Comprehensive Feature Selection for Machine Learning-Based Intrusion Detection in Healthcare IoMT Networks

Muaan Ur Rehman[1][a], Rajesh Kalakoti[1][b] and Hayretdin Bahşi[1,2][c]
[1]*Department of Software Science, Tallinn University of Technology, Tallinn, Estonia*
[2]*School of Informatics, Computing, and Cyber Systems, Northern Arizona University, U.S.A.*
{*muaan.ur, rajesh.kalakoti, hayretdin.bahsi*}*@taltech.ee*

Keywords: Feature Selection, Intrusion Detection, Machine Learning, Internet of Medical Things.

Abstract: The rapid growth of the Internet of Medical Things (IoMT) has increased the vulnerability of healthcare networks to cyberattacks. While Machine learning (ML) techniques can effectively detect these threats, their success depends on the quality and quantity of features used for training to improve detection efficiency in IoMT environments, which are typically resource-constrained. In this paper, we aim to identify the best-performing feature sets for IoMT networks, as measured by classification performance metrics such as F1-score and accuracy, while considering the trade-offs between resource requirements and detection effectiveness. We applied an ML workflow that benchmarks various filter-based feature selection methods for ML-based intrusion detection. To test and train our binary and multi-class models, we used two well-developed IoMT datasets (CICIoMT2024 and IoMT-TrafficData). We applied filter-based feature reduction techniques (Fisher Score, Mutual Information, and Information Gain) for different machine learning models, i.e., Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF). Our study demonstrates that 3-4 features can achieve optimal F1-score and accuracy in binary classification, whereas 7-8 features give reasonable performance in most of the multi-class classification tasks across both datasets. The combination of Information Gain and XGBoost with 15 features provides excellent results in binary and multi-class classification settings. Key features—protocol types, traffic metrics, temporal patterns, and statistical measures—are essential for accurate IoMT attack classification.

1 INTRODUCTION

The Internet of Medical Things (IoMT) is an interconnected network of sensors, wearable and medical devices, and clinical systems, enabling applications like remote monitoring, fitness tracking, chronic disease management, and elderly care while enhancing treatment quality, lowering costs, and facilitating prompt responses (Islam et al., 2015),(Dimitrov, 2016).

The security of IoMT is very crucial due to its role in healthcare, where sensitive patient data and critical medical systems are increasingly interconnected. IoMT devices are often targets of cyberattacks, posing risks to patient safety and data privacy (Kondeti and Bahsi, 2024). Intrusion detection systems (IDS) are essential to monitor and detect malicious activities, ensuring the reliability and security of these networks. Machine learning (ML) is vital for IDS in IoMT as it can identify complex attack patterns and adapt to evolving threats. However, IoMT devices have limited computational resources, making it essential to reduce data dimensions and select the most relevant features to ensure that ML-based IDS operates efficiently and effectively without overburdening the network. We applied filter-based feature reduction techniques (Fisher Score, Mutual Information, and Information Gain) for different machine learning models, i.e., XGBoost, KNN, Decision Tree, and Random Forest. present an analysis by utilising two benchmarking IoMT datasets CICIoMT2024 (Dadkhah et al., 2024) and IoMT-TrafficData (Areia et al., 2024) for training and testing our models. We applied filter-based feature reduction techniques (Fisher Score, Mutual Information, and Information Gain) for different machine learning models, i.e., XGBoost, KNN, Decision Tree, and Random Forest.

We evaluate the proposed model in terms of F1 score by focusing on both binary classification and multiclassification. Binary classification aims to dis-

[a] https://orcid.org/0009-0000-2656-0127
[b] https://orcid.org/0000-0001-7390-8034
[c] https://orcid.org/0000-0001-8882-4095

Rehman, M. U., Kalakoti, R. and Bahşi, H.
Comprehensive Feature Selection for Machine Learning-Based Intrusion Detection in Healthcare IoMT Networks.
DOI: 10.5220/0013313600003899

In *Proceedings of the 11th International Conference on Information Systems Security and Privacy (ICISSP 2025) - Volume 2*, pages 248-259
ISBN: 978-989-758-735-1; ISSN: 2184-4356

tinguish between benign and malicious traffic, providing a high-level detection mechanism, while multi-class classification goes further by categorizing traffic into specific attack types, enabling a granular understanding of threats. The CICIoMT2024 dataset includes traffic data for 18 types of cyberattacks (19 classes including benign traffic) grouped into five main categories (6 classes): DDoS, DoS, Reconnaissance, MQTT, and Spoofing. Similarly, the IoMT-TrafficData dataset comprises eight distinct cyberattack types, including Denial of Service, ARP Spoofing, and Network Scanning, alongside benign traffic, resulting in a 9-class classification problem.

Additionally, we evaluated the classification performance (accuracy, precision, recall, and F1) of the best-performing model, XGBoost, on both datasets, utilizing the top 15 features identified through the Information Gain (IG) feature selection. Furthermore, to address these security challenges, this study examines key network features within both datasets that are essential for identifying and classifying cyber-attacks in IoMT. Both datasets use network flow features extracted from benign and malicious traffic. Specifically, we focus on features, such as protocol type, traffic volume metrics, temporal patterns, and statistical attributes, in network flows to understand their role in distinguishing normal and attack traffic patterns.

There exists a line of research on feature selection for ML-based intrusion detection in IoT devices (Kalakoti et al., 2022; Bahşi et al., 2018). However, these studies present benchmarking results for IoT networks that include consumer IoT devices. It is necessary to understand the impact of feature selection and the best-performing features in IoMT networks, as benign traffic profiles and system components in these networks have distinct properties when compared to other IoT devices.

By highlighting critical features across IoMT datasets, this study contributes to more robust, feature-driven methods for accurate anomaly and attack detection in IoMT environments, ultimately aiming to strengthen the security and reliability of these healthcare networks. The uniqueness of our work is that we have conducted a cross-analysis between two well-developed datasets, which were released recently, to obtain more generalized findings regarding the best-performing features in IoMT networks. Our work puts a particular emphasis on feature selection in multi-class classification settings, which has not been elaborated well in the literature.

This paper is structured as follows. Section 2 reviews the related research. Section 3 presents the methodology used in our feature selection process. In Section 4, we show and discuss our results. Finally, Section 5 concludes the paper and discusses future directions.

