sustainability

MDPI

Article

Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models

Umar Islam [1], Ali Muhammad [2], Rafiq Mansoor [3], Md Shamim Hossain [4], Ijaz Ahmad [5,*], Elsayed Tag Eldin [6], Javed Ali Khan [7], Ateeq Ur Rehman [8] and Muhammad Shafiq [9,*]

[1] Department of Computer Science, IQRA National University, Swat Campus, Swat 19220, Pakistan; umar.koh@gmail.com
[2] Institute of Management Studies, University of Peshawar, Peshawar 25000, Pakistan; alimohmand@uop.edu.pk
[3] Department of Mechanical Engineering, International Islamic University Islamabad, Islamabad 44000, Pakistan; rafiq.mansoor@iiu.edu.pk
[4] Department of Marketing, Hajee Mohammad Danesh Science & Technology University, Dinajpur 5200, Bangladesh; shamim.mkt@hstu.ac.bd
[5] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518040, China
[6] Electrical Engineering Department, Faculty of Engineering & Technology, Future University in Egypt, New Cairo 11845, Egypt; elsayed.tageldin@fue.edu.eg
[7] Department of Software Engineering, University of Science & Technology, Bannu 28100, Pakistan; engr_javed501@yahoo.com
[8] Department of Electrical Engineering, Government College University, Lahore 54000, Pakistan; ateqrehman@gmail.com
[9] Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea
* Correspondence: ijaz@siat.ac.cn (I.A.); shafiq@ynu.ac.kr (M.S.)

check for updates

Citation: Islam, U.; Muhammad, A.; Mansoor, R.; Hossain, M.S.; Ahmad, I.; Eldin, E.T.; Khan, J.A.; Rehman, A.U.; Shafiq, M. Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models. *Sustainability* **2022**, *14*, 8374. https://doi.org/10.3390/su14148374

Academic Editors: Bhavik Bakshi and Andreas Kanavos

Received: 26 March 2022
Accepted: 29 June 2022
Published: 8 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Cyberattacks can trigger power outages, military equipment problems, and breaches of confidential information, i.e., medical records could be stolen if they get into the wrong hands. Due to the great monetary worth of the data it holds, the banking industry is particularly at risk. As the number of digital footprints of banks grows, so does the attack surface that hackers can exploit. This paper aims to detect distributed denial-of-service (DDOS) attacks on financial organizations using the Banking Dataset. In this research, we have used multiple classification models for the prediction of DDOS attacks. We have added some complexity to the architecture of generic models to enable them to perform well. We have further applied a support vector machine (SVM), K-Nearest Neighbors (KNN) and random forest algorithms (RF). The SVM shows an accuracy of 99.5%, while KNN and RF scored an accuracy of 97.5% and 98.74%, respectively, for the detection of (DDoS) attacks. Upon comparison, it has been concluded that the SVM is more robust as compared to KNN, RF and existing machine learning (ML) and deep learning (DL) approaches.

Keywords: machine learning; support vector machine; distributed denial-of-service

1. Introduction

An increase in ransomware attacks of 1318 percent in the first half of 2021 was disproportionately felt by the banking sector [1]. New COVID-19 opportunities for threat actors may be behind the 4 percent rise in business email compromise (BEC) attacks [2]. Banks are increasingly vulnerable to large-scale cyberattacks. A financial institution's solvency could be jeopardized by a cyberattack on another bank because of the banks' interconnectedness. State-sponsored cyberattacks on U.S. banks are particularly vulnerable [3].

Cybercrime has been steadily increasing over the years as more people use the internet and mobile banking. Types of fraud ranging from credit card scams and spamming to ATM robberies and identity theft are just some examples of cybercrime incidents [4]. Because of the high monetary value of the data it stores, the banking industry is particularly at risk. There are a number of ways that hackers can profit from the stolen financial information and banking credentials. As banks' digital footprints have grown, so too has the attack surface that could be exploited [5].

Cyberattacks can cause power outages, military equipment malfunctions, and leaks of classified information. They can lead to the theft of sensitive data, such as medical records, which can be extremely valuable. They can disrupt phone and computer networks, causing data to be unavailable, or paralyze systems [6].

As a result of the high value of the data it houses, banking is particularly vulnerable. There are a number of ways that hackers can profit from the stolen financial information and banking credentials. The machine learning models used address the above informative issues. ML/DL model play an important implemented different dataset such as biomedical [7–10], agriculture [11,12] and specially IoT dataset [13].

The main contributions of this research are as follows:

- This study proposed efficient machine learning model (SVM, KNN, and RF) for the classification of DDOS attacks using Banking Dataset based on their excellent performance. Further, no prior research has ever compared or used these (SVM, KNN, and RF) three approaches of DDoS attacks detection.
- We investigated the training parameters' influence on the classification accuracy (%), pres (%), recall (%) and F1-score (%) and time complexity (ms).
- To check the evaluation performance of ML models, we compared SVM, KNN, and RF in order to find out the most efficient model. The comparative result indicates that SVM is more robust as compared to KNN, RF and existing machine learning methods (ML/DL).

This research is divided into 5 sections. Section 1 shows the introduction of this study; Section 2 shows the related work. Section 3 shows the methodology of this research, while Section 4 shows the results of this study. At the end, our conclusion has been shown in Section 5.

2. Related Work

Previously, DDoS has been analyzed in many studies; we are detecting here Middlebox DDoS, which is a very severe and new type of DDoS. Its type is an in-network device that serves as a monitor, filter or transformer between two interacting hosts. Network devices, such as routers and switches, can only examine the header of a message and not the content of it, whereas middleboxes use Deep Packet Inspection (DPI) to examine both. Middleboxes have been used for a wide range of network functions, including firewalls, since their introduction. Censoring middleboxes are often installed at the nation's borders (or within the nation's ISPs) and are commonly implemented at huge scales to monitor all traffic passing through the censoring nation-state. In unencrypted traffic, DNS requests, or TLS server name indication (SNI) fields, censoring firewalls often identify prohibited keywords or domains that must be blocked. It is possible for a censoring middlebox to block connections in various ways, including by blocking packets, injecting RST packets to break the connection, or by injecting block pages in response to forbidden HTTP requests. Once a middlebox determines a connection should be blocked, it can do so in a variety of ways. When packets are re-ordered or lost, middleboxes often track the content of connections over numerous packets. Middleboxes, on the other hand, may not be able to see packets travelling in both ways. As a result, packets between two end hosts may take distinct routes across the Internet. As a result, a middlebox may only be able to see one end of a TCP connection e.g., (the packets from client to server). Even while middleboxes don't have access to all the packets in a connection, they can use TCP reassembly techniques to block connections even when they don't see all of them. Because of the middleboxes' tolerance to missing packets, attackers have an opening: a reflective attacker may be able to convince the middlebox that the three-way handshake has been completed without really completing it. As a result of the packets they send, middleboxes could be excellent targets for reflected amplification, especially for block pages. After that, we show how middleboxes can be hoodwinked and demonstrate how middleboxes can provide enormous amplification factors.

Chayomchai et al. [14] are looking into how cybercrime has impacted banking institutions and what steps have been taken to counteract those effects. The most recent victims have been banks. Massive malware assaults routinely target Indian banks, resulting in the theft of critical and private information and substantial financial losses. According to this study, the most vulnerable parts of a firm to cyber-attacks should be identified, and a customised cyber-security strategy should be developed to protect them. The research includes secondary data analysis of government websites, papers, and research studies, as well as case studies of previous cyber threats and crimes that have resulted in significant financial losses. To help banks, financial institutions, and the general public better comprehend the cyber regime, this study will be conducted. Poorly supervised models that do not require annotated data to estimate the impact of denial-of-service (DoS) attacks can be developed using Latent Dirichlet Allocation (LDA) and symmetric Kullback-Leibler divergence on tweets. There is a limit in the module that is only loosely supervised. Because fewer non-attack events on Twitter are likely to be mistaken for DoS attacks in the pre-specified detection window, this will become less of a problem. Another alternative is to use an extra classification layer, trained on manually annotated DoS attack tweets, to remove non-attack tweets from the dataset. Weakly-supervised learning approaches can be used to generate models that are precise and generalizable within the same industry [15,16].

Using Alimolaei et al. [16] though smart technology, an online bank customer can be identified if they are behaving abnormally. System designers used the fuzzy theory to take into consideration that users' activities are accompanied by some degree of uncertainty. The fuzzy expert system's performance was evaluated using a receiver operating characteristic curve, and the results show a 94% accuracy rate. E-banking security and service quality could benefit from the deployment of this expert system. To begin with, Refs. [17,18] describes the many cyber threats involved with online banking. Additionally, it provides an approach to cyber-banking security that focuses on protecting the borders of the application as well. Peripheral and application security are two different approaches to securing a system's infrastructure.

