Anomaly Detection of Distributed Denial of Service (DDoS) in IoT Network Using Machine Learning

Baydaa Hashim Mohammed (baidaa81@gmail.com)
Universiti Kebangsaan Malaysia
Hasimi Sallehudin
Universiti Kebangsaan Malaysia
Nurhizam Safie
Universiti Kebangsaan Malaysia
Mohd Satar
Universiti Kebangsaan Malaysia
Hamed Dhary Murhg
Al-Salam University College
Shaymaa Abdelghany Mohamed
University of Technology

Research Article

Keywords: Anomaly Detection, DDoS, IoT, Machine Learning, Network Security, Intrusion Detection, Cybersecurity

Posted Date: November 14th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3496063/v1

Additional Declarations: No competing interests reported.

Abstract

This research focuses on developing an anomaly detection system using machine learning to mitigate Distributed Denial of Service (DDoS) attacks in IoT networks. The study utilizes a diverse dataset from IoT environments to train and evaluate machine learning algorithms for DDoS detection. The dataset includes various IoT device types, communication protocols, and network configurations. The research aims to achieve several objectives, including dataset preprocessing, feature engineering, machine learning model selection, anomaly detection, and performance evaluation. The research team preprocesses the raw Internet of Things (IoT) network data by cleaning and transforming it to prepare it for analysis. They then extract relevant features from the data to effectively characterize normal and abnormal network behavior. Multiple machine learning algorithms are evaluated and compared to determine the most suitable models for DDoS detection in IoT networks. The selected machine learning models are then used to identify and classify abnormal traffic patterns associated with DDoS attacks. The performance of the developed anomaly detection system is evaluated by assessing its accuracy, precision, recall, and F1 score. The significance of this research lies in its potential to enhance the security of IoT networks by proactively detecting and mitigating DDoS attacks. By leveraging machine learning, the study aims to provide a robust defense mechanism against this pervasive threat, ensuring the reliability and availability of IoT services and applications.

1. Introduction

The IoT has completely changed how objects and systems communicate with one another, making it possible for seamless interactions to take place between the digital and physical worlds. However, the growing prevalence of IoT devices has also given rise to new security issues (Dawod et al., 2022). DDoS attacks, which involve flooding a target system or network with traffic to overload its resources and make it unavailable to legitimate users, are one example of these problems (Zakaria et al., 2020). These kinds of attacks may have significant repercussions, including the interruption of services, loss of financial resources, and compromising security and privacy. Device heterogeneity poses challenges in ensuring uniform security measures across the network due to the diverse types of devices (Tufail et al., 2021). Limited resources make implementing robust security protocols and encryption difficult, leaving IoT devices susceptible to attacks (Alahmadi et al., 2023). Communication protocols, such as HTTP, MQTT, and CoAP, may have their security considerations, making them vulnerable to attacks (Yin et al., 2020).

Data transmission is another area where IoT devices transmit sensitive data without proper encryption or authentication, leading to data breaches and privacy violations. Weak or absent authentication mechanisms can allow unauthorized access to IoT devices and networks, compromising network integrity (Ceron et al., 2019). Firmware updates may not be implemented, leaving devices vulnerable to known exploits. Physical access to IoT devices can lead to tampering or theft, potentially providing attackers with valuable information or manipulating devices (U. Islam et al., 2022). Insecure APIs can be exploited to gain unauthorized control over devices or extract sensitive data. A lack of security awareness among end-users and manufacturers can result in poor security practices and weak password choices

(Galchynsky et al., 2021). Scalability can also pose a vulnerability if not managed properly, making monitoring and securing all devices challenging. Often used by IoT devices, cloud-based services can be targeted with DDoS attacks, affecting the entire IoT ecosystem (Mohammad Shah et al., 2023).

It is very necessary to have a defense system that is both proactive and intelligent to fight against DDoS assaults and ensure the security of IoT networks. A powerful option for detecting DDoS attacks in IoT networks is anomaly detection driven by machine learning (Mohammed, Husairi et al., 2022). We can equip our models to differentiate between regular network behavior and the aberrant actions associated with DDoS attacks by training these models on various datasets that encapsulate the complexities of IoT traffic (Yusuf et al., 2022). Due to the proliferation of susceptible devices and the potential for broad disruption, attacks in IoT networks offer a serious security concern. Effective mitigation measures must understand these behaviors (Din et al., 2021). Amplification attacks use susceptible servers to boost DNS and NTP traffic to the target IoT device or network. Large data floods from these assaults may render IoT devices unusable and interrupt their functioning (Mohammed, Sallehuddin et al., 2022). A botnet of hacked IoT devices controlled by a hostile actor sends communication to the target concurrently. Malware may transform these devices into "zombies," which can overload IoT networks and impair services. Depletion attacks drain IoT devices' CPU power, preventing them from delivering valid requests (Mohammed, Sallehudin, et al., 2022). IoT services may be disrupted by application layer assaults like HTTP/HTTPS floods, which deplete processing capabilities. By exploiting IoT communication protocol weaknesses, protocol-based attacks cause service outages and network congestion. IoT device exploitation uses firmware or software flaws to take control, making attack traffic hard to track (Mohammed, Sallehudin, et al., 2022).

The purpose of a burst attack is to overload IoT devices by causing abrupt increases in the volume of data. These assaults bypass DDoS defenses and produce quick interruptions. Designing proactive security solutions requires knowing these kinds and features of DDoS assaults in IoT networks. Mitigation solutions should incorporate real-time traffic analysis, anomaly detection, and adaptive countermeasures to protect IoT devices and services. This research aims to build and implement an anomaly detection system that uses methods derived from machine learning to strengthen the security of IoT networks. IoT networks are becoming increasingly susceptible to DDoS attacks because they are becoming more integrated into vital infrastructure, residences, hospitals, and businesses. It presents a huge problem. The potential effect of a distributed denial of service assault on these networks is enormous, ranging from the interruption of services and financial losses to serious dangers to public safety and invasion of privacy (Tufail et al., 2021). The goals of the research include the preprocessing of data from IoT networks in preparation for analysis using machine learning, the engineering of relevant features to capture the essence of IoT network behavior, the selection and deployment of machine learning models that are capable of detecting anomalies indicative of DDoS attacks, and the evaluation of the performance of the anomaly detection system and its accuracy in distinguishing between normal and malicious network activities.

2. Literature Review

The literature review section of the paper provides an in-depth exploration of existing research, studies, and findings related to IoT network security and the specific threat of DDoS attacks in IoT environments. This section aims to build a comprehensive understanding of the state of the field, identify gaps in the current knowledge, and establish the foundation for the proposed research. This knowledge is a foundation for the proposed study on developing a machine learning-based anomaly detection system for DDoS attacks in IoT networks, aiming to contribute to the ongoing efforts to secure IoT ecosystems.

Machine learning algorithms are used to detect abnormal traffic patterns in IoT networks, allowing for the identification of potential DDoS attacks. For instance, Islam et al. (2021) proposed a " Towards Machine Learning Based Intrusion Detection in IoT Networks. " This paper specifically focuses on anomaly detection in IoT networks, a critical aspect of DDoS mitigation. It narrows down the research problem, making the contributions more targeted. The paper proposes a "two-tier" anomaly detection scheme, which suggests innovation in IoT network security. This innovation can contribute to more effective detection. While the specific focus can be a strength, it can also be a weakness if the research is too specialized and lacks a broader context. Considering how this research fits into the larger IoT security landscape is important. Depending on the complexity of the proposed two-tier scheme, it might be challenging to implement in real-world IoT environments. Practicality can be a concern. Dawod et al. (2022) introduce "IoT Device Integration and Payment via an Autonomic Blockchain-Based Service for IoT Device Sharing." This paper comprehensively surveys DDoS detection and mitigation in IoT networks. It offers a holistic view of the state of the field, which can be valuable for researchers and practitioners. Surveys are useful for synthesizing existing knowledge, identifying trends, and highlighting research gaps. They serve as a valuable resource for anyone interested in the topic. While surveys are helpful, they do not contribute new research findings. This paper may lack the depth of analysis found in primary research papers. The comprehensiveness of the survey can also be a weakness in terms of the paper's length and readability. It might cover a wide range of research areas but lack in-depth exploration.

Furthermore, flow-based analysis techniques monitor traffic flows within IoT networks, identifying deviations from normal behavior as potential DDoS attacks. For instance, Santos et al. (2023) proposed "A flow-based intrusion detection framework for IoT networks." The paper introduces a flow-based DDoS detection scheme, which can effectively identify abnormal traffic patterns in IoT networks. This approach aligns with real-world network data analysis. The paper's focus on flow-based DDoS detection might limit its applicability to broader DDoS attack scenarios. It may not address other types of attacks or provide a comprehensive solution. The paper could benefit from a more in-depth discussion of its limitations, including potential challenges in real-world implementations and scalability.

Similarly, Li et al., (2022) Proposed "FLEAM: A Federated Learning Empowered Architecture to Mitigate DDoS in Industrial IoT." This paper addresses DDoS attack detection in the context of fog computing, an emerging IoT paradigm. It considers resource constraints and can be more applicable to IoT environments. The paper's consideration of limited resources aligns with the resource-constrained nature

of many IoT devices and fog computing environments. The paper's focus on fog computing might limit its applicability to traditional IoT or cloud environments. It may not address DDoS detection in other IoT scenarios. Depending on the complexity of the proposed detection method, it might be challenging to implement in resource-constrained IoT environments. Practicality and efficiency need to be considered.

Additionally, real-time monitoring and analysis of network traffic patterns in IoT networks to identify unusual spikes or patterns indicative of DDoS attacks. As a result, Alhalabi et al. (2023) proposed a "Distinctive Measurement Scheme for Security and Privacy in IoT Applications Using Machine Learning Algorithms." The paper addresses the critical security and privacy issue in IoT data analysis, which is highly relevant given the vast amount of data generated by IoT devices. The paper's focus on security and privacy demonstrates a holistic approach to IoT data analysis, which is essential today. While the article provides valuable insights into big data security and privacy, its scope might be limited to these aspects. It may not comprehensively address other aspects of IoT security. Depending on the proposed solutions, the implementation complexity could be a concern, especially for resource-constrained IoT devices. In addition, Alhalabi et al., (2023) proposed a "Distinctive Measurement Scheme for Security and Privacy in the IoT Applications Using Machine Learning Algorithms. " The paper leverages machine learning techniques for detecting unauthorized IoT devices. This approach is effective in identifying anomalies and security threats. Unauthorized device detection is a practical and crucial security concern in IoT, making the paper's findings highly relevant. Similar to the first paper, the focus on unauthorized device detection might limit the paper's applicability to broader IoT security challenges. Resource-intensive algorithms may not be suitable for resource-constrained IoT devices depending on the machine learning techniques.

Also, behavioral analysis of IoT devices involves profiling normal behavior and identifying deviations from these profiles, which can indicate DDoS attacks. Elkhodr et al. (2023) proposed " BioChainReward: A Secure and Incentivised Blockchain Framework for Biomedical Data Sharing." The paper addresses the critical area of IoT-based healthcare and emphasizes the importance of intrusion detection, which is crucial for patient data security and privacy. By leveraging big data analytics, the paper demonstrates a sophisticated approach to intrusion detection, which is well-suited for handling the vast amount of data generated in healthcare IoT.