2 RELATED WORK

In the literature, various papers employ different feature selection techniques for machine learning-based attack classification. Some studies have adopted a filter approach to identify the best feature subsets, while others have applied wrapper or embedded methods. A few works combined both filter and wrapper techniques to determine the optimal feature set. This section provides a comprehensive review of the state-of-the-art methods for feature selection in machine learning-based intrusion detection systems, as reported in the literature. In (Khammassi and Krichen, 2017), a Genetic Algorithm (GA) combined with a Logistic Regression (LR) wrapper was applied to the UNSW-NB15 and KDDCup99 datasets. Using 20 features from UNSW-NB15, the GA-LR method with a Decision Tree (DT) classifier achieved 81.42% accuracy and a false alarm rate (FAR) of 6.39%. For KDDCup99, it achieved 99.90% accuracy with 18 features. In (Osanaiye et al., 2016), a filter-based approach using Information Gain, Chi-Square, and Relief was applied for Distributed Denial of Service (DDoS) detection on the NSL-KDD dataset. Using 13 features, the DT classifier reached 99.67% accuracy and a FAR of 0.42%. The work in (Ambusaidi et al., 2016) introduced a filter-inspired reduction approach with Flexible Mutual Information (FMI) and Least Square SVM (LS-SVM), achieving 99.94% accuracy on NSL-KDD with 18 features. In (Ingre and Yadav, 2015), a filter-based feature reduction method for IDS using correlation and DT was applied to NSL-KDD, reducing the feature set to 14 attributes and achieving 83.66% accuracy for multiclass classification. In (Alazzam et al., 2020), the Pigeon Inspired Optimizer (PIO) was used for feature reduction on multiple datasets. The Sigmoid and Cosine PIO methods selected features with accuracy rates between 86.9% and 96.0%.

Janarthanan and Zargari (Janarthanan and Zargari, 2017) implemented various feature selection algorithms on UNSW-NB15, selecting optimal subsets of 5 and 8 features. Using Random Forest (RF), they achieved up to 81.62% accuracy. Vikash and Ditipriya (Kumar et al., 2020) applied Information Gain for feature reduction on UNSW-NB15, selecting 22 attributes, and their IDS achieved 57.01% Attack Accuracy (AAc) and 90% F-Measure. In (Almomani, 2020), PSO, Firefly, Grey Wolf Optimization (GO), and GA were used on UNSW-NB15, with a 30-feature subset yielding 90.48% accuracy with the J48 classifier. Maajid and Nalina (Khan et al., 2020) used Random Forest (RF) to rank features on UNSW-NB15, selecting 11 attributes, with RF achieving 75.56% accuracy. In (Tama et al., 2019), a two-stage model combining PSO, GA, and Ant Colony Optimization (ACO) on UNSW-NB15 selected 19 features, achieving 91.27% accuracy. Some studies have also used feature selection methods prior to applying explainable techniques in IoT botnet detection problems(Kalakoti et al., 2024a; Kalakoti et al., 2024c; Kalakoti et al., 2024b; Kalakoti et al., 2023).

Zong et al. (Zong et al., 2018) proposed a two-stage model using Information Gain (IG) for feature selection on UNSW-NB15, achieving 85.78% accuracy. In (Kasongo and Sun, 2020), the authors applied a filter-based feature selection technique by utilizing the XGboost algorithm on the UNSW-NB15 intrusion detection dataset. The results illustrate that feature selection method based on XGBoost enables models like DT to improve test accuracy from 88.13% to 90.85% in the binary classification.

The domain of intrusion detection systems (IDS) within the Internet of Medical Things (IoMT) has attracted considerable attention in recent years due to the growing adoption of IoMT devices in healthcare systems. To protect the security and privacy of sensitive medical data, developing effective IDS is essential. While many researcheres have focused on IDS for traditional networks, there is a notable lack of studies dedicated to IDS for the IoMT (Alalhareth and Hong, 2023a).

Feature selection techniques are crucial for enhancing the performance of IDS in the Internet of Medical Things (IoMT) (Rbah et al., 2022),(Khalil et al., 2022). These techniques reduce the dimensionality of input features while retaining essential information (Wagan et al., 2023). Filter-based methods, like chi-square and Information Gain, evaluate features individually based on their contribution to the target variable (Awotunde et al., 2021). Wrapper-based methods, such as recursive feature elimination (RFE), use ML algorithms to iteratively select and remove features, assessing their impact on model performance.

Information theory-based feature selection methods, such as MIFS and MRMR, are commonly used in fields like intrusion detection for the Internet of Medical Things (IoMT) (Gökdemir and Calhan, 2022). However, these methods require large datasets to accurately estimate Mutual Information between features and the target variable, and limited data can lead to suboptimal results (Chaganti et al., 2022). Solutions to this issue include data augmentation techniques, like oversampling or synthetic data generation (Parimala and Kayalvizhi, 2021), and transfer learning, which applies knowledge from data-rich domains to improve performance in data-limited contexts (Awotunde et al., 2021). However, these approaches come with challenges, such as introducing bias or noise and increasing computational costs (Al-Sarem et al., 2021).

In (Alalhareth and Hong, 2023b) authors proposed an improved Mutual Information feature selection technique for IDS for the IoMT. This paper proposes a Logistic Redundancy Coefficient Gradual Upweighting MIFS (LRGU-MIFS) to enhance feature selection for IDS in the IoMT. LRGU-MIFS improves detection accuracy by addressing overfitting and non-linear feature redundancy, outperforming existing methods in identifying key features.

State-of-the-art IDS systems for IoMT, such as deep learning models, offer high accuracy but are computationally intensive and less adaptable to resource-constrained environments. In contrast, our integration of feature selection techniques with IDS significantly reduces computational overhead, enhancing suitability for IoMT applications. The studies on feature selection do not create or compare the optimal sets achievable for different multiclass problem formulations. They only focus on one dataset and derive conclusions. This paper addresses this gap by inducing various learning models, including various multi-class classification models, for two well-developed and comprehensive IoMT datasets (Dadkhah et al., 2024; Areia et al., 2024) released recently. These datasets contain a huge number of attack types, making them convenient for multi-class classification. This study also conducts a cross-analysis between two datasets to identify the commonalities.

3 METHODOLOGY

We applied an ML workflow that includes the stages, data preprocessing, feature selection, and model training/testing, as demonstrated in Figure 1. In the data pre-processing stage, we eliminated the correlated features using Pearson Correlation. We applied filter-based feature selection methods (i.e., Fisher Score, Information Gain, Mutual Information) to prioritize the features. In the last stage, we benchmarked various ML algorithms (i.e., k-NN, Decision Tree, XGBoost, Random Forest) with varying numbers of selected best features determined by filter-based selection methods.

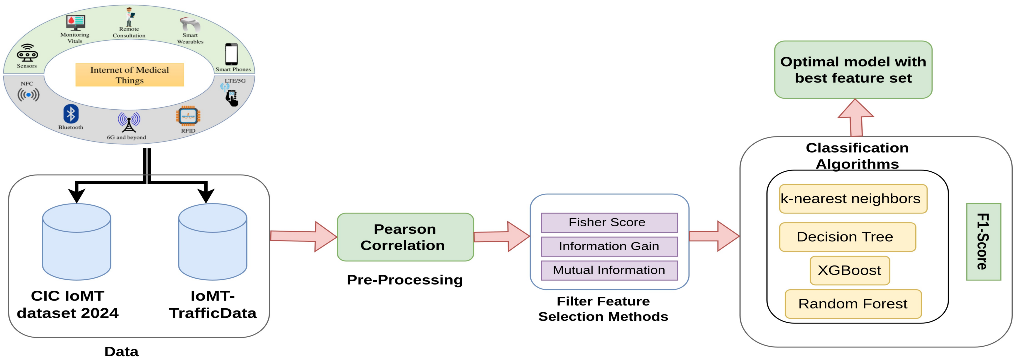


Figure 1: We employed filter methods for feature selection on the CICIoMT2024 Dataset (Dadkhah et al., 2024) and IoMT-TrafficData (Areia et al., 2024) to identify optimal features in IoMT networks. Four classifiers were used for evaluation: Decision Trees (DT), Random Forest (RF), k-Nearest Neighbors (k-NN), and XGBoost.