E-banking transactions can now be checked for potential fraud using a novel approach developed by Salem et al. [19]. A model with scoring criteria for online real-time transactions and offline historical transactions is combined for the goal of identifying fraud. Large-scale data processing: a framework for this using Kafka, Spark, and MPP Gbase, a method for analyzing large transaction logs is described. The author's experimental results show that their proposed technique is effective over a large dataset of electronic banking transactions. Future research by the author should fill in these gaps and resolve these difficulties. Cybercrime datasets are investigated and accessible problems are identified by Ref. [17], Ref. [20] using K-Means, Influenced Association Classifier and J48 Prediction Tree. Influenced Association Classification uses a clustering approach called K-Means. K-means classifiers can mine the record and make predictions about cybercrime using the J48 technique, which uses K-means selection to select the initial centroids. Bank cybercrime can be predicted better and more accurately utilizing information obtained via K-Means, Influenced Association Classifier, and J48 Prediction Tree combined. Author's law enforcement authorities must be well-equipped to combat and prevent cybercrime.

The authors of [18,21,22] described the challenges faced by a number of banks and card-based enterprises. It is necessary to investigate the issue in depth in order to come up with a solution that is both practical and effective. By sharing information, you can help keep a bank safe from cyber-attacks. Avoid allowing too many inquiries from the same source or user session at once [23,24]. The majority of automated attack sources request web pages faster than normal users. It's important to protect the network and its applications from DDoS attacks. Many DDoS attacks employ network strategies such as spoofing, fragmented packets, or failure to finish TCP handshakes. Attacks at the application level try to use up all of the server's resources. Anti-malware efforts can be circumvented by detecting abnormal user activity and employing documented application attack signatures. DDoS attacks can be detected by looking for well-known patterns or signatures. HTTP requests that don't conform to the protocol's requirements are also common in DDoS attacks. The Slowloris attack is known for its use of repetitive HTTP headers. It is possible for a DDoS client to try to access pages that are not there. Attacks may also be the source of a web server problem or slow response time.

Despite the numerous defense mechanisms in place, the authors of [25,26] have found that bandwidth availability and computer resource security remain problematic. Due to a surge in valid traffic and its resemblance to attack traffic, the DDoS problem became more urgent. According to this study, T-CAD, a distributed attack detection system, can be used on autonomous system routers to identify and mitigate the effects of DDoS attacks. For example, T-CAD can tell the difference between legitimate traffic, DDoS attacks, and flash events using the normalized router entropy. The proposed attack detection system has been proved to function on OMNeT++ and INET in tests so far. The T-CAD DDoS defense system has outperformed many existing thresholds and entropy-based DDoS detection approaches in simulation testing. In the research introduced by [27,28], DDoS attack's various models, as well as a timeline of defense measures and improvements to counter them, are all reviewed. Using the MapReduce programming architecture, a new DDoS attack detection mechanism has been built. As per the statement of the Internet banking platform, the procedures for authenticating customers differ. Several banks make use of passwords and PINs. Others use TANs and TAN lists to verify and authorize transactions (also known as scratch lists). Users can also be authenticated using one-time passwords or challenge-response systems, which are more complicated. As far as the author is aware, no major bank has successfully adopted public-key certificates for user authentication. The cryptographic strength of the SSL/TLS protocol is often used when advocating for the safety of online banking. Only a few theoretical faults and holes have been found in the SSL/TLS protocol's security. These investigations are related to the Dolev-Yao threat model. It's possible for an adversary to control the communication path between a client and server, but the end points of the channel remain secure. This misrepresents the realities of how an attacker might harm a customer.

For online banking fraud prevention, Mehmood and colleagues [29] employ a Hidden Markov Model (HMM). As a result, the bank's system has developed a one-time password that is sent directly to each customer's registered cell phone, ensuring that only legal transactions are rejected. Banks are implementing fraud detection and prevention technologies in an effort to avert enormous losses. Cutting-edge fraud technology is being used to detect and stop fraudulent Internet banking transactions from taking place at financial institutions around the world. The issue is that they are unable to efficiently identify and track legitimate users. As a solution, the author recommends using a Hidden Markov Model. To demonstrate the numerous attack methods used by cyber thieves against chosen Indian banks. In the research [30,31] has attempted to demonstrate how spoofing, brute force attacks, buffer overflow and cross-side scripting are all linked to Indian public and private sector banks. Cyber-attacks such as online identity theft [31], hacking [32] and harmful code [31,32]; DOS attack and credit card/ATM frauds [33] are also linked to Intruder Detection, as is System Monitoring.

Blockchain-based DDoS mitigation is a promising and viable approach. It is possible that the intrinsic and essential properties of blockchain, such as decentralization and internal and external trust lessness, immutability, verifiability and anonymity might combat this deadly cyber risk. In several businesses, we think there is no need for citation in such statements looking at how DDoS mitigation using blockchain technology is faring. In this research, we will take a close look at a number of strategies, focusing on their merits, drawbacks, and limits. A unified platform for learning about current strategies will be beneficial to DDoS mitigation research and development.

Collaborative DDoS attack detection methods that take into account the detection performance in different time zones is used to identify DDoS attacks on several networks more accurately. To get at a final conclusion for individual assaults on each of these networks, the detection and false positive rates for each network are weighted according to its time zone. To determine if a DDoS attack has taken place, [22] recommends using weighted detection data. It was found that the proposed technique reduces false positives by 35% while maintaining an extremely high detection rate. Yet to be identified and described are the ideal prerequisites for a protective solution despite this work's goal of identifying and describing those criteria [34,35]. All HTTP-based DoS and DDoS assaults have been thoroughly examined in this study, which aims to define and describe the ideal criteria for a protective architecture against these attacks in-depth. Several academics have developed various DDoS detection methods based on information theory entropy and divergence measures in the past [24]. Research suggests using a novel LeCam divergence measure based on flow similarity across network traffic flows to detect distinct types of DDoS attacks. Experiments on the datasets provided by MIT Lincoln and CAIDA show that the proposed techniques may be used effectively; according to the findings, the LeCam Divergence metric outperforms the more traditional Kullbeck-Leibler, Bhattacharyya, and Pearson Divergence measures.

DDoS attacks can be classified as either benign traffic or DDoS assault traffic by using a unique architecture that combines a well-posed sparse Auto Encoder (AE) for feature learning with a Deep Neural Network (DNN) for classification [36]. Assault detection is made easier by adjusting the parameters of DNN and AE in a way designed for this goal. To avoid overfitting, the author of this article explains how to reduce reconstruction error, remove gradient inflating or disappearing, and develop a more compact network with fewer nodes [37]. Performance criteria such as detection accuracy, precision, recall, and F1-Score were used to compare the proposed approach to ten current best practices. A series of tests on the CICIDS2017 and NSL-KDD standard datasets have been carried out in order to verify the results. The proposed method outperforms existing methods.

For decades, DDoS attacks have degraded network availability, and no effective security measure has yet been established to combat them. New defenses against DDoS attacks are now possible due to the growth of Software Defined Networking technology. Two methodologies have been developed in [37] to identify DDoS attacks. One technique is to identify the source of a DDoS assault to measure the intensity of the attack. To find the DDoS assault, the K-Nearest Neighbors (KNN) algorithm is developed using Machine Learning (ML). The author's proposed algorithms outperform those of other researchers in detecting DDoS attacks when tested on real-world datasets. If the system's security defenses are deceived by someone with valid access to the system, an insider attack is more likely [35]. Anti-DDoS assaults can be prevented by using EDIP (Early Detection and Isolation). EDIP is able to locate an insider among all of the system's legitimate clients by forwarding it to an attack proxy. As a result, a new algorithm has been developed to increase attack isolation while minimizing disruption to innocent customers. Using the load balancing approach, proxies can avoid being overloaded. Researchers have developed spectral gene set filtering (SGSF), a new method for filtering gene set collections prior to gene set testing in order to overcome the statistical power challenge posed by large gene set collections [4,35–37].

An important part of traditional IDS is blacklisting. These approaches are too time-consuming and ineffective for intruders to use. Machine learning and deep learning models have made it possible to automate and configure IDS such that they can be dynamic [37]. As a reminder, the model's effectiveness is strongly dependent on the training data that was used. There are a number of datasets that are regularly utilized in IDS research. Individual DDoS attack datasets have received very little attention. Huang et al. [37] investigates the feasibility of initiating an IoT-based DDoS attack at a cheap cost. In the first place, a new DDoS attack architecture is proposed. This design is ideal for resource-constrained DDoS attackers due to its negligible management costs, high detectability, and excellent robustness. It is simplified to a variation problem where the objective functional represents the predicted impact of the DDoS attack associated with a DDoS attack method in this architecture and is based on a novel botnet growth model [38]. Finally, the variation problem for three alternative DDoS defense systems has been solved. DDoS assaults based on IoT are now more understood as a result of this study.

3. Materials and Methods

This section describes the dataset, methodology, and performance indicators in depth. Figure 1 represents the proposed work of the research, furthermore, we have used data from an open-source platform (accessed on 2 February 2021) that contains data of Distributed Denial of Service (DDOS) attacks. Data collected from the public source which was raw dataset. We have preprocessed the dataset by using preprocessing techniques. Null values have been removed from datasets and then balancing techniques were applied to scale and balance the dataset. After extracting the best features, we have split the data into 70% training and 30% testing sets. The training set is used to train the ML models and the testing set is used to evaluate the models.