While the paper provides valuable insights into IoT-based healthcare and intrusion detection, its scope may be limited to this specific context. It may not comprehensively address broader IoT security challenges. Depending on the big data analytics techniques used, the complexity of implementation and resource requirements could be a concern, especially for resource-constrained IoT devices. Similarly, Hammad et al. (2023) introduced "Security Framework for Network-Based Manufacturing Systems with Personalized Customization: An Industry 4.0 Approach." The paper addresses DDoS resilience in the context of Industry 4.0, which is highly relevant in industrial IoT. Ensuring resilience against DDoS attacks is critical for industrial systems. DDoS resilience is a practical concern in industrial IoT, making the paper's findings directly applicable to real-world scenarios. The focus on DDoS resilience in Industry 4.0 might limit the paper's applicability to broader IoT security challenges or other industry sectors. The

resource requirements and scalability in industrial IoT environments must be considered depending on the proposed DDoS resilience techniques.

Furthermore, Collaborative DDoS detection involves IoT devices sharing information and collectively identifying DDoS attacks based on their observations. Thus, Hammad et al. (2023) proposed " Security Framework for Network-Based Manufacturing Systems with Personalized Customization: An Industry 4.0 Approach. " The paper addresses IoT security in the context of smart cities, which is highly relevant and reflects a real-world application of IoT security. The paper introduces a collaborative intrusion detection system for adaptive IoT, which can effectively identify security threats across a network of IoT devices. While the article provides valuable insights into IoT security in smart cities, its focus may limit its applicability to other IoT contexts outside of smart cities. Depending on the proposed collaborative intrusion detection system, the complexity of implementation and scalability in IoT environments must be considered. Al Rawajbeh et al., (2023) paper is based on " A new model for security analysis of network anomalies for IoT devices." The article is a survey that provides a comprehensive overview of network flow-based approaches for IoT security. It serves as a valuable resource for researchers and practitioners in the field. The survey's focus on network flow-based methods makes it relevant to various IoT security scenarios, not limited to specific contexts like smart cities. As a survey paper, it does not present original research findings but synthesizes existing research. It may not provide specific solutions but rather an overview of existing approaches. Given the breadth of topics covered, the paper may not delve deeply into each approach. Readers seeking in-depth analysis of specific techniques may need to refer to primary sources.

Lastly, cloud-based solutions leverage the scalability and resources of cloud platforms to analyze IoT network traffic for DDoS attacks. Accordingly, Safarov et al., (2023) proposed an "Explainable Lightweight Block Attention Module Framework for Network-Based IoT Attack Detection." The paper presents a lightweight DDoS detection framework for IoT, which is essential for resource-constrained IoT devices where efficiency is crucial. The framework is designed to be robust against DDoS attacks, a significant advantage in IoT environments where security is critical. While the paper provides valuable insights into lightweight DDoS detection for IoT, its focus may limit its applicability to other aspects of IoT security. Depending on the scale of IoT deployments, the scalability of the proposed framework needs to be considered. Similarly, Islam et al., (2023) presented " Real-Time Detection Schemes for Memory DoS (M-DoS) Attacks on Cloud Computing Applications." The paper introduces a hybrid cloud-based DDoS detection scheme, which can potentially handle large-scale IoT deployments effectively. Scalability and cloud-based solutions are highly relevant for security challenges in IoT environments with numerous devices. The hybrid cloud-based approach might introduce complexity regarding implementation and resource requirements, which could be challenging for certain IoT scenarios. The paper focuses on DDoS detection, which might limit its applicability to broader IoT security challenges.

However, there are still issues with how well application-layer DDoS assaults are detected. First, we must improve at spotting attacks and fewer false alarms throughout the detection process. Second, most studies have concentrated on finding ways to identify DDoS attacks at high rates. Very few academics

worry about low-rate DDoS attacks. Very few studies have offered methods for identifying either high-speed or low-speed DDoS assaults. It's also important to broaden the spectrum of detection. Third, some current data sets include information that has since become irrelevant, making comparisons and assessments impossible. Thus, this paper highlights the research gap in the need for effective anomaly detection methods to mitigate DDoS attacks in IoT networks using machine learning. Due to the diverse range of devices and communication patterns, traditional signature-based methods may not be sufficient to detect novel and sophisticated attacks in IoT environments. The proposed solution addresses this gap by leveraging machine learning techniques for anomaly detection in IoT networks. Machine learning can potentially adapt and identify previously unknown attack patterns, making it suitable for IoT security. The paper emphasizes the need for more advanced and adaptive approaches to DDoS detection in IoT, considering IoT environments' unique challenges and characteristics. Developing machine learning-based anomaly detection methods tailored to IoT networks can enhance the security and resilience of IoT deployments against DDoS threats.

3. Methodology and Materials

Using Machine Learning, the proposed method of anomaly DDoS in IoT networks is provided in this section. Thus, the following steps are used to achieve the proposed system. The study proposal is shown in Fig. 1, and we employed DDOS attack data from an open-source platform—data gathered from a publicly available source, which was an unprocessed dataset (Lange & Kettani, 2019). Using preprocessing methods, we preprocessed the dataset. The removal of null values from datasets was followed by strategies for balancing the dataset to be scaled and balanced. After selecting the most important characteristics, we partitioned the data into a training set of a 70% ratio and a testing set of 30%. The machine learning models are trained on the training set, whereas the testing set tests and assesses the models.

3.1 Data Collection and Description

The IoT devices generate network traffic, which includes information on communication patterns, data payloads, and source-destination pairs. Device metadata, such as device type, manufacturer, firmware version, and identifiers, can provide context for anomaly detection (Najafi et al., 2022). Sensor data, such as temperature, humidity, motion, or pollution levels, can detect anomalies when they deviate significantly from expected values. Server logs generated by IoT servers, gateways, or cloud platforms contain valuable information about device interactions, user activities, and system events (Mohammad Shah et al., 2023). Analyzing server logs can reveal abnormal patterns or suspicious activities. Subscribing to external threat intelligence feeds can provide real-time information about known attack sources, malware signatures, and emerging threats, which can be integrated into the anomaly detection system.

3.2 Description of the Dataset

The dataset used for fraud detection is called the "Fraud Detection Dataset" in academic literature. The dataset monitors network intrusions. DDoS is another feature of this harmful virus. Table 1 and Fig. 2

provide the dataset's properties and descriptions.

Table 1
A Detailed Description of the Feature Dataset.

Feature	Description	Variable Type
ID	ATM ID	Input Variable
State	“State of Railway (Connectivity)”	Input Variable
Spkts	“Source Packets (Sent to destination)”	Input Variable
Dpkts	“Destination Packets (Received at destination)”	Input Variable
Sbytes	“Source Bytes (Sent from Source)”	Input Variable
Dbytes	“Destination Bytes (Received from Source)”	Input Variable
Attack_Cat	“Category of an Attack” DDoS attacks apply here; if the label is 0, there is no attack; if the label is 1, there is an attack.	The variable that is dependent in this study comprises nine distinct classes.

Figure 2 illustrates the distribution of services provided by four personal computers (PCs) inside a banking institution. According to the heat map, it can be inferred that a repartition of services from a PC with a value over 0.5 indicates a heightened probability of a severe DDoS assault. When the value derived from a personal computer (PC) falls below 0.5, the likelihood of an attack occurring diminishes.

Figure 3 displays the overall distribution of the target class numbers throughout the dataset. The information reveals a frequency of 50,000 instances of DDoS harm.

These datasets are used to test the suggested approach. Preprocessed datasets can be used for deep learning. The homogeneity measure (k-means clustering) is an unsupervised approach for choosing important features from both data sets. Five-fold cross-validation can estimate and improve deep learning model performance. We employed three machine learning models to classify attacks. We have split the dataset into 70% training and 30% testing sets. Empirical studies show that the best results are obtained if we use 20–30% of the data for testing and the remaining 70–80% for training.

3.3 Data Preprocessing Techniques

Data cleaning is essential for analyzing IoT data, which can contain missing values, duplicates, or outliers. Feature engineering is crucial for anomaly detection, particularly in DDoS detection in IoT networks. Techniques like normalization, scaling, and dimensionality reduction can improve machine learning algorithms' performance. Handling imbalanced data and time-series decomposition are also essential. Labeled data is necessary for supervised learning, and datasets are divided into training, validation, and testing sets. Feature selection methods identify informative features and reduce

dimensionality, while interaction features reveal patterns not evident when considering individual features. Outlier detection methods identify unusual behavior, potentially indicating anomalies.

In summary, the data preprocessing includes;

The dataset is preprocessed to make it more appropriate for the ML classifier.

1. *Removal of Socket Information*: The source and destination IP addresses must be deleted to remove any potential bias in the identifying procedure. Instead of relying solely on data from a single socket, it can rule out hosts with similar packet information by examining packet characteristics.
2. *Remove White Spaces*: Labels with several classes may contain white spaces. Due to the varied labels for the other tuples in this class, it has two categories.
3. *Label Encoding*: Labeling labels into a numeric form so that a computer may read them is called label encoding. As a result, machine learning algorithms can make better decisions about how to use those labels. Supervised learning is a key step in preprocessing the structured dataset.
4. *Data Normalization*: Non-normalized data create inefficiencies in the predictions of outcomes, so we have normalized the dataset by using a standard scalar function. After the normalization of the dataset, feature ranking occurs.
5. *Feature Ranking*: Col_0 to Col_111 shows the number of features. Because of the large number of characters and strings, we have used the label col_0 to col_111. We have used k-means clustering in feature ranking of attributes; k-means used the weight of each feature and ranked according to the weight and checked the relevance of each feature in the outcome.

Thus, Fig. 4 presents the feature correlation matrix.

4. Machine Learning Algorithms

When selecting machine learning algorithms for IoT-based DDoS detection, consider measures like accuracy, recall, F1-score, and ROC-AUC to evaluate the model's ability to detect irregularities and reduce false positives. A comprehensive understanding of the data and unique anomaly detection job is crucial. Model training involves data collection, preprocessing, hyperparameter tuning, feature engineering, model training, validation, evaluation, threshold selection, deployment, monitoring and maintenance, documentation and interpretability, security, and scalability (Alahmadi et al., 2023; Yin et al., 2020). Data collection involves clean IoT data, preprocessing handles missing values, and feature engineering creates relevant features. Model training, validation, evaluation, deployment, monitoring, and maintenance ensure the model is scalable and capable of handling IoT data and devices (Al-Hadhrami & Hussain, 2021).

This study classified DDOS attacks in the dataset using three machine learning algorithms: K-Nearest Neighbours (KNN), Support Vector Machines (SVM), and Random Forests.

4.1 Support Vector Machine

Support Vector Machine (SVM) is a powerful machine learning algorithm for classification and regression tasks, particularly for binary classification problems. It focuses on margin maximization, which ensures robust generalization to new data (Kate et al., 2022). SVM can handle non-linearly separable data using the kernel trick, which transforms the data into a higher-dimensional space. Support vectors, the closest data points to the hyperplane, are crucial in defining the optimal hyperplane. A regularization parameter, "C," balances the desire for a large margin by correctly classifying all training data points (Xia et al., 2022). SVM can be extended to handle multi-class problems using one-vs-one or one-vs-all classification techniques. It can also be used for regression tasks, predicting a continuous target variable instead of discrete class labels. SVM is widely used in various domains due to its versatility and robustness in handling different data types (Kim & Meystre, 2020).