3.1 Datasets

We apply feature selection to the CICIoMTDataset 2024 dataset (Dadkhah et al., 2024) and IoMT-TrafficData (Areia et al., 2024), which focus on Internet of Medical Things devices in the healthcare sector. These datasets are designed to assess and improve the cybersecurity of IoMT devices through intrusion detection systems.

The CICIoMT2024 dataset (Dadkhah et al., 2024) includes traffic generated from 40 devices (25 real, 15 simulated) across multiple protocols like Wi-Fi, MQTT, and Bluetooth. The authors simulated 18 cyberattacks, categorized into five main categories i.e. DDoS, DoS, Recon, MQTT, and Spoofing.The features extracted from the attacks in the CICIoMT2024 dataset include Header Length, Duration, Rate, Srate, fin flag number, syn flag number, rst flag number, psh flag number, ack flag number, ece flag number, cwr flag number, syn count, ack count, fin count, rst count, IGMP, HTTPS, HTTP, Telnet, DNS, SMTP, SSH, IRC, TCP, UDP, DHCP, ARP, ICMP, IPv, LLC, Tot sum, Min, Max, AVG, Std, Tot size, IAT, Number, Radius, Magnitude, Variance, Covariance, Weight, and Protocol Type.

The IoMT-TrafficData dataset (Areia et al., 2024) is a comprehensive collection of network traffic data. It includes both benign and malicious traffic generated from eight different types of cyberattacks i.e. Denial of Service (DoS), Distributed Denial of Service (DDoS), ARP Spoofing, CAM Table Overflow, MQTT Malaria, Network Scanning, Bluetooth Reconnaissance, and Bluetooth Injection. The identified key features in the IP-based flows in the IoMT-TrafficData dataset cover various aspects of network communication. Protocol features include proto and service, which identify the transport and application protocols in use. Payload and packet metrics such as orig_bytes, resp_bytes, orig_pkts, and resp_pkts detail the volume and direction of data exchanged. Flow characteristics, including flow_duration and history, capture the overall session duration and connection state transitions. Packet directionality is covered by fwd_pkts_tot and bwd_pkts_tot for packet counts, and by fwd_pkts_payload and bwd_pkts_payload for payload bytes in each direction. Rate metrics (fwd_pkts_per_sec, bwd_pkts_per_sec, and flow_pkts_per_sec) provide packet transmission rates, while inter-arrival time features (fwd_iat, bwd_iat, and flow_iat) and active duration (active) reflect timing characteristics within the flow.

In this work, we used person correlation as the preprocessing step. The Pearson correlation coefficient, given by the equation (1) ,is used to compute the linear correlation between two variables. This technique involves calculating the collinearity matrix for all features to identify redundancy. The Pearson correlation coefficient P ranges from -1 to 1, where $P = 1$ indicates perfect positive correlation, $P = 0$ indicates no correlation, and $P = -1$ indicates perfect negative correlation. The formula for Pearson's correlation is:

$$P = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i - \mu_x)^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \mu_y)^2}} \tag{1}$$

Here, μ_x and μ_y represent the means of features x and y, respectively. Greater absolute values of P indicate a stronger linear relationship between the features.

3.2 Feature Selection Methods

Irrelevant features for classification problems are reduced to decrease the running time and improve the classification accuracy of machine learning algorithms. Feature selection methods are divided into three categories: wrapper, filter, and embedded techniques (Jović et al., 2015). Wrapper methods iteratively evaluate subsets of features using a machine learning algorithm, but they can be computationally intensive for high-dimensional data. In contrast, filter methods rank features independently of the learning algorithm, which may result in suboptimal selections due to the lack of guidance. To reduce computational complexity, we opted for filter-based methods, which are highly efficient and well-suited for resource-constrained IoMT environments. The following three primary filter-based feature methods are commonly employed for numeric-based feature inter-class and intra-class separation analysis and entropy-based methods as described below.

3.2.1 Fisher Score

The Fisher Score, also known as Fisher's ratio, measures the ratio of inter-class separation to intra-class separation for numeric features (Gu et al., 2012). The Fisher Score F_s is formally defined in equation 2 as:

$$F_s = \frac{\sum_{j=1}^{K} p_j(\mu_{ij} - \mu_i)^2}{\sum_{j=1}^{K} p_j \sigma_{ij}^2} \quad (2)$$

Where μ_{ij} and σ_{ij} represent the mean and standard deviation of the j-th class and i-th feature, while p_j denotes the proportion of data points in class j. A higher Fisher Score indicates greater discriminative power of a feature.

3.2.2 Mutual Information

Mutual Information (MI) quantifies the dependency between variables (Estévez et al., 2009). For continuous variables, MI is defined as:

$$I(X,Y) = \int\int p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \quad (3)$$

For discrete variables, MI is given by:

$$I(X;Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (4)$$

Here, $p(x,y)$ is the joint probability, and $p(x)$, $p(y)$ are the marginal probabilities. MI values range as follows:

$$0 \leq I(X;Y) \leq \min\{H(X), H(Y)\}$$

To enhance the Mutual Information feature selection, the following goal function is used:

$$G = I(C; f_i) - \frac{1}{|S|} \sum_{f_s \in S} NI(f_i; f_s) \quad (5)$$

Where $I(C; f_i)$ is the Mutual Information between class C and feature f_i, and S is the set of selected features. The algorithm selects features by maximizing this measure. Function $NI(f_i; f_s)$ is the Normalized Mutual Information between features f_i and f_s.

3.2.3 Information Gain

Information Gain helps quantify how much information a feature contributes to classification by utilizing the concept of entropy. It measures the reduction in dataset entropy after knowing the values of a particular feature (Velasco-Mata et al., 2021). The initial entropy of the dataset, $H(X)$, is given by the following equation, which is based on the probability $p(x)$ of a sample belonging to class x. The conditional entropy, $H(X|Y)$, after knowing the values of feature Y, is defined based on the probability $p(y)$ of a sample having feature value $y \in Y$, and the probability $p(x|y)$ of a class x sample having feature value $y \in Y$.

$$H(X) = -\sum_{x=1}^{X} p(x) \log(p(x)) \quad (6)$$

$$H(X|Y) = -\sum_{y} p(y) \sum_{x} p(x|y) \log(p(x|y)) \quad (7)$$

3.3 Machine Learning Work Flow

In our study, we employed four machine learning algorithms for classifying cyberattacks in IoMT network flow data: Decision Tree (DT), Random Forest (RF), XGBoost (XGB) and K-Nearest Neighbors (K-NN). Decision Tree (DT) is a non-parametric supervised method for classification and regression. DTs classify data by evaluating attributes at each node until reaching a decision. Random Forest (RF) is an ensemble method of decision trees, chosen for its robustness, ability to manage complex datasets, and compatibility with diverse features. XGBoost, a gradient-boosting algorithm, optimizes using second-order gradients and applies L1/L2 regularization to reduce overfitting and enhance performance. Its efficiency, interpretability, and scalability make it ideal for large datasets. Lastly, K-Nearest Neighbors (KNN) is a distance-based algorithm for classification and regression.