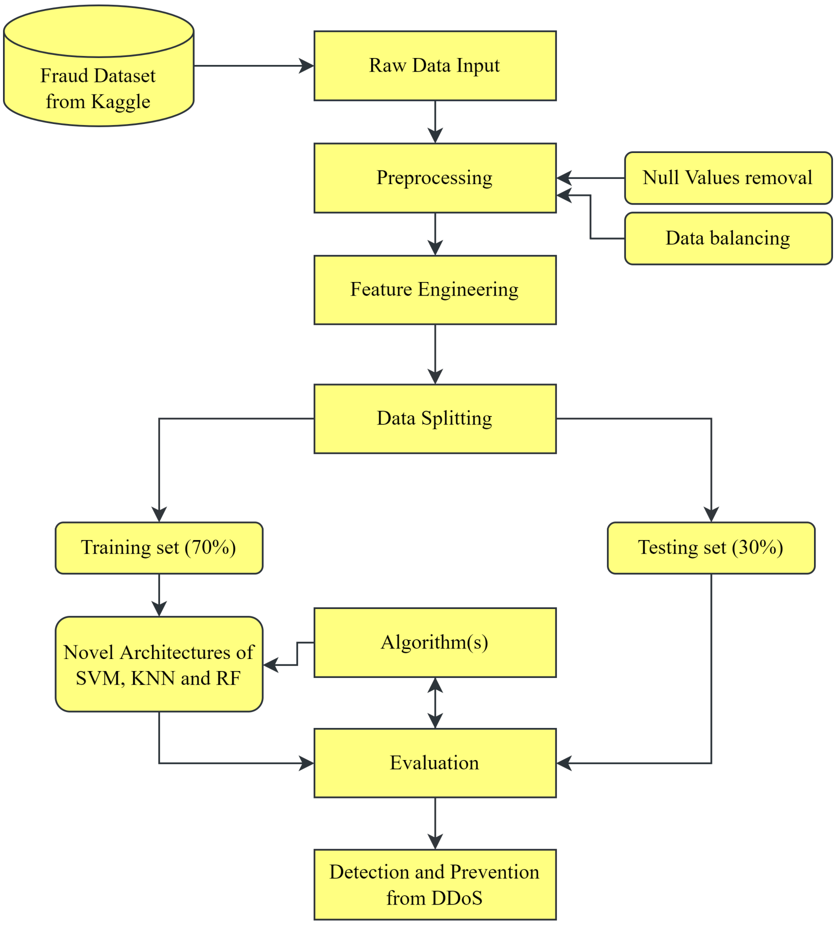


Figure 1. Flow chart of the Proposed ML models using Bank dataset.

3.1. Dataset Description

3.1.1. Fraud Detection Dataset

The Banking Dataset monitors network intrusions. This dangerous malware also includes DoS. Table 1 and Figure 2 exhibit the dataset attributes and their descriptions.

Figure 2 shows the repartition of services from 4 PCs in a bank. The heat map shows that if the value in the repartition of services from a PC is greater than 0.5, then there is a chance of severe DDoS attack. When the value from a PC is less than 0.5, there will be a lesser chance of an attack. Figure 3 shows the total counts of target class distribution in dataset. DDoS attacks occur 50,000 times in the dataset.

Table 1. The Bank dataset of features description.

Feature/Attribute	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 Here we have used DDoS attacks, if the label shows 0, there will be no attack, if label will be 1, there will be DDoS attack.	Output/Target Variable with Nine Classes

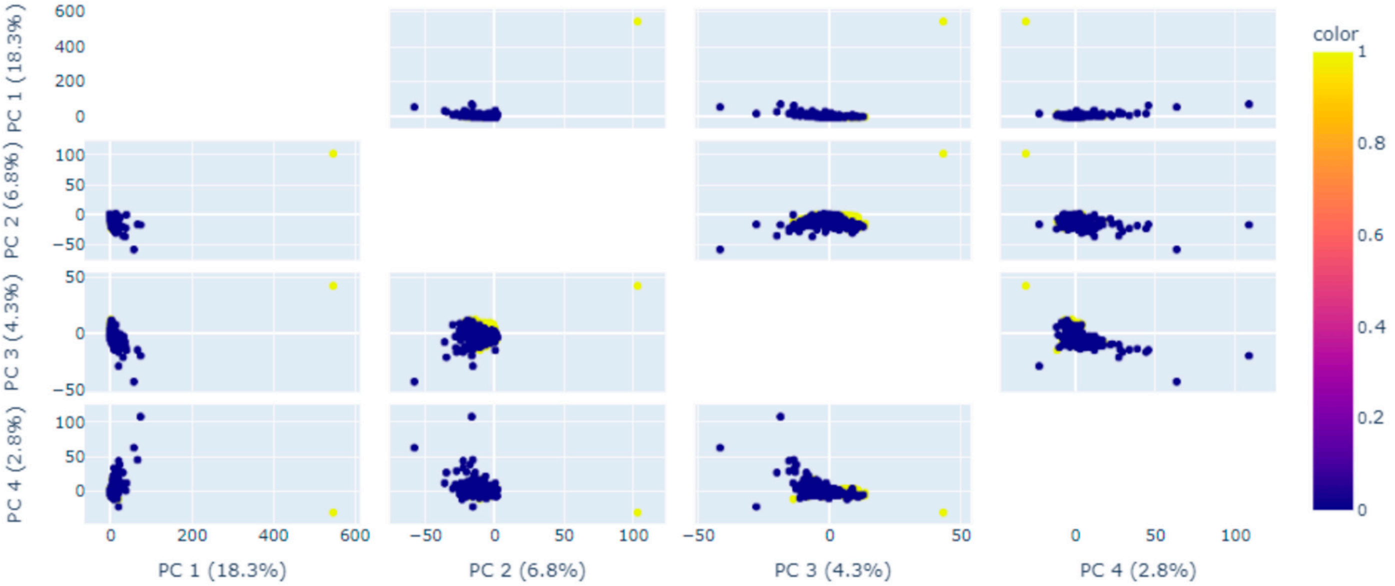


Figure 2. The repartition of services.

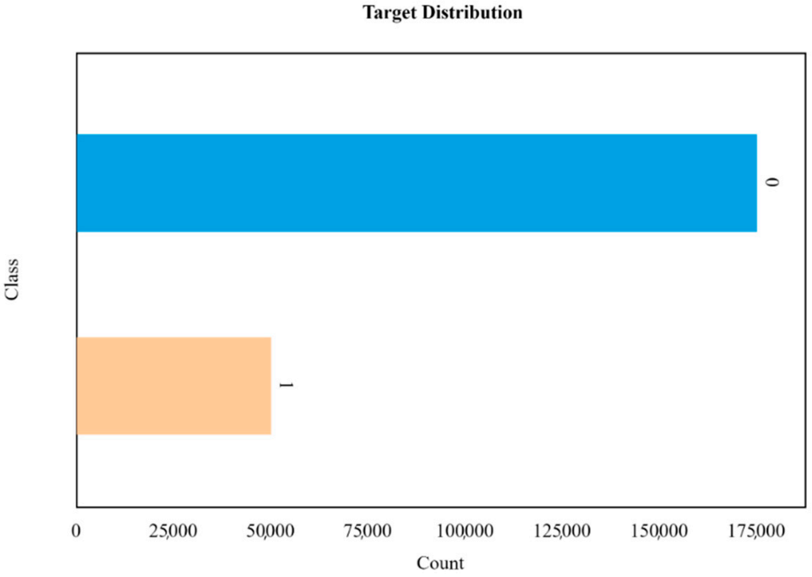


Figure 3. Target distribution of Bank dataset.

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 sets of data. 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% of the data for training.

3.1.2. Data Preprocessing

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

(a) *Removal of Socket Information*

In order to remove any potential bias in the identifying procedure, IP addresses of both the source and destination must be deleted. Instead of relying solely on data from a single socket, it can rule out hosts with similar packet information by examining packet characteristics.

(b) *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 classes.

(c) *Label Encoding*

Encoding the labels into a numeric form so that they may be read by a computer is called label encoding. As a result, machine learning algorithms can make better decisions about how to use those labels. In supervised learning, it's a key step in preprocessing the structured dataset.

(d) *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.

(e) *Feature Ranking*

Figure 4 presents the feature correlation matrix. Col_0 to Col_111 shows the number of features, because of the large number of characters and strings, we have used the label as 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 rank according to the weight and check the relevance of each feature in the outcome:

3.2. Machine Learning Models

Supervised Learning is an algorithm for classifying new observations based on previously learned data. Classification uses a dataset or set of observations to categorise fresh data into one of several categories. We used KNN, SVM, and Random Forests to classify DDOS attacks in the Banking Sector Dataset.