The SVM Classifier is used to categorize attacks in this work, using a new design based on the input signal vector (x) and the support vector (s) in the hidden layer (y). The output neuron calculates the total of the linear outputs generated by the neurons in the hidden layer. This research integrates two classification tasks, including using the SVM technique to analyze characteristic attributes and collecting switch data to extract relevant characteristic values. The primary objective is to identify the ideal classification hyperplane that distinguishes between normal network traffic and DDoS attack data, which is presented mathematically in Eq. 1 as;

$$SVM_{DDoS} = \sum_{i=1}^{n} G_i k\left(x_i s_i\right) \ldots (1)$$

Subsequently, the model is evaluated using test data to produce accurate classification findings, as illustrated in Fig. 5.

4.2 Random Forests

The random forests (RF) approach is an effective machine learning ensemble method that may be used for classification and regression applications. It combines the predictions of numerous models, each made up of decision trees, to generate more accurate predictions (Abdalzaher et al., 2023). Randomness is introduced on two levels, the first of which is random sampling, and the second is random feature selection. When doing classification tasks, each decision tree "votes" for the class it predicts, and the class that receives the most votes is the one that is selected as the final prediction (Islam et al., 2023). The predictions of each tree are summed before being used in regression tasks. RF can offer a measure of feature relevance and be highly parallelizable and less susceptible to outliers and noise. As a result of their capacity to process high-dimensional data, nonlinear connections, and noisy datasets, they find widespread usage in various applications (Krell et al., 2022).

Within the scope of this research, we are categorizing attacks by using the RF Classifier architecture. Figure 6 depicts its overall design.

The present study involves the integration of the logistic regression model with the RF classifier to enhance the accuracy of both models. The mathematical model may be expressed in the form that follows:

$$RF_{DDoS} = \sum_{i-1}^{n} f(x) \ldots (2)$$

Where RF_{DDoS} is equivalent to **D**, then we have;

$$\ln m + hD = m + hD \ldots (3)$$

$$\frac{V}{1-V} = e^{m+hD} \ldots (4)$$

$$V = \frac{e^{m+hD}}{1+e^{m+hD}} \ldots (5)$$

V is the Logistic Regression probability function, and *D* is the RF classification model result. $\sum_{i-1}^{n} f(x)$ Represents the boosting mechanism used by the RF Classifier. The result of D in RF will be sent to the logistic regression's probability function to determine the chances of a class containing either 0 or 1.

4.3 K-Nearest Neighbors

The K-Nearest Neighbours (KNN) method is a supervised machine-learning technique that may be used for classification and regression problems. It is an instance-based learning approach that makes all of the observations from the training dataset available as a point of reference for future observations (Ray, 2019; Roy et al., 2021). Both classification and regression issues are amenable to applying KNN, which gives a class label to a new data point based on the majority class found among that point's k closest neighbours. It uses a distance metric, commonly the Euclidean distance, to determine which data points are most comparable (Sokkalingam & Ramakrishnan, 2022). The number of nearest neighbours used to make predictions is determined by the hyperparameter K of the KNN algorithm, which must be selected before training begins (Sokkalingam & Ramakrishnan, 2022). Because it is non-parametric, it can deal with intricate patterns and conform to various data formats. KNN is easy to grasp, efficient in capturing complicated correlations in data, excellent for dealing with issues involving several classes, and does not need a model training process (Alkahtani & Aldhyani, 2022).

Within the scope of this research, we will be classifying attacks by using the recently developed KNN Classifier architecture. Figure 7 presents an overview of the device's architectural composition.

The vector input signal, denoted by x, is located in the input layer. Calculated between the input signal vector (k) and the neighbours (n) in the hidden layer (y), it is referred to as the y-intercept. The output neuron sums together the linear outputs O of the neurons in the hidden layer. The findings in Naseri & Gharehchopogh, (2022) demonstrate that KNN can achieve a high detection rate and a low percentage of

false positives. The paper Islam et al. (2023) presents a method that uses KNN and indicates that it is particularly helpful in detecting DDoS attacks.

$$KNN_{DDoS} = \sum_{i=1}^{n} G_i k\left(K_i s_i\right) \ldots (6)$$

5. Evaluation Criteria and Performance Metrics

The True Positive Rate measures the percentage of DDoS attacks correctly detected by a model, with higher rates indicating better detection. The False Positive Rate calculates the percentage of normal traffic incorrectly classified as DDoS attacks, with lower rates minimizing false alarms. Precision quantifies the accuracy of DDoS attack predictions, with higher precision resulting in fewer false alarms (Zhou et al., 2022). The F1-Score, the harmonic mean of precision and recall, provides a balanced measure of a model's performance considering false positives and negatives.

Accuracy, precision, recall, and F1 scores have been used to evaluate the performance of techniques. The Confusion Matrix has also demonstrated true and false positive rates. The potential solutions were evaluated using the criteria of accuracy, precision, recall, and F1 Score. Using a confusion matrix, we have shown how to distinguish between correctly categorized and incorrectly classified sentences. Table 2 shows the results of the computations used to determine the metrics in this research.

Table 2
Performance Metrics and Its Computation Model

Performance Metrics	Metrics Computation
Accuracy	$Accuracy = [\backslash \text{raisebox} 1ex\$TP\$ / \backslash \text{raisebox} -1ex\$ \left(TP + TN\right) \$] * 100$
Precision	$Precision = [\backslash \text{raisebox} 1ex\$TP\$ / \backslash \text{raisebox} -1ex\$ \left(TP + FP\right) \$] * 100$
Recall	$Recall = [\backslash \text{raisebox} 1ex\$TP\$ / \backslash \text{raisebox} -1ex\$ \left(TP + FN\right) \$] * 100$
F1 Score	$F1 = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right.$

6. Results

This section compares the results obtained by applying each model to the selected dataset. The SVM, RF, and KNN models have all been evaluated using their datasets, and the results are compared with the

research carried out by (Dawod et al., 2022). The dataset contains nine different attacks, and we will use them as examples to determine how well our machine can accurately categorize them.

6.1 The Performance Comparative Analysis of the Proposed SVM Vs Existing Model

A supervised machine learning model, an SVM, uses classification techniques to tackle classification issues involving two groups. Subsequently feeding an SVM model training data for each category, it is feasible to train the model to classify fresh text using the classes it has been taught to recognize. The effectiveness of the SVM model compared to the existing model has been shown in Fig. 8. Figure 9 provides the Confusion Matrix of Normalized and Non-Normalized SVM model.

Table 3
Performance Comparative Analysis of the Proposed SVM Vs Existing Model

	Proposed Method	Existing Method (Dawod et al., 2022)
Accuracy	99.85%	78.20%
Precision	99.60%	86.10%
Recall%	98.66%	83%
F1 Score	98.70%	87.70%

In terms of accuracy, precision, recall, and F1 score while employing SVM, the findings shown in Table 3 demonstrate that the proposed approach is superior to the currently used method. For instance, in terms of accuracy, the findings indicate that the approach proposed, which generates a score of 99.85%, exceeds the method that is currently being used, which generates a score of 78.20%. In addition, the results of the proposed approach in terms of accuracy, recall, and F1 score are 99.60%, 98.66%, and 98.70% respectively, which is higher than the results of the method.

6.2 The Comparative Analysis of the Proposed RF Vs. Existing Model

The RF approach to categorizing data is a classification method comprising several different decision trees. Using bagging and feature randomization, it aims to build an uncorrelated forest of trees for each tree. This is done to ensure that the prediction made by the committee is more accurate than the prediction provided by any one tree. Figures 10 and 11 provide the results of evaluating how well the RF model performed and how confusing its matrix appeared.

Table 4
Performance Comparative Analysis of the Proposed RF Vs Existing Model

	Proposed Method	Existing Method
Accuracy	96.80%	79.00%
Precision	97.58%	89.00%
Recall%	95.55%	86%
F1 Score	97.80%	82.00%

Regarding accuracy, precision, recall, and F1 score while utilizing RF, the analysis results provided in Table 4 suggest that the proposed approach is superior to the presently used technique. This is shown by the fact that the proposed approach is more effective than the currently used way. For instance, in terms of accuracy, the data reveal that the proposed method, which gives a score of 96.80%, is superior to the methodology presently being utilized, generating a score of 79.00%. This is because the proposed method uses a different algorithm. Similarly, the suggested technique's accuracy, recall, and F1 scores are 97.58%, 95.55%, and 97.80%, respectively, which is greater than the results of the method.

6.3 The Comparative Analysis of the Proposed KNN Vs Existing Model

The KNN method may be used for classification as well as regression analysis. It is a supervised machine learning method that is simple to implement. It is simple to install and comprehend, but one of its primary flaws is that its speed degrades as the data being processed increases. The successful performance of the KNN model and its confusion matrix is provided in Figs. 12 and 13, respectively.

Table 5
Performance Comparative Analysis of the Proposed KNN Vs Existing Model

	Proposed Method	Existing Method
Accuracy	98.90%	83.00%
Precision	98.88%	81.80%
Recall%	96.58%	86%
F1 Score	98.78%	84.00%

The analytical findings in Table 5 reveal that the suggested strategy is superior to the currently utilized technique. This is the case regarding accuracy, precision, recall, and the F1 score when KNN is employed. This is shown by the fact that the method that is offered is more effective than the one that is presently being employed. For instance, the statistics demonstrate that the suggested approach, which provides a score of 98.90%, is better than the methodology now being used, which generates a score of 83.00%. This is because the recommended method has a higher rate of accuracy. This is because the suggested

approach uses a different algorithm. In a similar vein, the results of the recommended strategy in terms of accuracy, recall, and F1 score are respectively 98.88%, 96.58%, and 98.70%, which are higher than the results of the method.

7. Discussion

The proposed machine learning-based approach has improved considerably over current techniques for detecting Distributed Denial of Service (DDoS) in IoT networks. Due to its superior capacity to analyze massive datasets and detect tiny abnormalities, the methodology greatly outperforms conventional techniques in terms of accuracy. As a result, DDoS attacks may be pinpointed with greater accuracy. Greater accuracy means less chance of false alarms and less time and energy spent monitoring for DDoS assaults. The method also demonstrates improved recall, which enables DDoS assaults to be detected and countered in real-time. By combining accuracy and recall, the F1 Score allows for accurate detection of DDoS assaults with minimal false positives. The capacity of machine learning algorithms to learn and adapt from data, spotting intricate patterns and anomalies that are difficult to capture with more conventional approaches, is largely responsible for these advancements. In addition, machine learning models may be updated and improved regularly, making them flexible enough to deal with the constantly shifting challenges encountered in IoT contexts. Finally, machine learning is a major improvement over traditional approaches for detecting anomalies in IoT networks and DDoS assaults in particular. It is a helpful tool for strengthening IoT ecosystems' safety and robustness because of its higher accuracy, precision, recall, and F1 Score.

8. Conclusion

The study conducted on "Anomaly Detection of DDoS in IoT Networks Using Machine Learning" has yielded findings that demonstrate the substantial improvement in the accuracy of DDoS attack identification in IoT networks via the use of machine learning-based anomaly detection methods. These strategies demonstrate a high level of effectiveness in differentiating between typical network activity and abnormal activities that are indicative of malicious intent. As a consequence, they achieve enhanced accuracy and increased recall rates. The integration of enhanced precision and recall results in an elevated F1 Score, which signifies a more optimal equilibrium between the accuracy of detection and the reduction of false alarms.

