000	0.010	0.000
	MLP	0.548±0.164	0.579±0.152	0.566±0.298	0.579±0.152	0.443±0.209	0.155±0.305	0.579±0.152	29.691	0.013
	NB	0.701±0.000	0.722±0.000	0.804±0.000	0.722±0.000	0.686±0.000	0.425±0.000	0.722±0.000	0.000	0.000
	RF	0.464±0.037	0.499±0.035	0.349±0.200	0.499±0.035	0.343±0.063	-0.002±0.07	0.499±0.035	0.329	0.017
	SVM	0.463±0.000	0.500±0.000	0.231±0.000	0.500±0.000	0.316±0.000	0.000±0.000	0.500±0.000	0.053	0.003
	XGB	0.567±0.000	0.597±0.000	0.758±0.000	0.597±0.000	0.503±0.000	0.183±0.000	0.597±0.000	1.888	0.002
HTTP	DT	0.747±0.086	0.853±0.050	0.687±0.044	0.853±0.050	0.681±0.080	0.418±0.124	0.853±0.050	0.000	0.006
	KNN	0.614±0.000	0.775±0.000	0.633±0.000	0.775±0.000	0.566±0.000	0.256±0.000	0.775±0.000	0.000	0.192
	LR	0.957±0.000	0.975±0.000	0.882±0.000	0.975±0.000	0.920±0.000	0.841±0.000	0.975±0.000	0.020	0.003
	MLP	0.556±0.015	0.326±0.008	0.401±0.002	0.326±0.008	0.359±0.006	-0.23±0.005	0.326±0.008	1.241	0.309
	NB	0.985±0.000	0.991±0.000	0.952±0.000	0.991±0.000	0.971±0.000	0.941±0.000	0.991±0.000	0.000	0.014
	RF	0.651±0.003	0.797±0.002	0.643±0.001	0.797±0.002	0.596±0.003	0.291±0.003	0.797±0.002	0.060	0.044
	SVM	0.860±0.000	0.500±0.000	0.430±0.000	0.500±0.000	0.462±0.000	0.000±0.000	0.500±0.000	0.141	4.179
	XGB	0.719±0.000	0.837±0.000	0.666±0.000	0.837±0.000	0.652±0.000	0.366±0.000	0.837±0.000	1.869	0.176
PS	DT	0.989±0.021	0.962±0.015	0.948±0.101	0.962±0.015	0.949±0.077	0.898±0.151	0.962±0.015	0.000	0.002
	KNN	0.948±0.000	0.713±0.000	0.681±0.000	0.713±0.000	0.696±0.000	0.391±0.000	0.713±0.000	0.005	0.113
	LR	0.999±0.000	0.994±0.000	0.994±0.000	0.994±0.000	0.994±0.000	0.987±0.000	0.994±0.000	0.007	0.000
	MLP	0.904±0.112	0.753±0.092	0.672±0.076	0.753±0.092	0.685±0.110	0.394±0.170	0.753±0.092	3.587	0.041
	NB	0.964±0.000	0.710±0.000	0.780±0.000	0.710±0.000	0.739±0.000	0.478±0.000	0.710±0.000	0.002	0.002
	RF	0.997±0.000	0.969±0.000	0.992±0.000	0.969±0.000	0.980±0.000	0.960±0.000	0.969±0.000	0.044	0.012
	SVM	0.960±0.000	0.530±0.000	0.793±0.000	0.530±0.000	0.546±0.000	0.106±0.000	0.530±0.000	0.033	0.094
	XGB	0.997±0.000	0.969±0.000	0.992±0.000	0.969±0.000	0.980±0.000	0.960±0.000	0.969±0.000	1.531	0.004
SYN	DT	0.185±0.288	0.592±0.144	0.450±0.158	0.592±0.144	0.127±0.141	0.002±0.005	0.592±0.144	0.067	0.084
	KNN	0.580±0.000	0.510±0.000	0.500±0.000	0.510±0.000	0.367±0.000	0.000±0.000	0.510±0.000	0.005	18.839
	LR	0.731±0.000	0.765±0.000	0.500±0.000	0.765±0.000	0.422±0.000	0.000±0.000	0.765±0.000	0.918	0.054
	MLP	0.333±0.333	0.548±0.156	0.498±0.004	0.548±0.156	0.208±0.188	0.000±0.000	0.548±0.156	7.714	1.842
	NB	0.592±0.000	0.676±0.000	0.500±0.000	0.676±0.000	0.372±0.000	0.000±0.000	0.676±0.000	0.017	0.505
	RF	0.000±0.000	0.500±0.000	0.000±0.000	0.500±0.000	0.000±0.000	0.000±0.000	0.500±0.000	1.063	2.673
	SVM	1.000±0.000	0.500±0.000	0.500±0.000	0.500±0.000	0.500±0.000	0.000±0.000	0.500±0.000	30.999	404.215
	XGB	0.123±0.000	0.562±0.000	0.500±0.000	0.562±0.000	0.110±0.000	0.000±0.000	0.562±0.000	0.681	0.338
UDP	DT	0.001±0.000	0.363±0.000	0.000±0.000	0.363±0.000	0.001±0.000	-0.001±0.00	0.363±0.000	0.001	0.028
	KNN	0.001±0.000	0.500±0.000	0.001±0.000	0.500±0.000	0.001±0.000	0.000±0.000	0.500±0.000	0.000	1.290
	LR	0.001±0.000	0.500±0.000	0.001±0.000	0.500±0.000	0.001±0.000	0.000±0.000	0.500±0.000	0.140	0.009
	MLP	0.001±0.000	0.500±0.000	0.001±0.000	0.500±0.000	0.001±0.000	0.000±0.000	0.500±0.000	29.840	4.562
	NB	0.001±0.000	0.499±0.000	0.001±0.000	0.499±0.000	0.001±0.000	-0.00±0.000	0.499±0.000	0.002	0.126
	RF	0.001±0.000	0.418±0.071	0.001±0.000	0.418±0.071	0.001±0.000	-0.00±0.000	0.418±0.071	0.010	0.193
	SVM	0.001±0.000	0.500±0.000	0.001±0.000	0.500±0.000	0.001±0.000	0.000±0.000	0.500±0.000	0.027	12.958
	XGB	0.001±0.000	0.363±0.000	0.000±0.000	0.363±0.000	0.001±0.000	-0.001±0.00	0.363±0.000	0.392	0.207

Flow approach Test results.