Our classification models were evaluated for IoMT attack detection using confusion matrices for both binary and multi-class classification. For binary classification, True positives(TP) (correctly classified attacks), True negatives (TN) (correctly classified benign traffic), False negatives (FN) (misclassified attacks), and False Positives (FP) (misclassified benign traffic) were recorded. In this study, we have utilized the F1 score metric to evaluate distinct subsets of features. The F1 score is defined as the harmonic mean of precision (P) and recall (R). It provides a more appropriate measure of incorrectly classified cases compared to accuracy. We have employed the harmonic mean of the F1 score, as it penalizes extreme values.

$$\text{F1 score} = \frac{2 \times \text{P} \times \text{R}}{\text{P} + \text{R}} \tag{8}$$

To train the models in binary classification, we have taken 5,000 samples of each class label. This results in a total of 10,000 samples from two labels. On the other hand, for the multi-class classification involving different distinct classes, we ensured an equal number of samples from each label, even for the classes with fewer instances, to maintain a balanced representation across all attack types. In the preprocessing step, after applying the Pearson correlation, balanced samples were drawn from the dataset of interest. Then, the datasets were divided into training and testing subsets in an 80/20 ratio. For evaluating each feature set with models, Random Search hyperparameter tuning was used for training the classification algorithms.

4 RESULTS AND DISCUSSIONS

This study analyzed the discriminatory power of network traffic flow features using filter-based feature selection techniques, including Fisher Score, Mutual Information, and Information Gain, for a machine learning-based intrusion detection function in IoMT healthcare networks. The analysis was conducted for binary and multiclass classification tasks on the CICIoMT2024 and IoMT-TrafficData datasets.

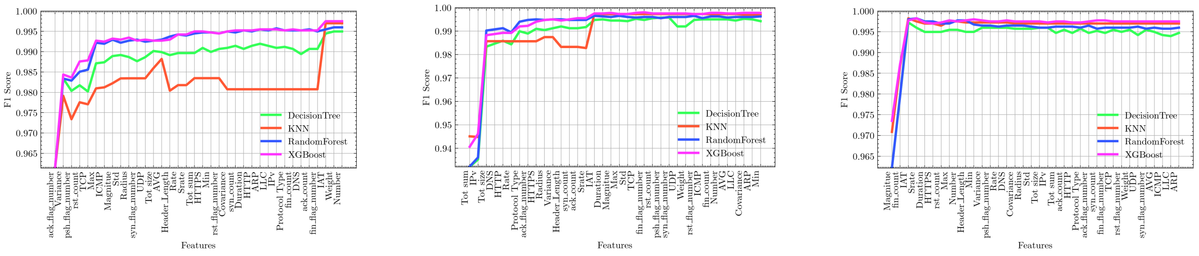
First, we applied Pearson's linear correlation coefficient (r) as a data preprocessing step to remove redundant and irrelevant data features. Any feature highly correlated with another feature ($|r| > 0.80$) was removed, keeping only one. As a result, out of the initial set of 44 features used to describe each sample in the dataset, 36 features remained in the final feature set. After removing the Pearson co-related features, in IoMT-Traffic dataset, we get 21 features, however, we also removed is_attack feature as it represent binary label. The final feature list contains 20 features.

After applying Pearson correlation and excluding unnecessary features, we applied three filter-based feature selection methods, i.e., Fisher Score, Mutual Information, and Information Gain. These methods were used to rank the importance of the remaining reduced features. An iterative, stepwise approach was used to train the ML models for each filter-based feature selection method (Fisher Score, Mutual Information, and Information Gain). Starting with the highest-ranked feature, we added one feature at a time, trained the model, and evaluated its performance progressively. For example, if the features were ranked as $f = \{f_1, f_2..f_n\}$, the model was first trained using only the top-ranked feature subset $\{f_1\}$, followed by training with $\{f_1$ and $f_2\}$, then with $\{f_1, f_2, \ldots\ f_n\}$ This process was repeated for all (n) ranked features in each method for both datasets. At each step, we added the next highest-ranked feature, as determined by the feature selection method, to the feature set incrementally to assess its impact on the model performance. The performance classifiers—Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), and XGBoost (XGB)—were evaluated based on the F1 score for both binary and multiclass classification tasks.

From the CICIoMT-2024 dataset, Binary classification was used to differentiate between benign and attack traffic. Two types of studies were performed for multi-class classification: category-based and attack-based classification. In the category-based classification, we identified six categories of network traffic: benign, MQTT attacks, DDoS, DoS, Reconnaissance, and ARP spoofing attacks, referred to as the 6-class classification. In the attack-based classification, there were 19 classes, which included various attack types such as ARP Spoofing, Ping Sweep Scan, Reconnaissance VulScan, OS Scan, Port Scan, Malformed Data Packets, Connect Flood (DoS), Publish Flood (DDoS), Publish Flood (DoS), Connect Flood (DDoS), TCP (DoS), ICMP (DoS), SYN (DoS), UDP (DoS), SYN (DDoS) , TCP (DDoS) , ICMP (DDoS) , and UDP (DDoS). Attack-based detection is referred to as a 19-class classification.

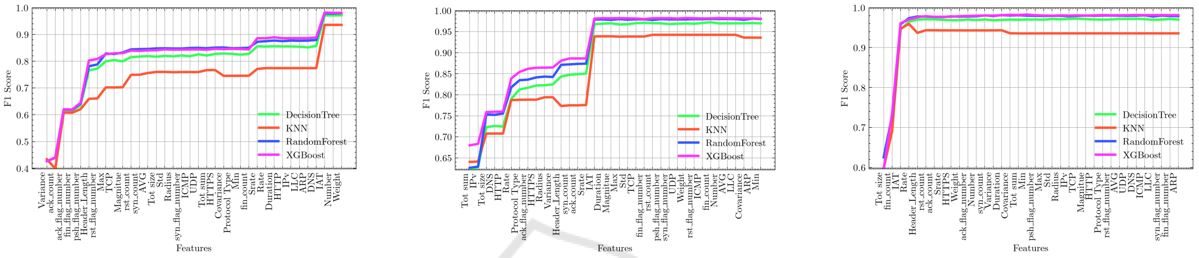
Fig. 2 shows the algorithm's performance comparison using different feature selection methods on the CICIoMT2024 dataset for binary classification. Across all three feature selection methods, the classifiers' performance rapidly improves by adding the first few features. However, the performance plateau shows only marginal improvements as more features are added. When all 36 features were included, a small subset of highly informative features had already achieved high performance across models. Most classifiers achieved high F1 scores (above 0.99) with only 5-10 features. Notably, XGB and RF consistently reached near-optimal performance with fewer than five features, while DT and KNN demonstrated more gradual improvements as features were added, achieving their best results after more features were incorporated into the model.