This study significantly enhances anomaly detection approaches, specifically in IoT ecosystems. Machine learning methodologies provide a more advanced and flexible resolution for detecting DDoS attacks, which pose an escalating threat in IoT networks. This study makes a valuable contribution to boosting the security and resilience of IoT networks by improving the accuracy of detection and measures such as precision, recall, and the F1 Score. Ensuring the integrity and operation of IoT systems is of paramount importance. Another major improvement is a reduction in false positives, which reduces the number of erroneous security warnings and related expenses and interruptions to business operations. Adapting machine learning models to evolving threats in IoT networks is a significant

addition, as it guarantees the long-term effectiveness of security solutions. The models above can undergo continual updates to effectively combat emerging and developing DDoS attack patterns, ensuring their prolonged efficacy.

The study results and contributions underscore the significance of using machine learning-based anomaly detection techniques to address DDoS assaults in IoT networks. These methodologies' increased accuracy, precision, recall, and F1 Score are significant.

Furthermore, future research directions focus on enhancing anomaly detection capabilities in IoT networks by combining multiple machine learning models, using deep learning approaches, developing real-time threat response systems, focusing on behavioral analysis, explaining AI, leveraging cross-domain threat intelligence, implementing edge computing solutions, addressing scalability and resource efficiency, developing datasets, and researching regulatory and compliance aspects. These efforts aim to make IoT networks more accurate, efficient, and adaptable to evolving security threats. Directions for future work in IoT security and machine learning.

Declarations

Author's Contribution

Hasimi Sallehudin and Nurhizam Safie Mohd Satar proofread the entire manuscript. Hamed Dhary Murhg, and Shaymaa Abdelghany Mohamed developed the introductory and literature review section, respectively. Baydaa Hashim Mohammed wrote the other sections and implemented the proposed technique.

Competing Interest

This article has no competing interests.

Availability of Data and Materials

All data and materials will be made available.

References

1. Abdalzaher, M. S., Fouda, M. M., Elsayed, H. A., & Salim, M. M. (2023). Towards Secured IoT-based Smart Systems Using Machine Learning. *IEEE Access*, *11*(November 2022), 20827–20841. https://doi.org/10.1109/ACCESS.2023.3250235
2. Al-Hadhrami, Y., & Hussain, F. K. (2021). DDoS attacks in IoT networks: a comprehensive systematic literature review. In *World Wide Web* (Vol. 24, Issue 3). World Wide Web. https://doi.org/10.1007/s11280-020-00855-2

3. Al Rawajbeh, M., Alzyadat, W., Kaabneh, K., Afaneh, S., Alrwashdeh, D. F., Albayaydah, H. S., & Alhadid, I. H. (2023). A new model for security analysis of network anomalies for IoT devices. *International Journal of Data and Network Science*, *7*(3), 1241–1248. https://doi.org/10.5267/j.ijdns.2023.5.001
4. Alahmadi, A. A., Aljabri, M., Alhaidari, F., Alharthi, D. J., Rayani, G. E., Marghalani, L. A., Alotaibi, O. B., & Bajandouh, S. A. (2023). DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions. *Electronics (Switzerland)*, *12*(14), 1–24. https://doi.org/10.3390/electronics12143103
5. Alhalabi, W., Al-Rasheed, A., Manoharan, H., Alabdulkareem, E., Alduailij, M., Alduailij, M., & Selvarajan, S. (2023). Distinctive Measurement Scheme for Security and Privacy in Internet of Things Applications Using Machine Learning Algorithms. *Electronics (Switzerland)*, *12*(3). https://doi.org/10.3390/electronics12030747
6. Alkahtani, H., & Aldhyani, T. H. H. (2022). Artificial Intelligence Algorithms for Malware Detection in Android-Operated Mobile Devices. *Sensors*, *22*(6), 1–26. https://doi.org/10.3390/s22062268
7. Ceron, J. M., Steding-Jessen, K., Hoepers, C., Granville, L. Z., & Margi, C. B. (2019). Improving iot botnet investigation using an adaptive network layer. *Sensors (Switzerland)*, *19*(3), 1–16. https://doi.org/10.3390/s19030727
8. Dawod, A., Georgakopoulos, D., Jayaraman, P. P., Nirmalathas, A., & Parampalli, U. (2022). IoT Device Integration and Payment via an Autonomic Blockchain-Based Service for IoT Device Sharing. *Sensors*, *22*(4), 1–22. https://doi.org/10.3390/s22041344
9. Din, Z., Jambari, D. I., Yusof, M. M., & Yahaya, J. (2021). Challenges in IoT Technology Adoption into Information System Security Management of Smart Cities: A Review. *Advances in Science, Technology and Engineering Systems Journal*, *6*(2), 99–112. https://doi.org/10.25046/aj060213
10. Elkhodr, M., Gide, E., & Darwish, O. (2023). BioChainReward : A Secure and Incentivised Blockchain Framework for Biomedical Data Sharing. *International Journal of Environmental Research and Public Health*, *6*(12), 1–14.
11. Galchynsky, L., Graivoronskyi, M., & Dmytrenko, O. (2021). Evaluation of Machine Learning Methods to Detect DoS / DDoS Attacks on IoT. *CEUR Workshop Proceedings*, *3241*, 225–236.
12. Hammad, M., Jillani, R. M., Ullah, S., Namoun, A., Tufail, A., Kim, K. H., & Shah, H. (2023). Security Framework for Network-Based Manufacturing Systems with Personalized Customization: An Industry 4.0 Approach. *Sensors*, *23*(17). https://doi.org/10.3390/s23177555
13. Islam, N., Farhin, F., Sultana, I., Kaiser, S., Rahman, S., Mahmud, M., Hosen, S., & Cho, G. H. (2021). Towards Machine Learning Based Intrusion Detection in IoT Networks. *Computers, Materials and Continua*, *69*(2), 1801–1821. https://doi.org/10.32604/cmc.2021.018466
14. Islam, U., Al-Atawi, A., Alwageed, H. S., Ahsan, M., Awwad, F. A., & Abonazel, M. R. (2023). Real-Time Detection Schemes for Memory DoS (M-DoS) Attacks on Cloud Computing Applications. *IEEE Access*, *11*(July), 74641–74656. https://doi.org/10.1109/ACCESS.2023.3290910

15. Islam, U., Muhammad, A., Mansoor, R., Hossain, M. S., Ahmad, I., Eldin, E. T., Khan, J. A., Rehman, A. U., & Shafiq, M. (2022). Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models. *Sustainability (Switzerland)*, *14*(14). https://doi.org/10.3390/su14148374
16. Kate, P., Ravi, V., & Gangwar, A. (2022). FinGAN: Chaotic generative adversarial network for analytical customer relationship management in banking and insurance. *Neural Computing and Applications*, 1–22. https://doi.org/10.1007/s00521-022-07968-x
17. Kim, Y., & Meystre, S. M. (2020). Ensemble method-based extraction of medication and related information from clinical texts. *Journal of the American Medical Informatics Association*, *27*(1), 31–38. https://doi.org/10.1093/jamia/ocz100
18. Krell, N., Davenport, F., Harrison, L., Turner, W., Peterson, S., Shukla, S., Marter-Kenyon, J., Husak, G., Evans, T., & Caylor, K. (2022). Using real-time mobile phone data to characterize the relationships between small-scale farmers' planting dates and socio-environmental factors. *Climate Risk Management*, *35*(January), 100396. https://doi.org/10.1016/j.crm.2022.100396
19. Lange, T., & Kettani, H. (2019). On Security Threats of Botnets to Cyber Systems. *2019 6th International Conference on Signal Processing and Integrated Networks, SPIN 2019*, 176–183. https://doi.org/10.1109/SPIN.2019.8711780
20. Li, J., Lyu, L., Liu, X., Zhang, X., & Lyu, X. (2022). FLEAM: A Federated Learning Empowered Architecture to Mitigate DDoS in Industrial IoT. *IEEE Transactions on Industrial Informatics*, *18*(6), 4059–4068. https://doi.org/10.1109/TII.2021.3088938
21. Mohammad Shah, I. N., Ismail, E. S., Samat, F., & Nek Abd Rahman, N. (2023). Modified Generalized Feistel Network Block Cipher for the Internet of Things. *Symmetry*, *15*(4). https://doi.org/10.3390/sym15040900
22. Mohammed, B. H., Husairi, A., Sallehudin, H., Alaba, F. A., & Safie, N. (2022). A Conceptual Framework for Securing IoT-BIM. *Proceedings - AiIC 2022: 2022 Applied Informatics International Conference: Digital Innovation in Applied Informatics during the Pandemic*, *May*, 68–71. https://doi.org/10.1109/AiIC54368.2022.9914592
23. Mohammed, B. H., Sallehuddin, H., Yadegaridehkordi, E., Safie Mohd Satar, N., Hussain, A. H. Bin, & Abdelghanymohamed, S. (2022). Nexus between Building Information Modeling and Internet of Things in the Construction Industries. *Applied Sciences (Switzerland)*, *12*(20). https://doi.org/10.3390/app122010629
24. Mohammed, B. H., Sallehudin, H., Mohamed, S. A., Satar, N. S. M., & Hussain, A. H. Bin. (2022). Internet of Things-Building Information Modeling Integration: Attacks, Challenges, and Countermeasures. *IEEE Access*, *10*(July), 74508–74522. https://doi.org/10.1109/ACCESS.2022.3190357
25. Najafi, S. E., Nozari, H., & Edalatpanah, S. A. (2022). Artificial intelligence of things (AIoT) and industry 4.0-based supply chain (FMCG Industry). *A Roadmap for Enabling Industry 4.0 by Artificial Intelligence*, *December*, 31–42. https://doi.org/10.1002/9781119905141.ch3

26. Naseri, T. S., & Gharehchopogh, F. S. (2022). A Feature Selection Based on the Farmland Fertility Algorithm for Improved Intrusion Detection Systems. *Journal of Network and Systems Management*, *30*(3). https://doi.org/10.1007/s10922-022-09653-9
27. Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, 35–39. https://doi.org/10.1109/COMITCon.2019.8862451
28. Roy, K., Ahmad, M., Waqar, K., Priyaah, K., Nebhen, J., Alshamrani, S. S., Raza, M. A., & Ali, I. (2021). An Enhanced Machine Learning Framework for Type 2 Diabetes Classification Using Imbalanced Data with Missing Values. *Complexity*, *2021*. https://doi.org/10.1155/2021/9953314
29. Safarov, F., Basak, M., Nasimov, R., Abdusalomov, A., & Cho, Y. I. (2023). Explainable Lightweight Block Attention Module Framework for Network-Based IoT Attack Detection. *Future Internet*, *15*(9), 297. https://doi.org/10.3390/fi15090297
30. Santos, L., Gonçalves, R., Rabadão, C., & Martins, J. (2023). A flow-based intrusion detection framework for internet of things networks. *Cluster Computing*, *26*(1), 37–57. https://doi.org/10.1007/s10586-021-03238-y
31. Sokkalingam, S., & Ramakrishnan, R. (2022). An intelligent intrusion detection system for distributed denial of service attacks: A support vector machine with hybrid optimization algorithm based approach. *Concurrency and Computation: Practice and Experience*, *34*(27). https://doi.org/10.1002/cpe.7334
32. Tufail, S., Parvez, I., Batool, S., & Sarwat, A. (2021). A survey on cybersecurity challenges, detection, and mitigation techniques for the smart grid. *Energies*, *14*(18), 1–22. https://doi.org/10.3390/en14185894
33. Xia, W., Neware, R., Kumar, S. D., Karras, D. A., & Rizwan, A. (2022). An optimization technique for intrusion detection of industrial control network vulnerabilities based on BP neural network. *International Journal of System Assurance Engineering and Management*, *13*, 576–582. https://doi.org/10.1007/s13198-021-01541-w
34. Yin, L., Luo, X., Zhu, C., Wang, L., Xu, Z., & Lu, H. (2020). ConnSpoiler: Disrupting C&C Communication of IoT-Based Botnet through Fast Detection of Anomalous Domain Queries. *IEEE Transactions on Industrial Informatics*, *16*(2), 1373–1384. https://doi.org/10.1109/TII.2019.2940742
35. Yusuf, M. M., Sahrani, S., Saad, M. H., Sarker, M., & Samah, M. Z. (2022). Design and Development of An Internet of Things (IoT) Based Real Time Monitoring and Control System for Smart Indoor Hydroponic Vertical Farming System With ESP32 and Adafruit IO. *Journal of Information System and Technology Management*, *7*(28), 155–163. https://doi.org/10.35631/JISTM.728010
36. Zakaria, M. S., Abdul Ghani, A. T., Yahya, M. S., & Jamali, S. N. (2020). Information Technology Risk Management for Water Quality Monitoring IoT Infrastructure: A Case Study at Tasik Chini Unesco Biosphere Reserve. *Asia-Pacific Journal of Information Technology & Multimedia*, *09*(02), 94–102. https://doi.org/10.17576/apjitm-2020-0902-07