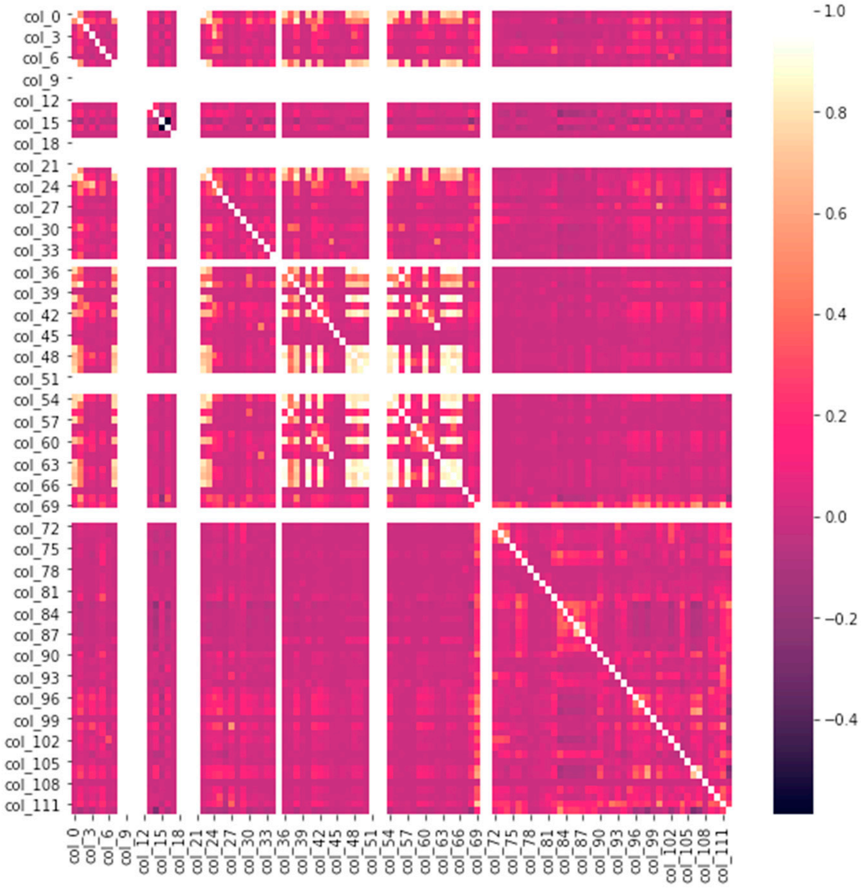


Figure 4. Presents feature correlation matrix.

3.2.1. Support Vector Machine

Classification, regression, and outlier detection can all be accomplished with the help of supervised learning techniques known as support vector machines (SVMs). There are numerous benefits to using support vector machines. Useful in high-dimensional environments, even if the number of dimensions exceeds the number of samples, the method is still effective. In this study, we are classifying attacks by using novel architecture of the SVM Classifier. Its general architecture is shown in Figure 5. The input layer contains the vector input signal (x). It is computed between the input signal vector (x) and the support vector (s) in the hidden layer (y). The output neuron sums the linear outputs O of the hidden layer neurons. DDoS attack detection is equivalent to the two-classification problem; we use the SVM algorithm characteristics, collect switch data to extract the characteristic values to train, find the optimal classification hyperplane between the normal data and DDoS attack data, and then use the test data to test our model and get the classification results.

$$O = \sum W_i k(x_i s_i) \tag{1}$$

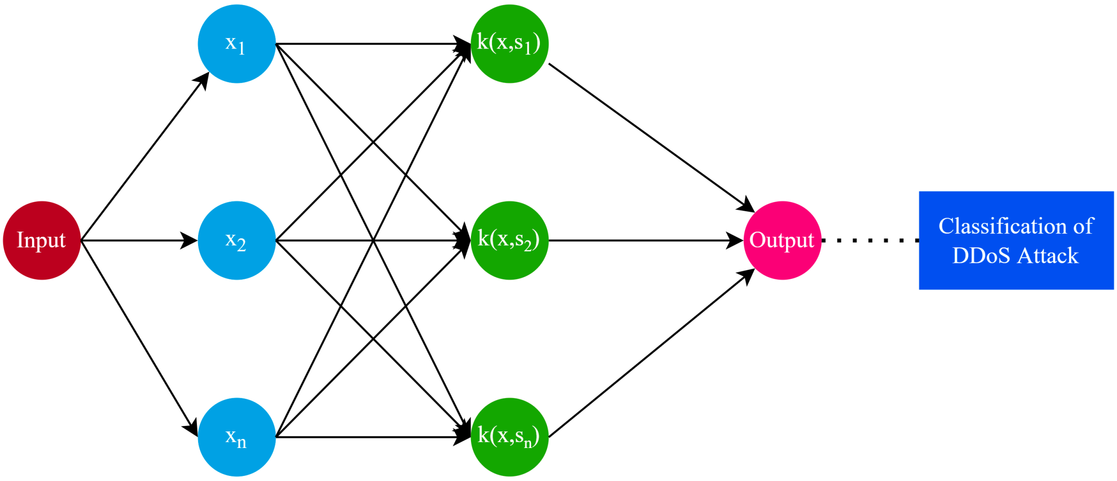


Figure 5. Basic architecture of SVM model for classification of DDoS attacks.

3.2.2. Random Forests

Classification and regression tasks can benefit from the use of random forests or random decision forests because they are an ensemble learning method that builds many decision trees at once. However, the accuracy of random forests is lower than that of gradient-boosted trees despite the fact that they outperform the former. In this study, random forests have been used for DDOS attack detection. In this study, we are classifying attacks by using architecture of RF Classifier. Its general architecture is shown in Figure 6. This model has been developed by assembling the logistic regression model into RF Classifier to improve both models' accuracy. The mathematical model is as follows:

$$y = \sum_{k=1}^{n} f(x) \tag{2}$$

$$\ln \frac{P}{1-P} = a + by \tag{3}$$

$$\frac{P}{1-P} = e^{a+by} \tag{4}$$

$$P = \frac{e^{a+by}}{1+e^{a+by}} \tag{5}$$

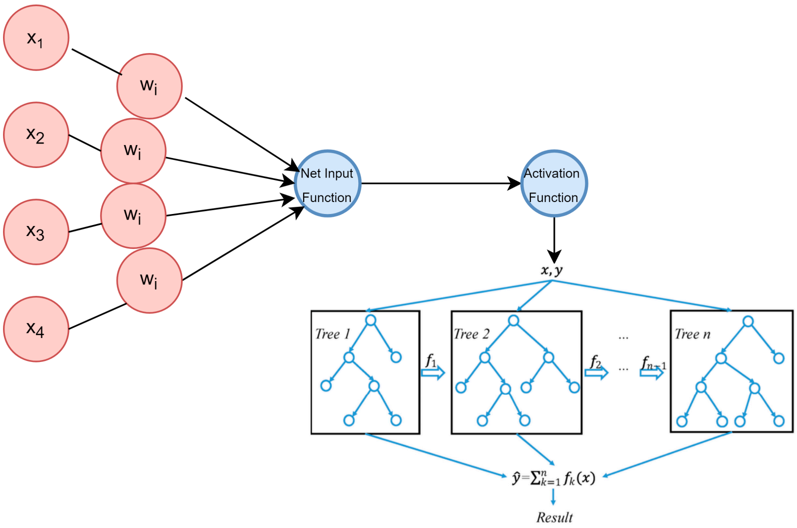


Figure 6. The structure of RF model for classification of DDoS attacks.

Here, P is the probability function of Logistic Regression and y is the output of RF classification model. $\sum_{k=1}^{n} f(x)$ shows the boosting function of the RF Classifier. When RF takes the output of y it will be sent to probability function of logistic regression to check the changes of a class whether it is from 0 or 1.

3.2.3. K-Nearest Neighbors

It stands for "K-Nearest Neighbor", and the abbreviation is KNN. To put it another way, it is part of the field of supervised machine learning. Use this algorithm to solve both classification and regression problems. "K" stands for the number of nearest neighbors to a newly discovered unknown variable that must be predicted or classified. In this study we are classifying attacks by using novel architecture of KNN Classifier. Its general architecture is shown in Figure 7. The input layer contains the vector input signal (x). It is computed between the input signal vector (k) and the neighbors (n) in the hidden layer (y). The output neuron sums the linear outputs O of the hidden layer neurons. The results mentioned in [8] also shows that KNN provides a high detection rate and low false positive rate. An approach using KNN in [9] shows that it is very useful in detecting DDoS attacks.

$$O = \sum W_i k(K_i n_i) \tag{6}$$

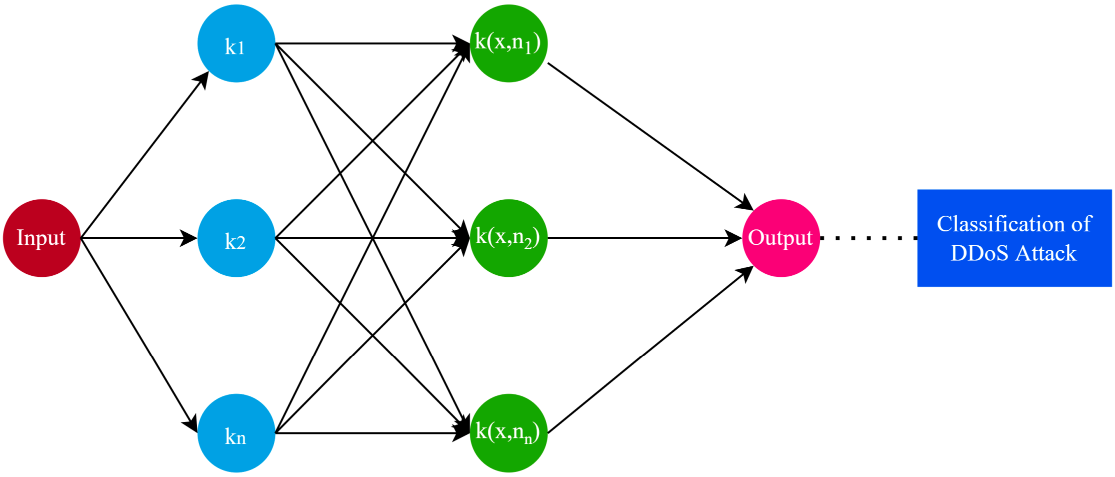


Figure 7. The Block diagram of KNN model for classification of DDoS attacks.