Attacks categories (Figure 3) and 19-classes based classification (Figure 4) show almost the same performance in comparison with binary classification as XGBoost and Random Forest again perform best, particularly when fewer features are used. However, Mutual Information demonstrates overall higher model



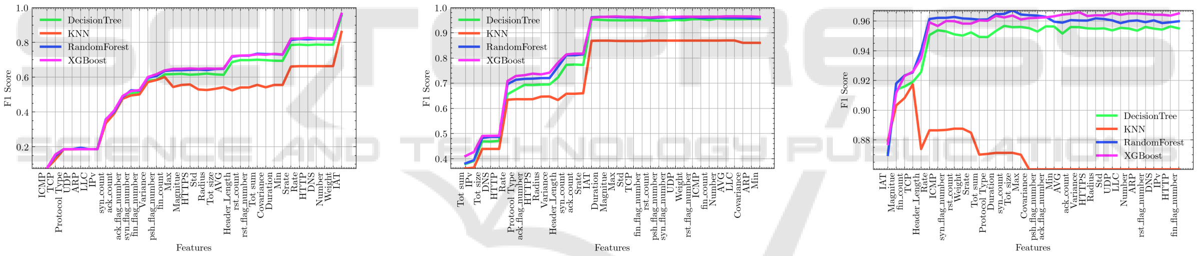
(a) Fisher Score (b) Mutual Information (c) Information Gain

Figure 2: Comparison of algorithms performance using Feature selection methods over CICIoMT2024 data set for Binary Classification.



(a) Fisher Score (b) Mutual Information (c) Information Gain

Figure 3: Comparison of algorithms performance using Feature selection methods over CICIoMT2024 data set for 6-class Classification.



(a) Fisher Score (b) Mutual Information (c) Information Gain

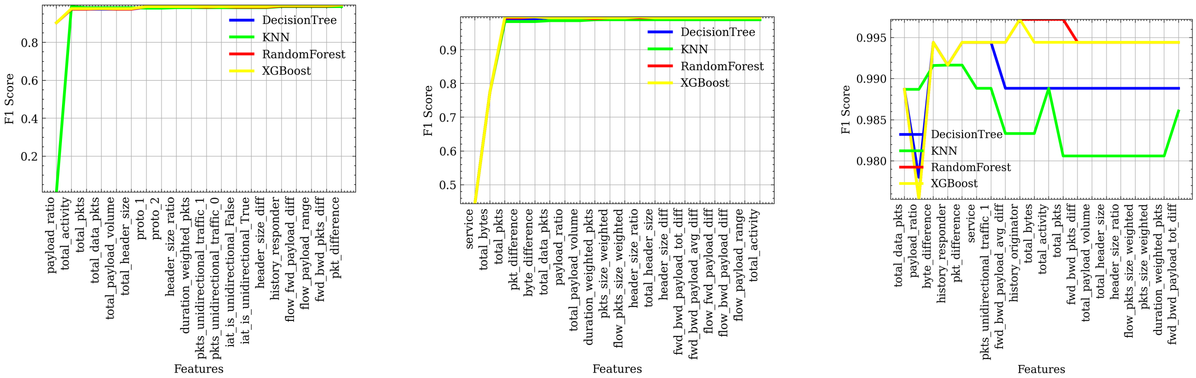
Figure 4: Comparison of algorithms performance using Feature selection methods over CICIoMT2024 data set for 19-class Classification.

performance early on, i.e., for the first three features in binary classification compared to multi-class classifications. Figure 4 shows that KNN performance drops dramatically after the first 4 features in the case of Information Gain. This shows that KNN does not work well for multi-class classification in the CICIoMT2024 dataset.

From the IoMT-TrafficData dataset, binary classification was used to differentiate between benign and attack traffic and multi-class classification was used to classify the attack traffic further. In the attack-based classification, there were 9 classes, which included 8 different types of cyberattacks i.e. DoS (Apachekiller, Slowread, Rudeadyet, Slowloris), Distributed Denial of Service (DDoS), ARP Spoofing, Buffer Overflow (Camoverflow), MQTT Malaria, and Network Scanning (Netscan).

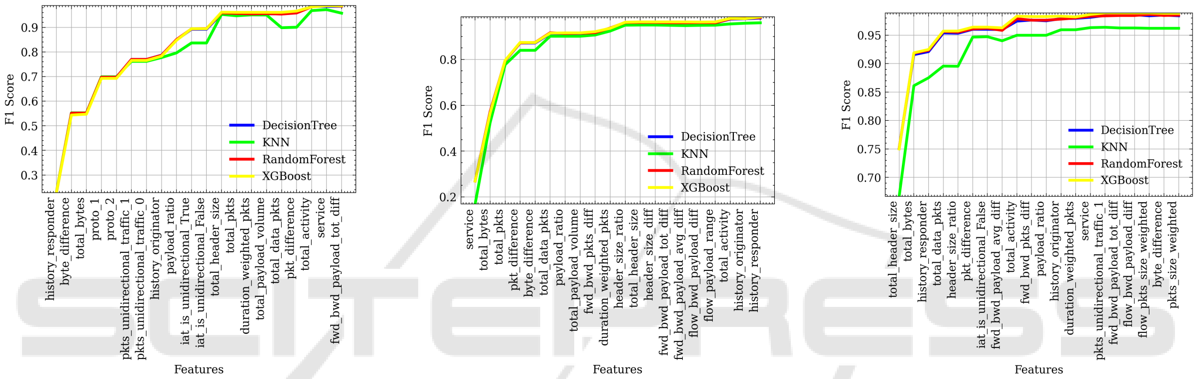
Table 1 presents the classification performance report of the XGBoost model on both CICIoMT2024 and IoMT-TrafficData, using selected top-15 features from Information Gain (IG) feature selection. Both datasets show excellent binary classification performance, with accuracy, precision, recall, and F1-score around 0.997 for both classes (attack and benign), indicating strong classification ability. On the CICIoMT2024 dataset, the model performs well across 6 and 19-class classifications, with high accuracy (0.977) and consistent metrics, though performance slightly drops for complex classes like Recon and ARP Spoofing.

On the IoMT-TrafficData dataset, accuracy remains high at 0.987, with perfect precision and recall



(a) Fisher Score Features (b) Mutual Information (c) Information Gain

Figure 5: Comparison of algorithms performance using Feature selection methods over IoMT-TrafficData dataset for Binary Classification.



(a) Fisher Score (b) Mutual Information (c) Information Gain

Figure 6: Comparison of algorithms performance using Feature selection methods over IoMT-TrafficData dataset for Multi-class classification (9 classes).

Table 1: Classification report of selected top-15 features from Information Gain (IG) feature selection for the CICIoMT2024 dataset & IoMT-TrafficData , using the XGBoost model for all three classification types.