37. Zhou, Q., Li, R., Xu, L., Nallanathan, A., Yanga, J., & Fu, A. (2022). *Sufficient Reasons for A Zero-Day Intrusion Detection Artificial Immune System*. http://arxiv.org/abs/2204.02255

Figures

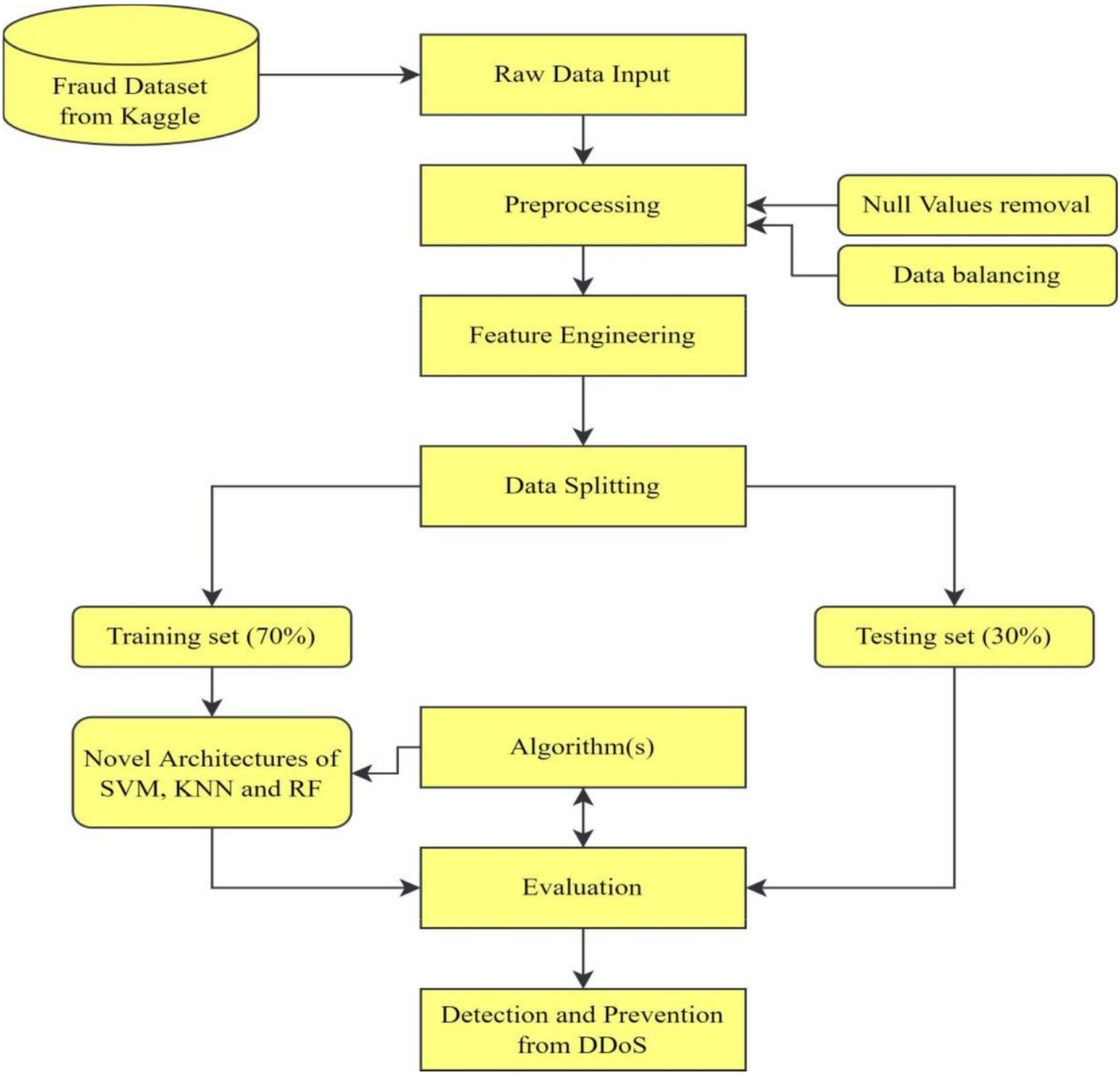

Figure 1

Proposed ML Models Dataset Flow Chart.