3.3. Performance Metrics

The performance of algorithms has been calculated by showing accuracy precision, recall and F1 score. True positive rate and false positive rates has also been shown by using Confusion Matrix. Accuracy, precision, recall, and F1 Score criteria were utilized to assess the effectiveness of the strategies under consideration. The distinction between classified and misclassified clauses has been demonstrated through the use of a confusion matrix. The computations of the metrics utilized in this study are shown in the Table 2 below.

Table 2. Presents the performance metrics.

Metric	Description
Accuracy	$Accuracy = [\mathrm{TP}/(\mathrm{TP}+\mathrm{TN})] \times 100$
Precision	$Precision = (\mathrm{TP})/(\mathrm{TP}+\mathrm{FP}) \times 100$
Recall	$Recall = (\mathrm{TP})/(\mathrm{TP}+\mathrm{FN}) \times 100$
F1 Score	$F1 = 2\left(\frac{precision \times recall}{precision + recall}\right)$

4. Results

This section demonstrates how each model performs when applied to the chosen dataset. SVM, RF, and KNN models have each been tested on a distinct dataset. There are nine assaults in the Fraud Dataset that we will use to test our model's ability to correctly classify them.

4.1. Performance of SVM Model

A supervised machine learning model known as a support vector machine (SVM) employs classification algorithms to solve classification problems involving two groups. It is possible to train an SVM model to categorize new text after providing it with training data for each category. The performance of SVM model has been shown in the Figures 8 and 9 below. Accuracy and other performance metrics of SVM is 99.8%:

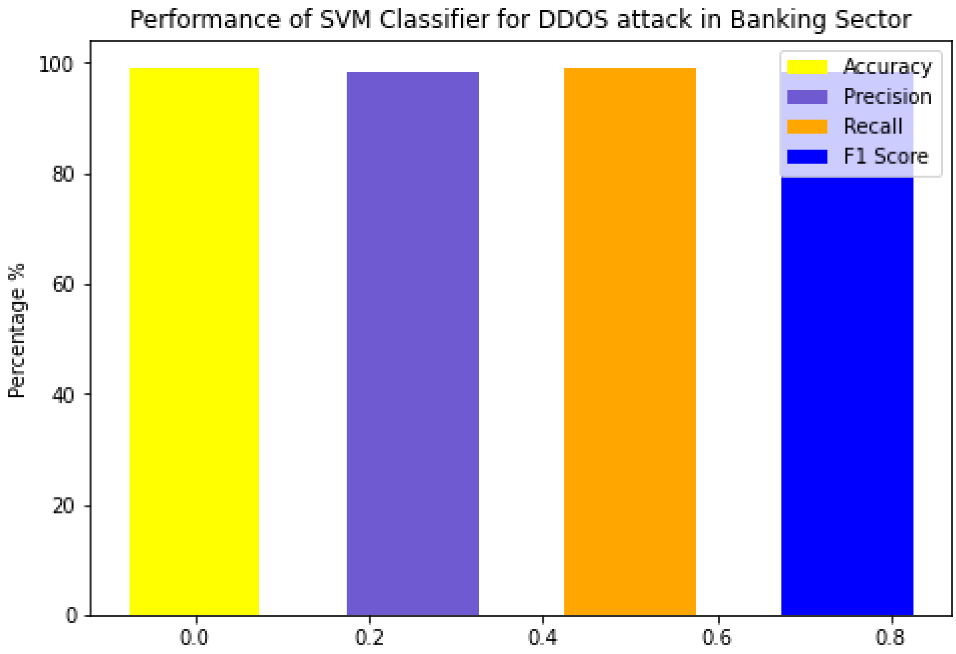
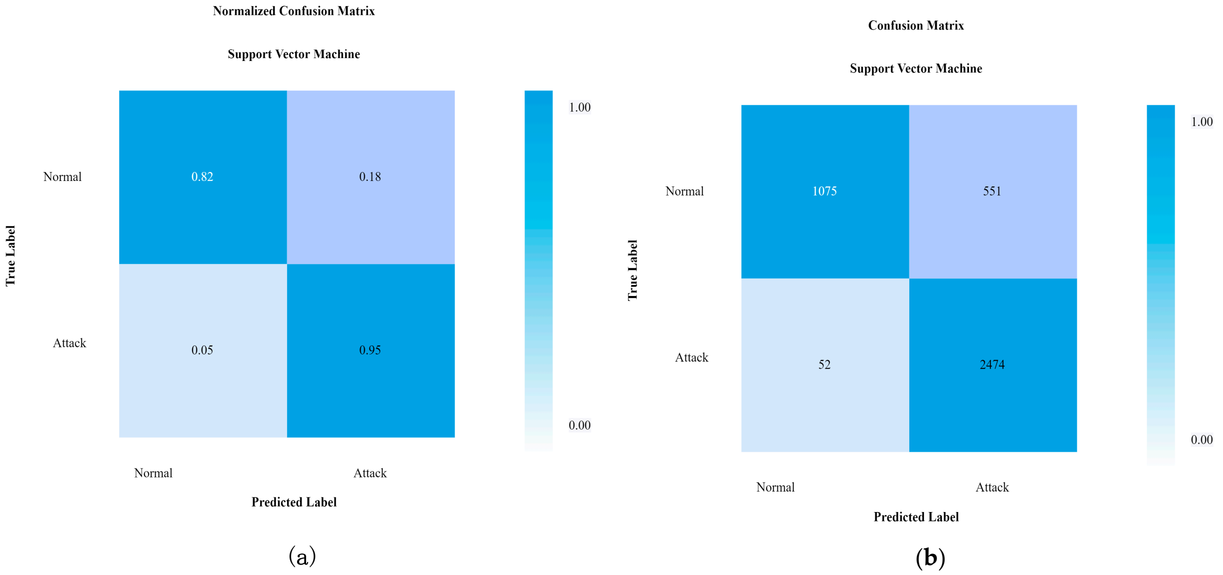


Figure 8. The performance evaluation metrics of SVM model.



(a) (**b**)

Figure 9. Confusion Matrix of SVM model (**a**) Normalized (**b**) Non-Normalized.

4.2. *Performance of Random Forests*

A classification algorithm known as a "random forest" is made up of a large number of decision trees. For each tree, it attempts to create an uncorrelated forest of trees whose prediction by committee is more accurate than any individual tree's using bagging and feature randomness. The performance of RF model has been shown in the Figures 10 and 11 below. Accuracy and other performance metrics of RF is 97.5%