Dataset	Classification type	Binary				6-Class classification report														
CICIoMT2024 dataset	Metric\class	Attack	Benign		Metric\class	ARP Spoofing	Benign	DDoS	DoS	MQTT	Recon									
	Accuracy	0.997	0.997		Accuracy	0.977	0.977	0.977	0.977	0.977	0.977									
	Precision	0.999	0.994		Precision	0.918	0.959	0.999	1.000	0.997	0.994									
	Recall	0.995	0.999		Recall	0.967	0.945	1.000	0.998	0.993	0.962									
	F1-Score	0.997	0.997		F1-Score	0.942	0.952	1.000	0.999	0.995	0.978									
		19-Class classification																		
	Metric/Class	ARP Spoofing	Benign	MQTT-DDoS -Connect Flood	MQTT-DDoS -Publish Flood	MQTT-DoS -Connect Flood	MQTT-DoS -Publish Flood	MQTT-Malformed Data	Recon-OS Scan	Recon-Ping Sweep	Recon-Port Scan	Recon-VulScan	TCP IP-DDoS-ICMP	TCP IP-DDoS-SYN	TCP IP-DDoS-TCP	TCP IP-DDoS-UDP	TCP IP-DoS-ICMP	TCP IP-DoS-SYN	TCP IP-DoS-TCP	TCP IP-DoS-UDP
	Accuracy	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967	0.967
	Precision	0.902	0.931	1.000	1.000	1.000	1.000	0.899	0.870	0.947	0.902	0.915	1.000	0.997	1.000	0.997	1.000	1.000	1.000	1.000
	Recall	0.877	0.922	1.000	0.987	0.997	0.997	0.958	0.831	0.967	0.902	0.927	1.000	1.000	1.000	1.000	0.997	0.997	1.000	0.997
	F1-Score	0.889	0.926	1.000	0.994	0.998	0.998	0.927	0.850	0.957	0.902	0.921	1.000	0.998	1.000	0.998	0.998	0.998	1.000	0.998
IoMT-Traffic dataset	Classification type	Binary				9-Class classification report														
	Metric\class	Attack	Benign		Metric\class	Apachekiller	Arpspoofing	Camoverflow	Mqttmalaria	Netscan	Normal	Rudeadyet	Slowloris	Slowread						
	Accuracy	0.997	0.997		Accuracy	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987	0.987						
	Precision	0.997	0.997		Precision	0.993	1.0	1.0	0.996	1.0	0.974	0.982	0.977	0.965						
	Recall	0.997	0.996		Recall	0.981	0.987	1.0	0.989	1.0	0.993	0.966	0.981	0.997						
	F1-Score	0.997	0.997		F1-Score	0.987	0.993	1.0	0.993	1.0	0.983	0.974	0.979	0.977						

for many attack types (e.g., Camoverflow, Netscan), but slightly lower performance for some attack classes like Slowread and Rudeadyet.

Fig. 5 shows the algorithm's performance comparison using different feature selection methods on the IoMT-TrafficData dataset for binary classification. Across Mutual Information, the classifiers' performance rapidly improves with the addition of the first few features. Fisher Score follows the same pattern; however, it gives a slightly lower Model performance. Information Gain effectively ranks features by their usefulness, as the performance improves significantly with the first few features. XGBoost emerges as the best-performing classifier, while KNN performance tends to decrease after the first 5 features. The results suggest that focusing on the top-ranked features can optimize classifier performance while reducing computational costs. Figure 6 compares algorithm per-

Table 2: Binary classification top 15 best features.

Type of features	CICIoMT2024 dataset	IoMT-TrafficDat dataset
Protocol	TCP UDP ICMP IPv DNS HTTP Protocol Type HTTPS	proto_2 service
Traffic Volume	Tot sum Tot size	total_bytes payload_ratio total_activity
Temporal metrics	Duration IAT	iat_is_unidirectional_True duration_weighted_pkts
Packets rate	Rate srate	
Flags	ack flag number psh flag number syn flag number rst count syn count ack count fin count	history_responder hisoty_originator
Other statistical features	Std variance Max Magnitude Min Radius AVG	byte_difference fwd_bwd_pkts_diff fwd_bwd_payload_avg_diff fwd_bwd_payload_tot_diff flow_bwd_payload_diff flow_payload_range pkt_difference
Other features	Header Length	pkts_unidirectional_traffic_1

Table 3: 19-class classification top 15 best features in CICIoMT2024 dataset.

Type of features	Fisher Score	Mutual information	Information Gain
Protocol	ICMP TCP UDP ARP LLC Ipv Protocol Type	IPv DNS HTTP HTTPS Protocol type	ICMP TCP Protocol type
Traffic Volume		tot sum tot size	tot sum
Temporal metrics			IAT Duration
Packets rate		Rate srate	Rate srate
Flags	ack flag number psh flag number syn flag number fin flag number	ack flag number	syn flag number
Header Attributes		Header_Length	Header_Length
Other statistical features	Variance	Variance	Variance Mangnitude Weight
Other	syn_count ack_count fin_count	syn_count ack_count	syn_count rst_count fin_count

Table 4: 9-class classification top 15 best features in IoMT-TrafficData dataset.

Type of features	Fisher Score	Mutual information	Information Gain
Protocol	proto_2	service	service
Traffic Volume	total_bytes payload_ratio total_pkts total_header_size total_payload_volume	total_bytes payload_ratio total_data_pkts total_header_size total_payload_volume	total_bytes payload_ratio total_data_pkts
Temporal metrics	duration_weighted_pkts iat_is_unidirectional_True iat_is_unidirectional_False	duration_weighted_pkts	duration_weighted_pkts total_activity iat_is_unidirectional_False
Flags	history_responder history_originator		history_responder history_originator
Header Attributes	header_size_diff	header_size_ratio	header_size_ratio
Other statistical features	byte_difference pkts_unidirectional_traffic_1 pkts_unidirectional_traffic_0	byte_difference pkt_difference fwd_bwd_pkts_diff fwd_bwd_payload_avg_diff fwd_bwd_payload_tot_diff	pkt_difference pkts_unidirectional_traffic_1 fwd_bwd_pkts_diff fwd_bwd_payload_avg_diff

formance for multi-class classification (9 classes) on the IoMT-TrafficData dataset. Performance gradually improves with the increase in features in the case of the Fisher Score. However, models obtain higher performance earlier (after four features) in the case of Mutual Information and Information Gain. All models show comparable performance across the feature selection methods, except KNN, which lags behind when using the Information Gain method.