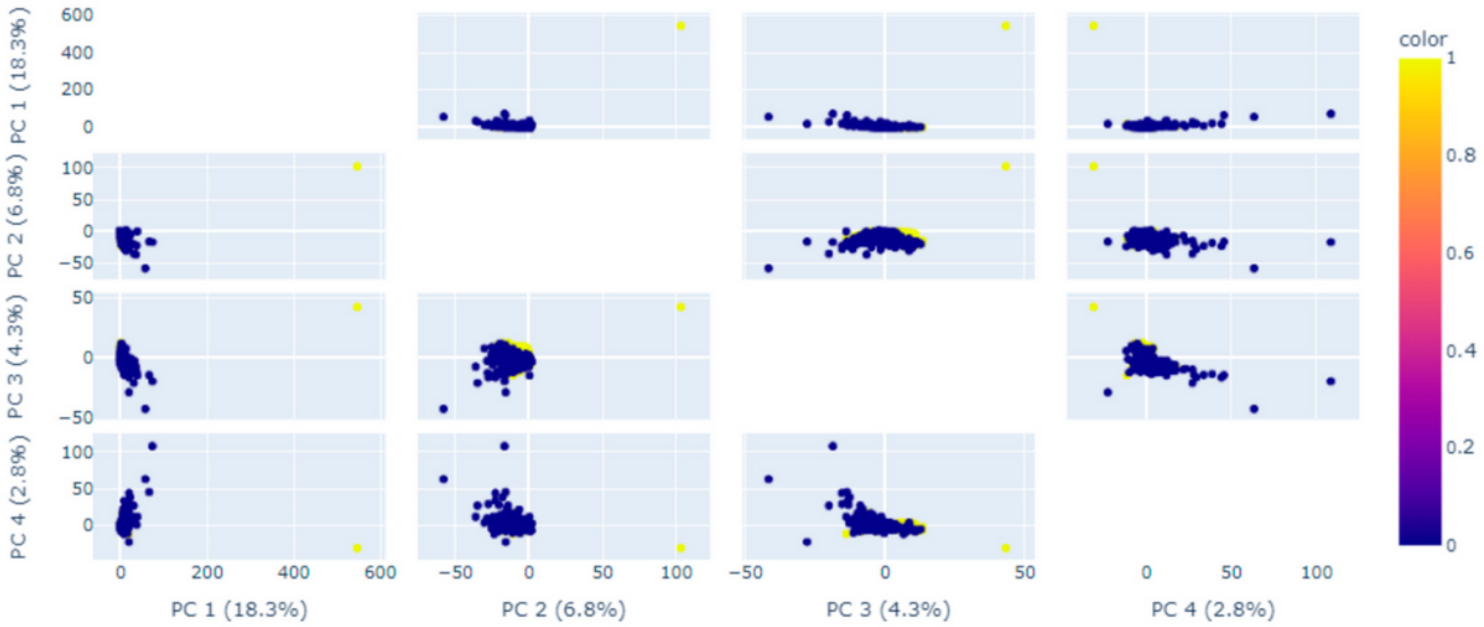

Figure 2

The Repartition of Services

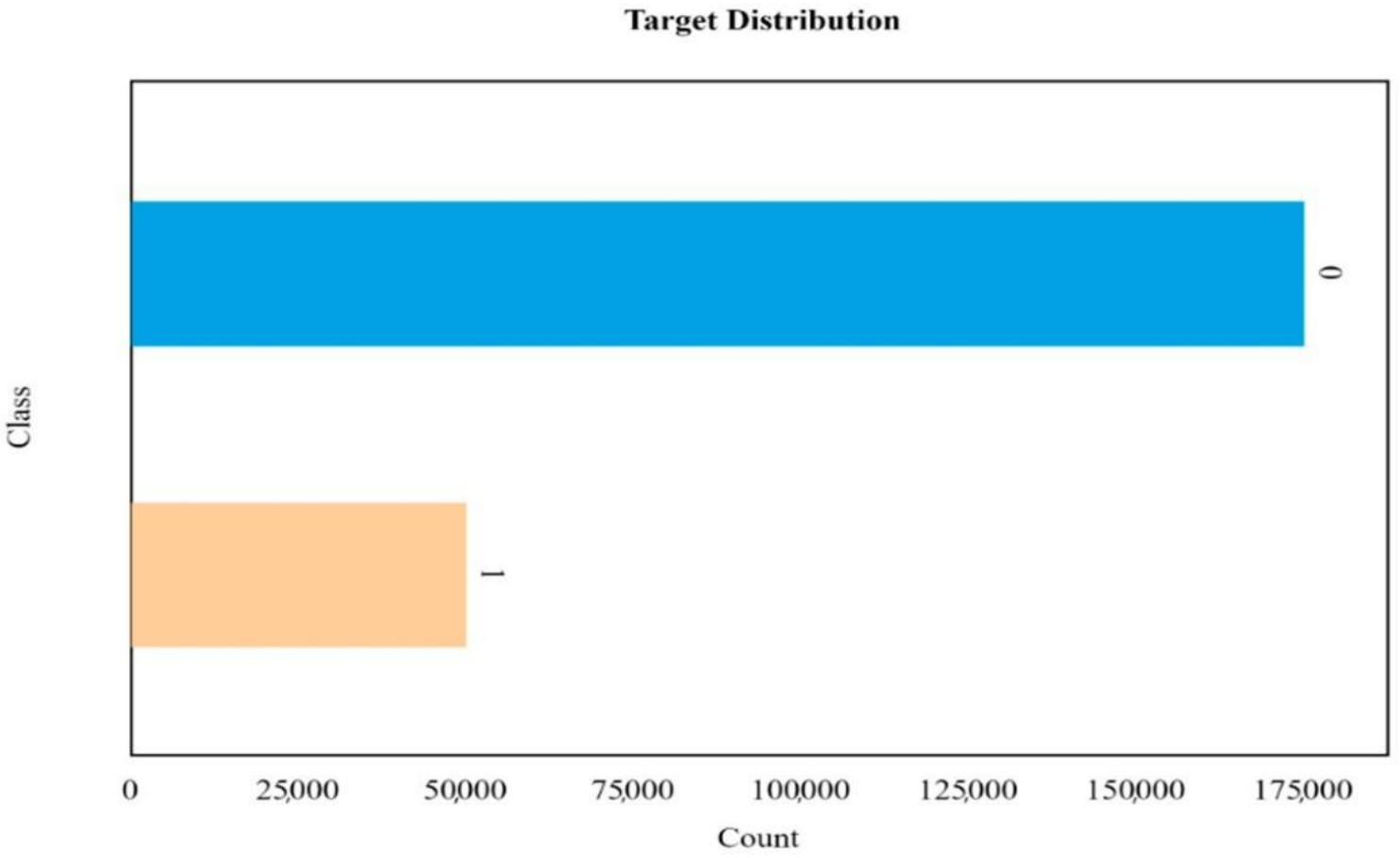

Figure 3

Target Distribution of Bank Dataset

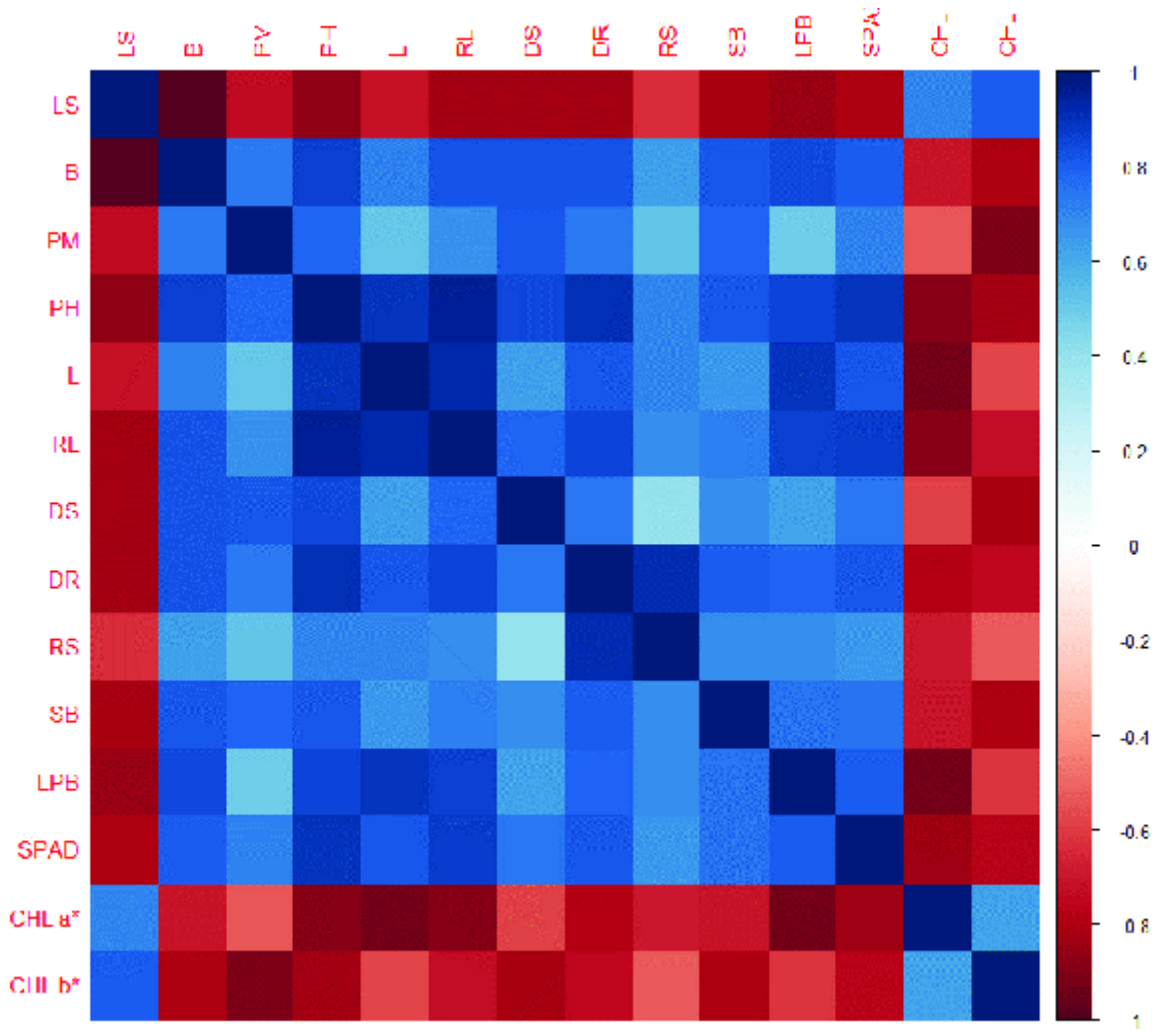

Figure 4

The Feature Correlation Matrix

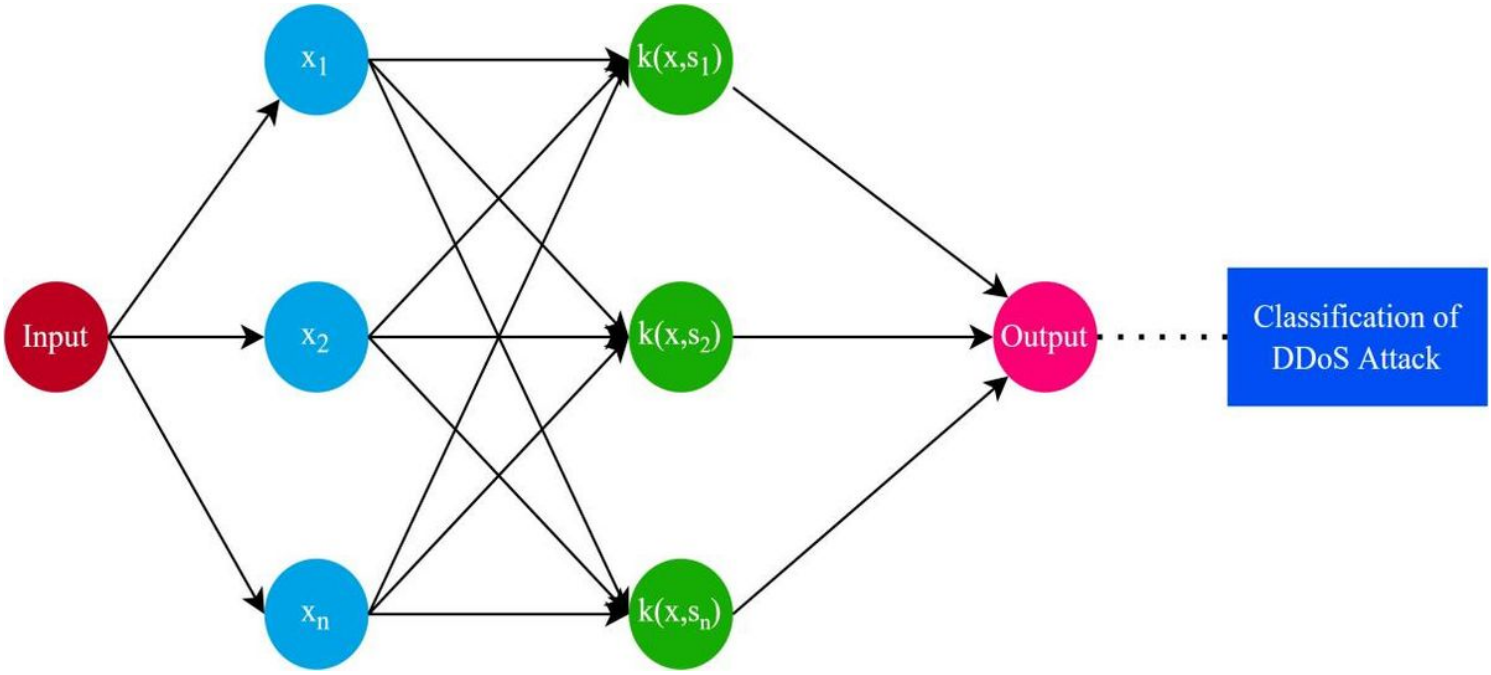

Figure 5

An Overview of the Conceptual Framework of an SVM Model for Identifying DDoS Attacks