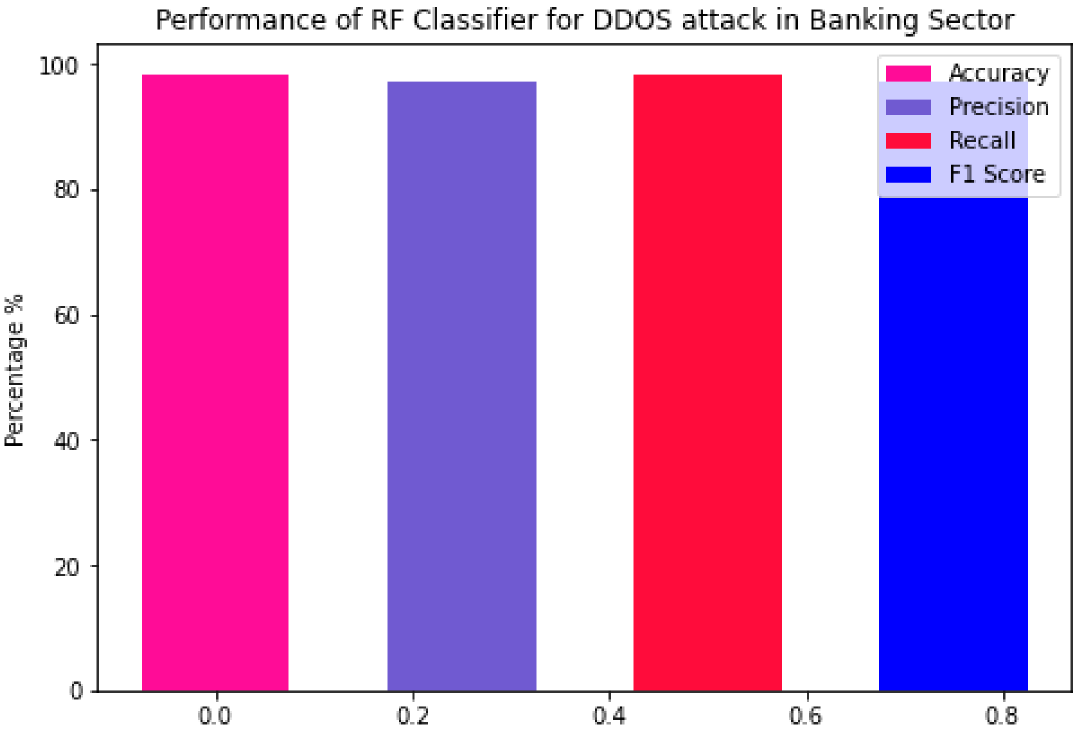
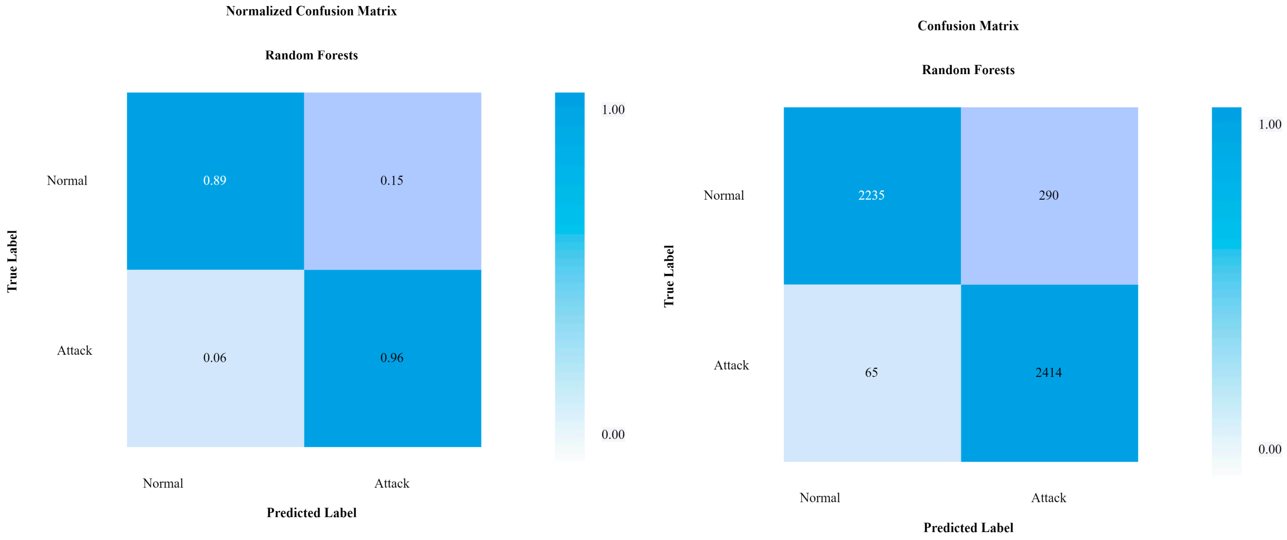


Figure 10. The performance evaluation metrics of RF model.



(a) (**b**)

Figure 11. Confusion matrix of RF(**a**) Normalized (**b**) non-Normalized.

4.3. *Performance of KNN*

For both classification and regression, the KNN algorithm can be used. It's an easy-to-use supervised machine learning algorithm. It slows down as the amount of data used grows, making it easy to implement and understand, but a major drawback. The performance of KNN model has been shown in the Figures 12 and 13 below. Accuracy and other performance metrics of KNN is 98.74%:

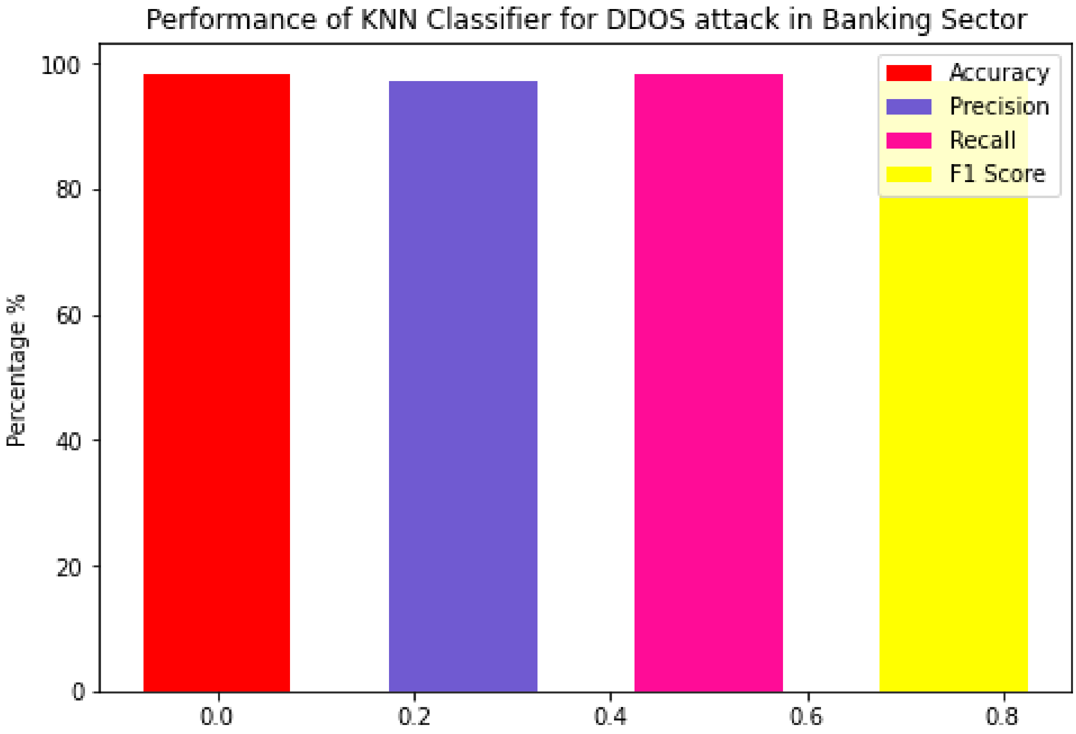


Figure 12. The performance evaluation metrics of KNN model.

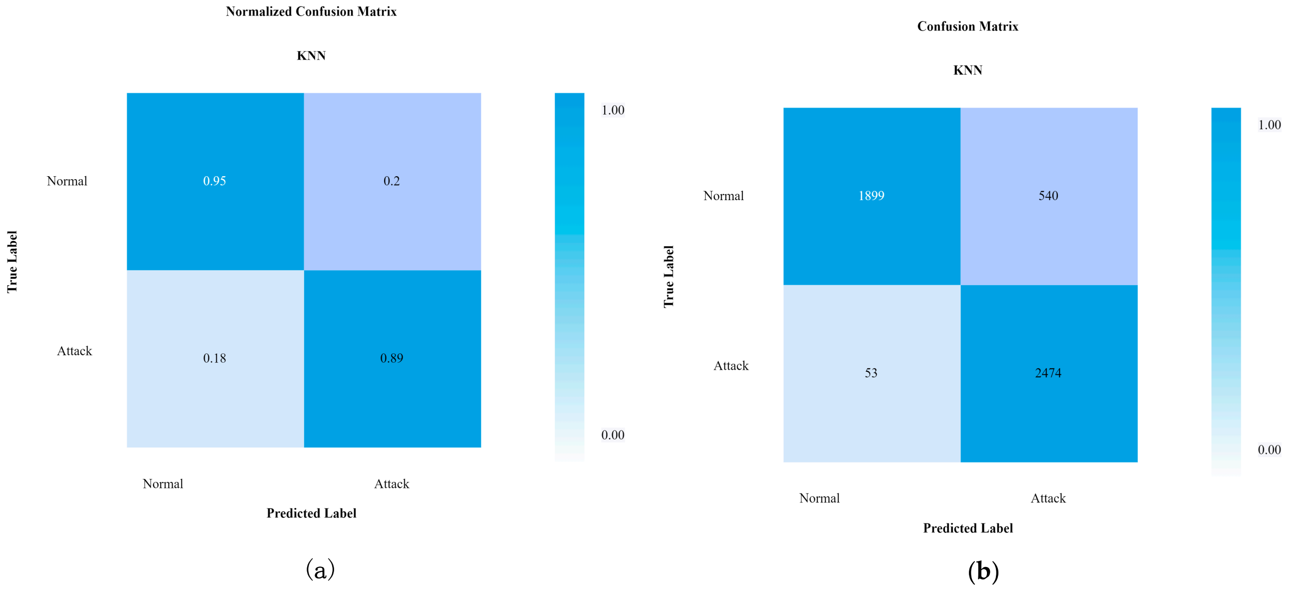


Figure 13. The Confusion matrix of KNN (**a**) Normalized (**b**) Non-normalized.