By examining the important features in both datasets, it is possible to identify the important network characteristics for attack detection in IoMT traffic. Table 2 illustrates the union of the top 15 features selected by different feature selection methods. The CICIoMT2024 dataset includes transport-layer protocol features, TCP, and UDP, while the IoMT-TrafficData dataset uses proto_2, which also represents transport-layer protocols. Therefore, we conclude that TCP and UDP are important features. Both datasets also emphasize application-layer protocols, such as HTTP, DNS, and SMTP (in CICIoMT2024) and service (in IoMT-TrafficData dataset), which identify application-layer protocols as well. Tot sum (CICIoMT2024) provides a key metric to understand traffic volume when considered alongside total_pkts and total_bytes in the IoMT-TrafficData dataset. Flags in CICIoMT2024 directly capture counts of specific TCP flags, while IoMT-TrafficData's history_responder encapsulates the sequence of connection states, reflecting the flags' transitions. Variability measures in packet lengths in a flow, such as Std and Variance (Ratio of the variances of incoming to outgoing packet lengths in the flow) in CICIoMT2024, along with pkt_diff, byte_difference and fwd_bwd_payload_tot_diff, which capture the fluctuations and differences in packet lengths in IoMT-TrafficData, are essential metrics in identifying anomalies in IoMT networks. Temporal features such as Duration and IAT (Inter-arrival time) in the CICIoMT2024 dataset can be compared with duration_weighted_pkts and iat_is_unidirectional_False in the IoMT TrafficData dataset, which provides additional directional features, i.e., unidirectional/bi-directional that enhance understanding of packet arrival patterns.

Table 3 identifies several common features across Fisher Score, Mutual Information, and Information Gain methods (selected by at least two methods) for 19-class classification in CICIoMT2024 dataset. Features that relate to protocols (ICMP, TCP, Protocol type), traffic volume (tot sum), packet transmission rate (Rate, Srate) , flags (ack flag number, syn flag number, syn_count), and statistical properties (variance), are pivotal for distinguishing patterns and detecting multiple attacks in IoMT traffic, highlighting their relevance in network security analysis.

The common important features underlined in Table 4 reveal crucial insights into detecting multiple attacks in IoMT-TrafficData dataset for 9-class classification. In traffic volume, features like total_bytes, payload_ratio, total_header_size, and total_payload_volume appear frequently, emphasizing the significance of overall data transferred and packet structure. Temporal metrics are also prominent, with duration_weighted_pkts, which capture the rate or proportion of packets over time within a flow, and iat_is_unidirectional_False, capturing consistency of IAT with bidirectional traffic flowing traffic. Among flags, history_responder and history_originator recur, reflecting connection state transitions. It should also be noted that Mutual Information could not grasp any flag information. The header attribute feature, header_size_ratio, refers to the proportion of the header size relative to the total size of a packet also highlights the significance of packet header size. Lastly, statistical features like byte_difference (difference in payload bytes between the originator and responder), fwd_bwd_pkts_diff (difference in the number of packets sent forward and backward in the connection), fwd_bwd_payload_avg_diff (difference in average payload size per packet between forward and backward traffic), and pkts_unidirectional_traffic_1 (indication of unidirectional traffic) show significance in multi-class attack classification in IoMT traffic.

The analysis of Figures 2, 3, 4, 5, and 6 reveals that filter-based methods exhibit excellent performance for binary classification across both datasets. Notably, these models achieve higher performance levels early on, often after selecting just 3 to 4 features using information gain for feature selection. This indicates that these methods are effective in differentiating between benign and malicious traffic with a minimal set of features. As the number of selected features increases, the models' performance steadily improves. Significant accuracy is attained with 7 to 8 features, particularly for multi-class classifications (6-class, 9-class, and 19-class) using information gain. Among the evaluated models, the XGBoost model achieved the highest performance with fewer features selected through information gain feature selection.

5 CONCLUSION AND FUTURE WORK

In this work, we performed filter-based feature selection methods (Fisher Score, Mutual Information, Information Gain) to identify the best features in two IoMT datasets (CICIoMT2024 and IoMT-TrafficData,). We compared the performance of four machine learning algorithms (Decision Tree, Random Forest, K-Nearest Neighbors, and XGBoost) in both datasets. We checked the performance for binary and multi-class classifications in both datasets.

Fisher Score works well for both datasets, especially for classifiers like Decision Tree and KNN, which show gradual improvements as more features are added. Mutual Information is highly effective across both datasets, particularly for Random Forest and XGBoost, which reach optimal performance with fewer features. For the CICIoMT2024 dataset in binary classification, XGBoost and Random Forest perform best with Fisher Score or Mutual Information, requiring fewer features for optimal results, while Multi-class (6-class & 19-class) observed a similar trend with XGBoost and Random Forest consistently outperforming other models when using with the mentioned methods.

Information Gain works better for CICIoMT2024 datasets but shows a different pattern for binary classification in the IoMT-TrafficData dataset, where performance does not improve as rapidly compared to the other methods. Furthermore, the binary classification of IoMT-TrafficData with XGBoost and Random Forest shows superior performance with Mutual Information and Fisher Score, achieving near-optimal results with only a few features. Fisher Score and Mutual Information are again the most effective in Multi-class classification, especially for Random Forest and XGBoost in IoMT-TrafficData.

Our paper highlights key features for IoMT attack detection across both datasets, including essential transport-layer protocols (TCP, UDP), application-layer identifiers (e.g., HTTP, DNS), and traffic volume metrics (e.g., total_bytes, payload_ratio). Temporal and directional metrics, like Duration, IAT, and connection-state flags (history_responder), enhance understanding of packet flows, while variability and statistical measures (variance, byte_difference) are crucial for identifying attack patterns, underscoring their importance in multi-class attack classification in IoMT traffic. Furthermore, the XGBoost model demonstrates excellent performance in both binary and multi-class classification across the CICIoMT2024 and IoMT-TrafficData datasets, with minor variations in handling certain attack types. Our study shows that filter-based methods perform well in binary classification with 3-4 features, while multi-class classification achieves significant accuracy with 7-8 features across both datasets. Furthermore, this study also illustrates that using the top-15 features of the selection of information gains (IG) features for the XGboost model, achieving excellent binary classification results (0.997 accuracy, precision, recall, and F1 score) and very good performance in multiclass classifications, with slight drops for few complex attacks, thereby opening doors for further research.

In future work, exploring hybrid feature selection methods, such as combining Mutual Information with optimization techniques like Genetic Algorithms, could improve feature relevance. Implementing non-stationary models to dynamically adapt to new features and unseen attacks would also enhance the robustness of intrusion detection systems in healthcare IoMT networks. Furthermore, extending the work to include other types of datasets, such as telemetry, software, hardware threats, or monitored data from implantable devices, could broaden the applicability of the results.

ACKNOWLEDGMENT

This study was co-funded by the European Union and Estonian Research Council via project TEM-TA5.

REFERENCES

Al-Sarem, M., Saeed, F., Alkhammash, E. H., and Alghamdi, N. S. (2021). An aggregated mutual information based feature selection with machine learning methods for enhancing iot botnet attack detection. *Sensors*, 22(1):185.

Alalhareth, M. and Hong, S.-C. (2023a). An improved mutual information feature selection technique for intrusion detection systems in the internet of medical things. *Sensors*, 23(10).

Alalhareth, M. and Hong, S.-C. (2023b). An improved mutual information feature selection technique for intrusion detection systems in the internet of medical things. *Sensors*, 23(10):4971.