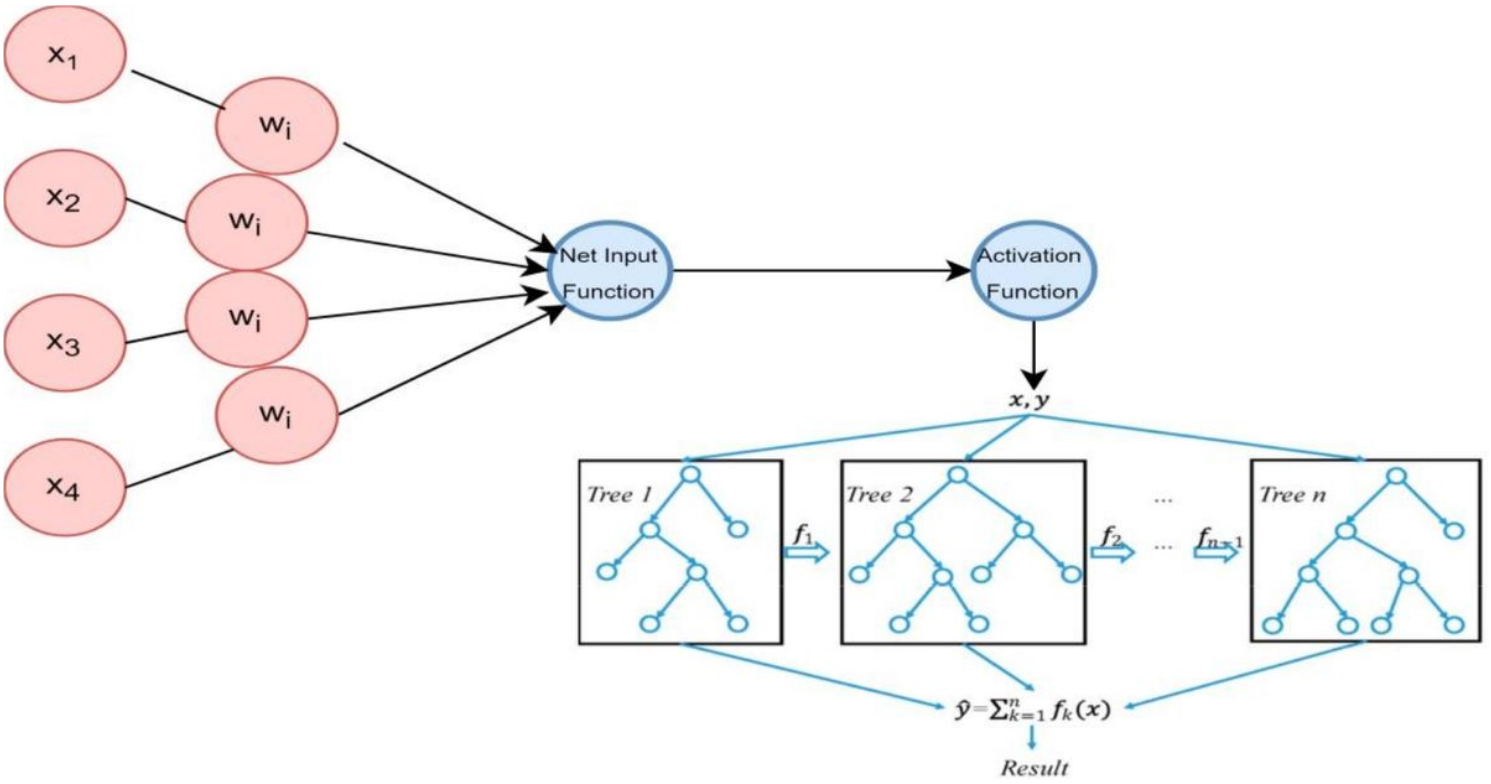

Figure 6

RF DDoS Classification Model Structure

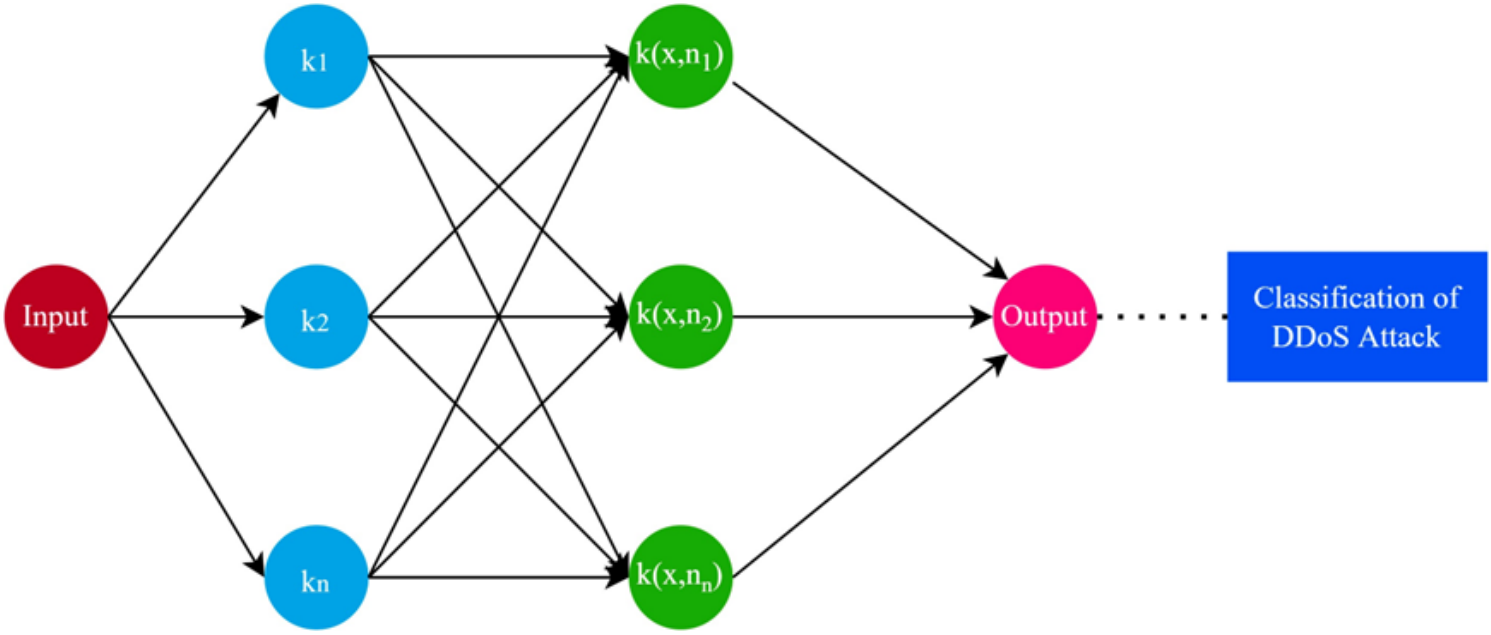

Figure 7

DDoS Attack Classification Using the KNN Model Block Schematic.

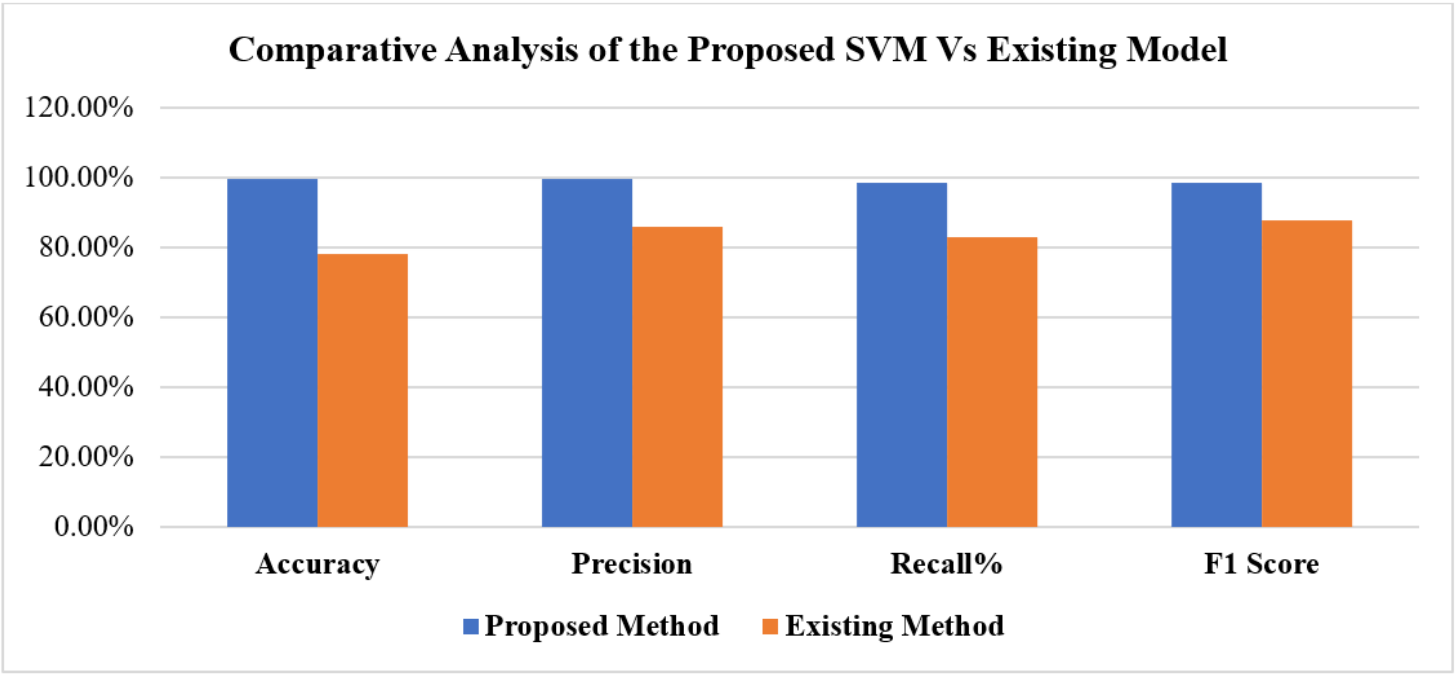

Figure 8

Comparative Analysis of the Proposed SVM Vs Existing Model

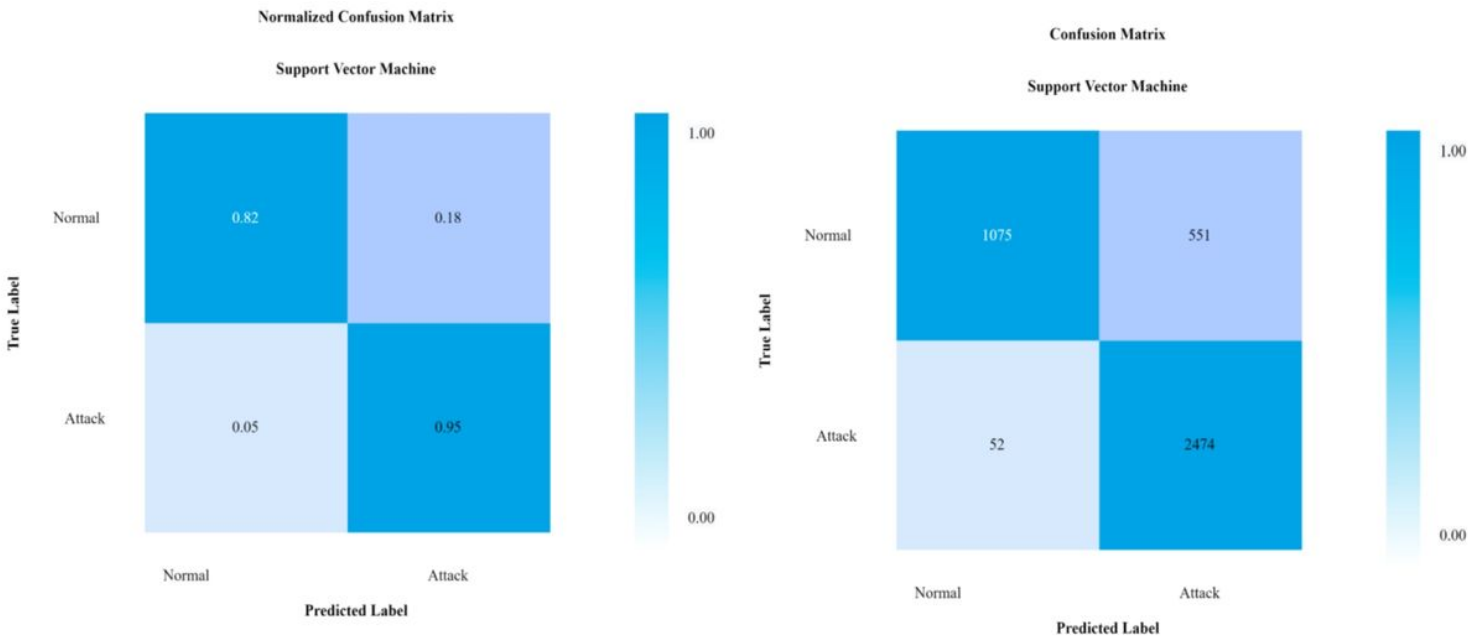

Figure 9

Confusion Matrix of Normalized and Non-Normalized SVM Model.