4.4. Time Complexity (sec)

In order to check the efficiency of the models (SVM, KNN, and RF) time complexity is very important. Figure 14 indicated the time complexity (sec) of the models. SVMs are more efficient than other models (KNN, and RF).

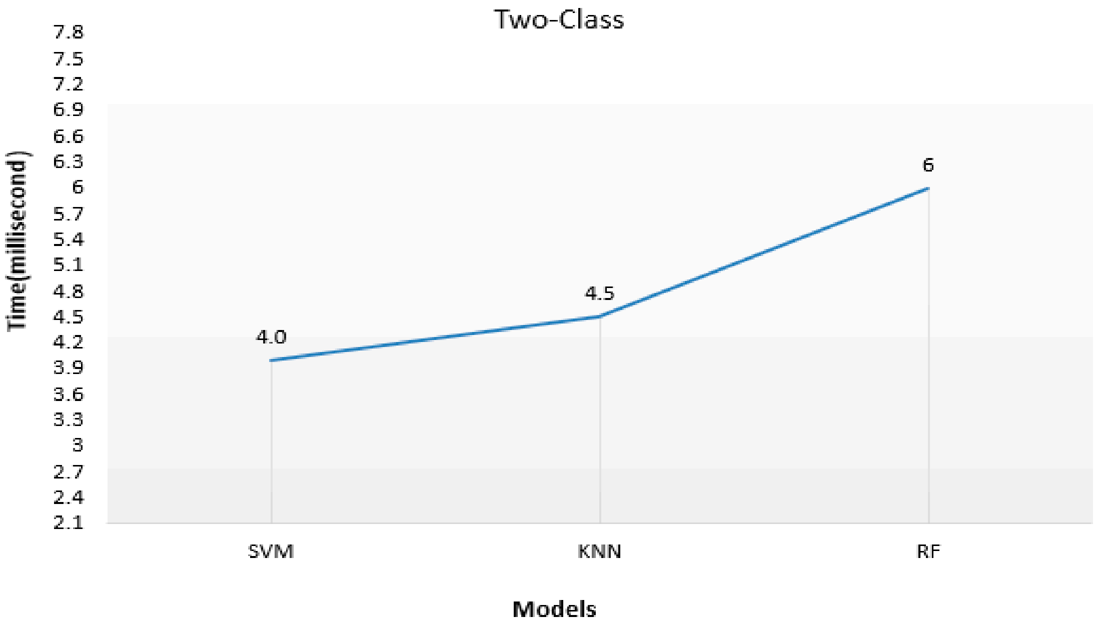


Figure 14. Presents the time complexity (sec) of the SVM, KNN, RF model.

Table 3 shows the comparative analysis of all models in the current study.

Table 3. Presents all the performance metrics of SVM, KNN, and RF model of DDoS detection.

Model	Accuracy%	Precision%	Recall%	F1 Score%
SVM	99.8	99.07	98.32	98.5
RF	97.5	97.23	96.5	97.0
KNN	98.74	98.53	97.33	98.53

It can be seen from our proposed work that the proposed framework is more accurate and efficient as compared to previous studies. Table 4 below shows the comparative analysis of the current study with previous state of art methods (ML/DL) of studies.

Table 4. The comparison study of SVM, KNN, RF with existing ML/DL.

Reference	Model	Accuracy %	Dataset
Current Study	SVM, RF, KNN	99.8, 97.5, 98.74	Banking Fraud Detection (Kaggle)
Dawod et al. [39]	ANN Model	83.5%	IoT Banking Devices Datasets
Hanafizadeh et al. [40]	CNN-LSTM	78%, 79%	Banking Fraud Time Series Data
Yan et al. [41]	SVM	86.7%	DDoS Datasets
Mishra et al. [42]	Trees	85.55%	DDoS Datasets
Gao, Aljuhani, et al. [43,44]	ML (KNN, SVM, ANN)	83%, 84%, 81%	Banking Datasets
Rehman et al. [45]	GRU	81.7%	DDoS Datasets
Guo et al. [46]	ANN, SVM	88.5%, 91%	Real Time Dataset

Additionally, the proposed models have some disadvantages. The models required high computational power and specialized hardware, e.g., they needed a good GPU to accomplish the training process.

5. Conclusions

Financial institutions are particularly vulnerable because of the great monetary worth of the data they hold. Hackers can make a fortune by selling the financial information and banking passwords they have stolen. Similarly, the attack surface that may be exploited by hackers has risen in tandem with the expansion of banks' digital footprints. Using the Banking Dataset, we intend to uncover distributed denial-of-service (DDOS) attempts against financial institutions and other organizations. For the detection of assaults on the financial industry, machine learning algorithms have been implemented. We have applied SVM, KNN and RF. Each model performed with 99.5%, 97.5%, and 98.74% accuracy, respectively, for the detection of DDOS attacks. Comparative results indicated that SVM is more robust as compared to KNN, RF and existing machine learning (ML/DL) approaches. This model is limited to offline datasets, and if we want to work on real-time datasets, we must move to real-time fraud detection applications which are based on these supervised learning models.

Author Contributions: Conceptualization, U.I., A.M., R.M., M.S.H. and I.A.; methodology, U.I., A.M., R.M., M.S.H., I.A., E.T.E., J.A.K., A.U.R. and M.S.; software, U.I., A.M. and R.M.; validation, U.I., A.M., I.A. and M.S.; formal analysis, U.I., R.M., I.A., J.A.K. and E.T.E.; investigation, I.A., A.U.R. and M.S.; resources, R.M., J.A.K., M.S.H., E.T.E., M.S. and I.A.; data curation, E.T.E., I.A. and M.S.; writing—original draft preparation, U.I., A.M., R.M., M.S.H., I.A., E.T.E., J.A.K., A.U.R. and M.S.; writing—review and editing, U.I., A.M., R.M., M.S.H., I.A., E.T.E., J.A.K., A.U.R. and M.S.; visualization, J.A.K. and E.T.E.; supervision, I.A., A.U.R. and M.S.; project administration, E.T.E. and I.A.; funding acquisition, I.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Future University Researchers Supporting Project Number FUESP-2020/48 at Future University in Egypt, New Cairo 11845, Egypt. However, this research also supported by Taif University Researchers supporting Project number (TURSP-2020/159), Taif University-Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this investigation are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this paper:

AI	Artificial intelligence
ML	Machine learning
DL	Deep learning
DDoS	Distributed Denial-of-Service
LDA	Latent Dirichlet Allocation
DNN	Deep Neural Network
HTTP	Hypertext Transfer Protocol
DNN	Deep Neural Network

References

1. Sahingoz, O.K.; Buber, E.; Demir, O.; Diri, B. Machine learning based phishing detection from URLs. *Expert Syst. Appl.* **2019**, *117*, 345–357. [CrossRef]
2. Kambourakis, G.; Moschos, T.; Geneiatakis, D.; Gritzalis, S. Detecting DNS amplification attacks. In *CRITIS 2007: Critical Information Infrastructures Security*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5141 LNCS, pp. 185–196. [CrossRef]
3. Ezekiel, S.; Divakaran, D.M.; Gurusamy, M. Dynamic attack mitigation using SDN. In Proceedings of the 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), Melbourne, VIC, Australia, 22–24 November 2017; pp. 1–6. [CrossRef]
4. Javeed, D.; Gao, T.; Khan, M.T. SDN-enabled hybrid DL-driven framework for the detection of emerging cyber threats in IoT. *Electronics* **2021**, *10*, 918. [CrossRef]
5. Kushwah, G.S.; Ranga, V. Voting extreme learning machine based distributed denial of service attack detection in cloud computing. *J. Inf. Secur. Appl.* **2020**, *53*, 102532. [CrossRef]