Alazzam, H., Sharieh, A., and Sabri, K. E. (2020). A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. *Expert systems with applications*, 148:113249.

Almomani, O. (2020). A feature selection model for network intrusion detection system based on pso, gwo, ffa and ga algorithms. *Symmetry*, 12(6):1046.

Ambusaidi, M. A., He, X., Nanda, P., and Tan, Z. (2016). Building an intrusion detection system using a filter-based feature selection algorithm. *IEEE transactions on computers*, 65(10):2986–2998.

Areia, J., Bispo, I., Santos, L., and Costa, R. L. d. C. (2024). Iomt-trafficdata: Dataset and tools for benchmarking intrusion detection in internet of medical things. *IEEE Access*.

Awotunde, J. B., Abiodun, K. M., Adeniyi, E. A., Folorunso, S. O., and Jimoh, R. G. (2021). A deep learning-based intrusion detection technique for a secured iomt system. In *International Conference on Informatics and Intelligent Applications*, pages 50–62. Springer.

Bahşi, H., Nõmm, S., and La Torre, F. B. (2018). Dimensionality reduction for machine learning based iot botnet detection. In *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 1857–1862. IEEE.

Chaganti, R., Mourade, A., Ravi, V., Vemprala, N., Dua, A., and Bhushan, B. (2022). A particle swarm optimization and deep learning approach for intrusion detection system in internet of medical things. *Sustainability*, 14(19):12828.

Dadkhah, S., Neto, E. C. P., Ferreira, R., Molokwu, R. C., Sadeghi, S., and Ghorbani, A. A. (2024). Ciciomt2024: A benchmark dataset for multi-protocol security assessment in iomt. *Internet of Things*, 28:101351.

Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, 22(3):156–163.

Estévez, P. A., Tesmer, M., Perez, C. A., and Zurada, J. M. (2009). Normalized mutual information feature selection. *IEEE Transactions on neural networks*, 20(2):189–201.

Gökdemir, A. and Calhan, A. (2022). Deep learning and machine learning based anomaly detection in internet of things environments. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 37(4):1945–1956.

Gu, Q., Li, Z., and Han, J. (2012). Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.

Ingre, B. and Yadav, A. (2015). Performance analysis of nsl-kdd dataset using ann. In *2015 international conference on signal processing and communication engineering systems*, pages 92–96. IEEE.

Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., and Kwak, K.-S. (2015). The internet of things for health care: a comprehensive survey. *IEEE access*, 3:678–708.

Janarthanan, T. and Zargari, S. (2017). Feature selection in unsw-nb15 and kddcup'99 datasets. In *2017 IEEE 26th international symposium on industrial electronics (ISIE)*, pages 1881–1886. IEEE.

Jović, A., Brkić, K., and Bogunović, N. (2015). A review of feature selection methods with applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1200–1205.

Kalakoti, R., Bahsi, H., and Nõmm, S. (2024a). Improving iot security with explainable ai: Quantitative evaluation of explainability for iot botnet detection. *IEEE Internet of Things Journal*.

Kalakoti, R., Bahsi, H., and Nõmm, S. (2024b). Explainable federated learning for botnet detection in iot networks. In *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 01–08.

Kalakoti, R., Nõmm, S., and Bahsi, H. (2022). In-depth feature selection for the statistical machine learning-based botnet detection in iot networks. *IEEE Access*, 10:94518–94535.

Kalakoti, R., Nõmm, S., and Bahsi, H. (2023). Improving transparency and explainability of deep learning based iot botnet detection using explainable artificial intelligence (xai). In *2023 International Conference on Machine Learning and Applications (ICMLA)*, pages 595–601. IEEE.

Kalakoti, R., Nõmm, S., and Bahsi, H. (2024c). Enhancing iot botnet attack detection in socs with an explainable active learning framework. In *2024 IEEE World AI IoT Congress (AIIoT)*, pages 265–272. IEEE.

Kasongo, S. M. and Sun, Y. (2020). Performance analysis of intrusion detection systems using a feature selection method on the unsw-nb15 dataset. *Journal of Big Data*, 7(1):105.

Khalil, A. A., E Ibrahim, F., Abbass, M. Y., Haggag, N., Mahrous, Y., Sedik, A., Elsherbeeny, Z., Khalaf, A. A., Rihan, M., El-Shafai, W., et al. (2022). Efficient anomaly detection from medical signals and images with convolutional neural networks for internet of medical things (iomt) systems. *International Journal for Numerical Methods in Biomedical Engineering*, 38(1):e3530.

Khammassi, C. and Krichen, S. (2017). A ga-lr wrapper approach for feature selection in network intrusion detection. *computers & security*, 70:255–277.

Khan, N. M., Madhav C, N., Negi, A., and Thaseen, I. S. (2020). Analysis on improving the performance of machine learning models using feature selection technique. In *Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2*, pages 69–77. Springer.

Kondeti, V. and Bahsi, H. (2024). Mapping cyber attacks on the internet of medical things: A taxonomic review. In *2024 19th Annual System of Systems Engineering Conference (SoSE)*, pages 84–91. IEEE.

Kumar, V., Sinha, D., Das, A. K., Pandey, S. C., and Goswami, R. T. (2020). An integrated rule based intrusion detection system: analysis on unsw-nb15 data set and the real time online dataset. *Cluster Computing*, 23:1397–1418.

Osanaiye, O., Cai, H., Choo, K.-K. R., Dehghantanha, A., Xu, Z., and Dlodlo, M. (2016). Ensemble-based multi-filter feature selection method for ddos detection in cloud computing. *EURASIP Journal on Wireless Communications and Networking*, 2016:1–10.

Parimala, G. and Kayalvizhi, R. (2021). An effective intrusion detection system for securing iot using feature selection and deep learning. In *2021 international conference on computer communication and informatics (ICCCI)*, pages 1–4. IEEE.

Rbah, Y., Mahfoudi, M., Balboul, Y., Fattah, M., Mazer, S., Elbekkali, M., and Bernoussi, B. (2022). Machine learning and deep learning methods for intrusion detection systems in iomt: A survey. In *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, pages 1–9. IEEE.

Tama, B. A., Comuzzi, M., and Rhee, K.-H. (2019). Tse-ids: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. *IEEE access*, 7:94497–94507.

Velasco-Mata, J., González-Castro, V., Fernández, E. F., and Alegre, E. (2021). Efficient detection of botnet traffic by features selection and decision trees. *IEEE Access*, 9:120567–120579.

Wagan, S. A., Koo, J., Siddiqui, I. F., Qureshi, N. M. F., Attique, M., and Shin, D. R. (2023). A fuzzy-based duo-secure multi-modal framework for iomt anomaly detection. *Journal of King Saud University-Computer and Information Sciences*, 35(1):131–144.

Zong, W., Chow, Y.-W., and Susilo, W. (2018). A two-stage classifier approach for network intrusion detection. In *Information Security Practice and Experience: 14th International Conference, ISPEC 2018, Tokyo, Japan, September 25-27, 2018, Proceedings 14*, pages 329–340. Springer.