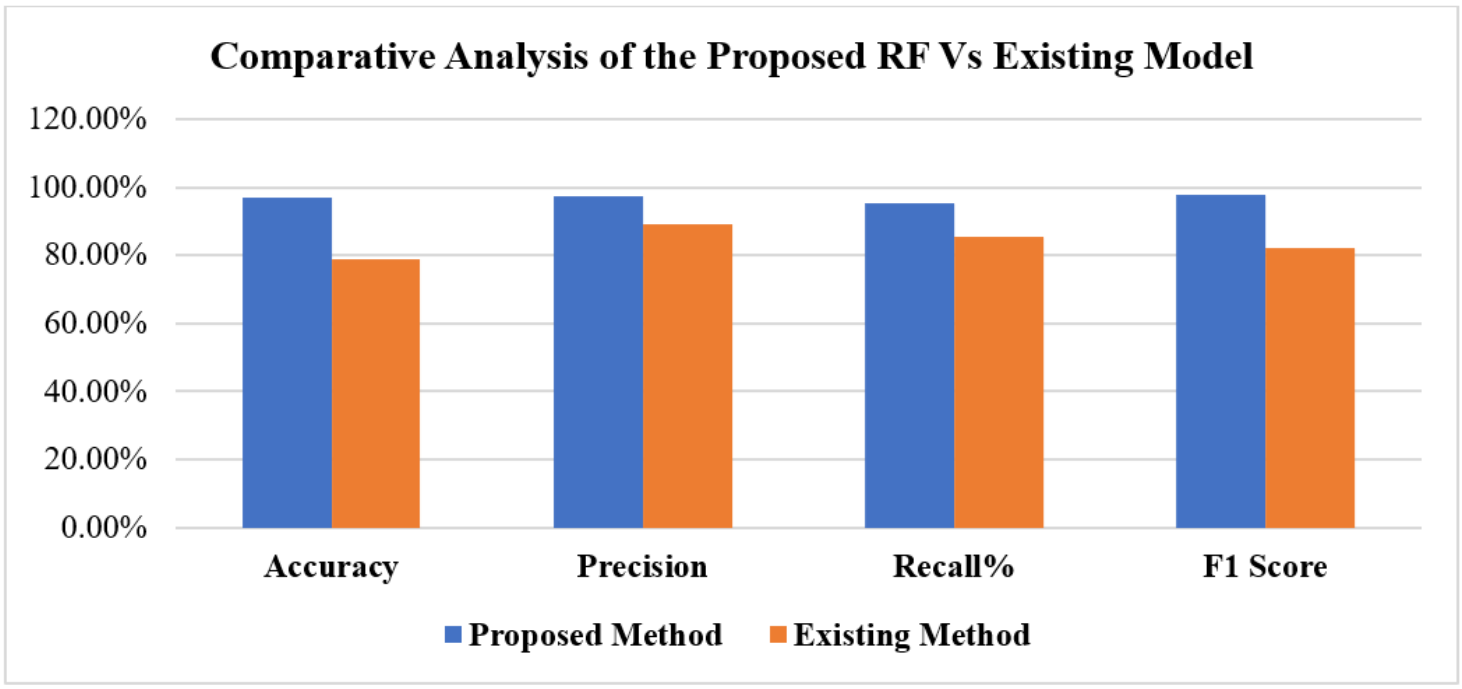

Figure 10

Comparative Analysis of the Proposed RF Vs Existing Model

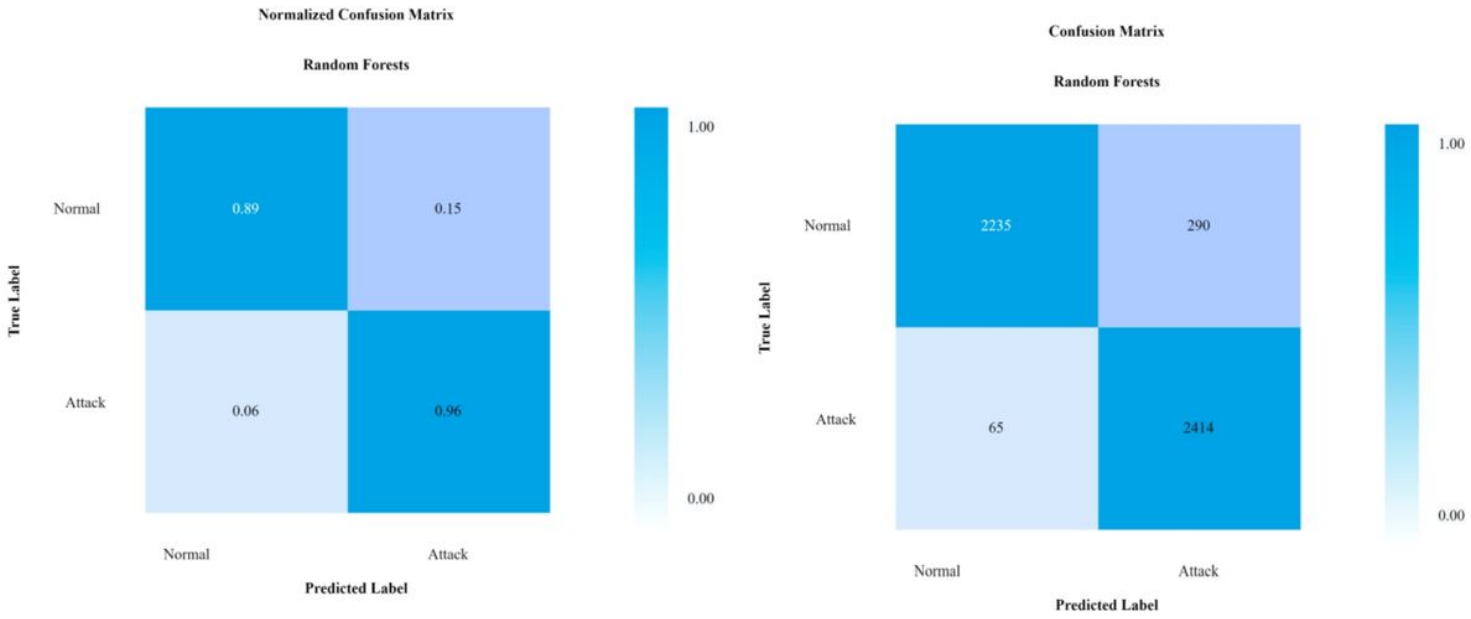

Figure 11

Confusion Matrix of Normalized and Non-Normalized RF Model

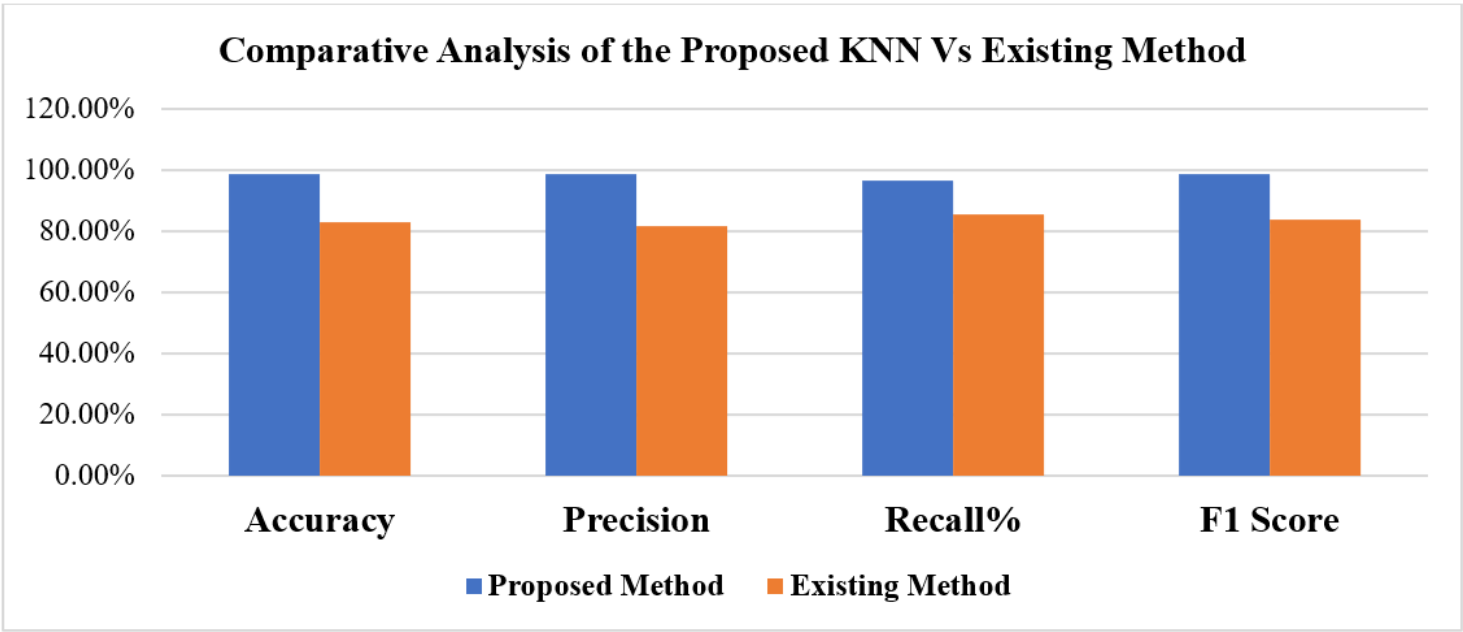

Figure 12

Comparative Analysis of the Proposed KNN Vs Existing Method

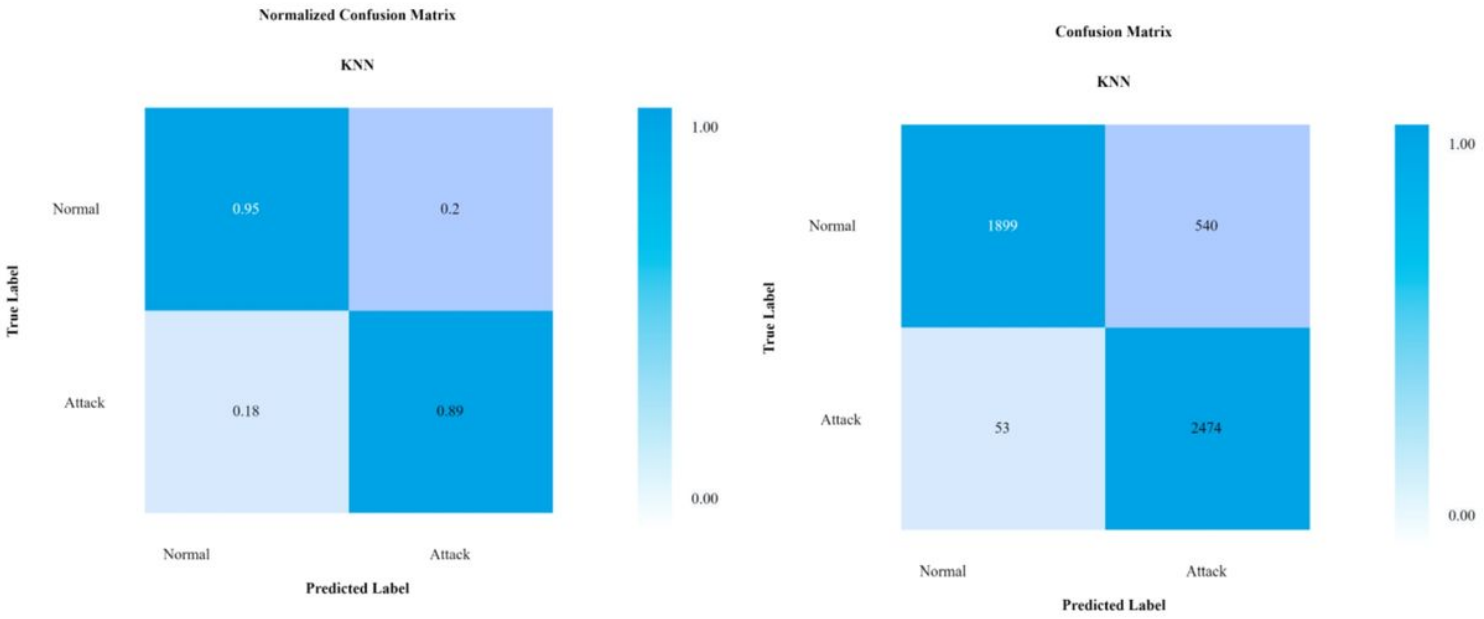

Figure 13

Confusion Matrix of Normalized and Non-Normalized KNN Model