6. Osanaiye, O.; Choo, K.-K.R.; Dlodlo, M. Analysing Feature Selection and Classification Techniques for DDoS Detection in Cloud. In Proceedings of the Southern Africa Telecommunication Networks and Applications Conference (SATNAC) 2016, George, South Africa, 7 September 2016; pp. 198–203.
7. Ahmad, I.; Wang, X.; Zhu, M.; Wang, C.; Pi, Y.; Khan, J.A.; Li, G. EEG-Based Epileptic Seizure Detection via Machine/Deep Learning Approaches: A Systematic Review. *Comput. Intell. Neurosci.* **2022**, *2022*, 6486570. [CrossRef] [PubMed]
8. Ahmad, S.; Ullah, T.; Ahmad, I.; AL-Sharabi, A.; Ullah, K.; Khan, R.A.; Ali, M. A Novel Hybrid Deep Learning Model for Metastatic Cancer Detection. *Comput. Intell. Neurosci.* **2022**, *2022*, 8141530. [CrossRef]
9. Ahmad, I.; Ullah, I.; Khan, W.U.; Ur Rehman, A.; Adrees, M.S.; Saleem, M.Q.; Shafiq, M. Efficient algorithms for E-healthcare to solve multiobject fuse detection problem. *J. Healthc. Eng.* **2021**, *2021*, 9500304. [CrossRef]
10. Ahmad, I.; Liu, Y.; Javeed, D.; Ahmad, S. A decision-making technique for solving order allocation problem using a genetic algorithm. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *853*, 012054. [CrossRef]
11. Wang, Y.; Wang, W.; Ahmad, I.; Tag-Eldin, E. Multi-Objective Quantum-Inspired Seagull Optimization Algorithm. *Electronics* **2022**, *11*, 1834. [CrossRef]
12. Ahmad, I.; Liu, Y.; Javeed, D.; Shamshad, N.; Sarwr, D.; Ahmad, S. A review of artificial intelligence techniques for selection & evaluation. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *853*, 012055.
13. Ali, S.; Javaid, N.; Javeed, D.; Ahmad, I.; Ali, A.; Badamasi, U.M. A blockchain-based secure data storage and trading model for wireless sensor networks. In Proceedings of the International Conference on Advanced Information Networking and Applications, Caserta, Italy, 15–17 April 2020.
14. Chayomchai, A.; Phonsiri, W.; Junjit, A.; Boongapim, R.; Suwannapusit, U. Factors affecting acceptance and use of online technology in Thai people during COVID-19 quarantine time. *Manag. Sci. Lett.* **2020**, *10*, 3009–3016. [CrossRef]
15. Mhamane, S.S.; Lobo, L.M.R.J. Internet banking fraud detection using HMM. In Proceedings of the 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Coimbatore, India, 26–28 July 2012. [CrossRef]
16. Alimolaei, S. An intelligent system for user behavior detection in Internet Banking. In Proceedings of the 2015 4th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Zahedan, Iran, 9–11 September 2015. [CrossRef]
17. Fang, L.; Li, Y.; Liu, Z.; Yin, C.; Li, M.; Cao, Z.J. A Practical Model Based on Anomaly Detection for Protecting Medical IoT Control Services against External Attacks. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4260–4269. [CrossRef]
18. Using, N.; Learning, M. A Comprehensive Study of Anomaly Detection Schemes in IoT Networks Using Machine Learning Algorithms. *Sensors* **2021**, *21*, 8320. [CrossRef]
19. Salem, O.; Alsubhi, K.; Shaafi, A.; Gheryani, M.; Mehaoua, A.; Boutaba, R. Man-in-the-Middle Attack Mitigation in Internet of Medical Things. *IEEE Trans. Ind. Inform.* **2022**, *18*, 2053–2062. [CrossRef]
20. Gupta, D.; Gupta, M.; Bhatt, S.; Tosun, A.S. Detecting Anomalous User Behavior in Remote Patient Monitoring. In Proceedings of the 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), Las Vegas, NV, USA, 10–12 August 2021; pp. 33–40. [CrossRef]
21. Saeedi, K. Machine Learning for Ddos Detection in Packet Core Network for IoT. Master's Thesis, Luleå University of Technology, Luleå, Sweden, 2019.
22. Tahir Ullah, K. Internet of Things (IOT) systems and its security challenges. *Int. J. Adv. Res. Comput. Eng. Technol.* **2019**, *8*, 12.
23. Kamruzzaman, M.M. New Opportunities, Challenges, and Applications of Edge-AI for Connected Healthcare in Smart Cities. In Proceedings of the 2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain, 7–11 December 2021.
24. Jegadeesan, S.; Azees, M.; Ramesh Babu, N.; Subramaniam, U.; Almakhles, J.D. EPAW: Efficient Privacy Preserving Anonymous Mutual Authentication Scheme for Wireless Body Area Networks (WBANs). *IEEE Access* **2020**, *8*, 48576–48586. [CrossRef]
25. Oppliger, R.; Rytz, R.; Holderegger, T. Internet banking: Client-side attacks and protection mechanisms. *Computer* **2009**, *42*, 27–33. [CrossRef]
26. Zachos, G.; Essop, I.; Mantas, G.; Porfyrakis, K.; Ribeiro, J.C. An Anomaly-Based Intrusion Detection System Internet of Medical Things Networks. *Electronics* **2021**, *10*, 2562. [CrossRef]
27. Lange, T.; Kettani, H. On Security Threats of Botnets to Cyber Systems. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019; pp. 176–183.
28. Aski, V.; Dhaka, V.S.; Kumar, S.; Parashar, A.; Ladagi, A. A multi-factor access control and ownership transfer framework for future generation healthcare systems. In Proceedings of the 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 6–8 November 2020; pp. 93–98. [CrossRef]
29. Mehmood, M.; Javed, T.; Nebhen, J.; Abbas, S.; Abid, R.; Bojja, G.R.; Rizwan, M. A hybrid approach for network intrusion detection. *Comput. Mater. Contin.* **2021**, *70*, 91–107. [CrossRef]
30. Ramapatruni, S.; Narayanan, S.N.; Mittal, S.; Joshi, A.; Joshi, K. Anomaly Detection Models for Smart Home Security. In Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Washington, DC, USA, 27–29 May 2019; pp. 19–24. [CrossRef]
31. Hameed, M.; Yang, F.; Ghafoor, M.I.; Jaskani, F.H.; Islam, U.; Fayaz, M.; Mehmood, G. IOTA-Based Mobile Crowd Sensing: Detection of Fake Sensing Using Logit-Boosted Machine Learning Algorithms. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 6274114. [CrossRef]

32. Kaushik, I.; Sharma, N. Black hole attack and its security measure in wireless sensors networks. In *Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's*; Springer: Cham, Switzerland, 2020; Volume 1132.
33. Dilraj, M.; Nimmy, K.; Sankaran, S. Towards Behavioral Profiling Based Anomaly Detection for Smart Homes. In Proceedings of the TENCON 2019–2019 IEEE Region 10 Conference (TENCON), Kochi, India, 17–20 October 2019; pp. 1258–1263.
34. Javeed, D.; Khan, M.T.; Ahmad, I.; Iqbal, T.; Badamasi, U.M.; Ndubuisi, C.O.; Umar, A. An efficient approach of threat hunting using memory forensics. *Int. J. Comput. Netw. Commun. Secur.* **2020**, *8*, 37–45. [CrossRef]
35. Javeed, D.; Gao, T.; Khan, M.T.; Shoukat, D. A hybrid intelligent framework to combat sophisticated threats in secure industries. *Sensors* **2022**, *22*, 1582. [CrossRef] [PubMed]
36. Shaikh, H.; Khan, M.S.; Mahar, Z.A.; Anwar, M.; Raza, A.; Shah, A. A conceptual framework for determining acceptance of internet of things (IoT) in higher education institutions of Pakistan. In Proceedings of the 2019 International Conference on Information Science and Communication Technology (ICISCT), Karachi, Pakistan, 9–10 March 2019; pp. 1–5. [CrossRef]
37. Huang, K.; Yang, L.X.; Yang, X.; Xiang, Y.; Tang, Y.Y. A Low-Cost Distributed Denial-of-Service Attack Architecture. *IEEE Access* **2020**, *8*, 42111–42119. [CrossRef]
38. Razib, A.M.; Javeed, D.; Khan, M.T.; Alkanhel, R.; Muthanna, M.S.A. Cyber Threats Detection in Smart Environments Using SDN-Enabled DNN-LSTM Hybrid Framework. *IEEE Access* **2022**, *10*, 53015–53026. [CrossRef]
39. Dawod, A.; Georgakopoulos, D.; Jayaraman, P.P.; Nirmalathas, A.; Parampalli, U. IoT Device Integration and Payment via an Autonomic Blockchain-Based Service for IoT Device Sharing. *Sensors* **2022**, *22*, 1344. [CrossRef]
40. Hanafizadeh, P.; Amin, M.G. *The Transformative Potential of Banking Service Domains with the Emergence of FinTechs*; Palgrave Macmillan: London, UK, 2022; No. 0123456789.
41. Yan, W. Security Optimization Management for loT-Assisted Bank Liquidity Risk Emergency Using Big Data Analytic-Based Case Reasoning. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 8396931. [CrossRef]
42. Mishra, P.; Guru Sant, T. Role of Artificial Intelligence and Internet of Things in Promoting Banking and Financial Services during COVID-19: Pre and Post Effect. In Proceedings of the 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 22–23 October 2021.
43. Javeed, D.; Gao, T.; Khan, M.T.; Ahmad, I. A Hybrid Deep Learning-Driven SDN Enabled Mechanism for Secure Communication in Internet of Things (IoT). *Sensors* **2021**, *21*, 4884. [CrossRef]
44. Aljuhani, A. Machine learning approaches for combating distributed denial of service attacks in modern networking environments. *IEEE Access* **2021**, *9*, 42236–42264. [CrossRef]
45. ur Rehman, S.; Khaliq, M.; Imtiaz, S.I.; Rasool, A.; Shafiq, M.; Javed, A.R.; Jalil, Z.; Bashir, A.K. Diddos: An approach for detection and identification of distributed denial of service (ddos) cyberattacks using gated recurrent units (gru). *Future Gener. Comput. Syst.* **2021**, *118*, 453–466. [CrossRef]
46. Guo, C.; Wang, H.; Dai, H.N.; Cheng, S.; Wang, T. Fraud risk monitoring system for e-banking transactions. In Proceedings of the 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, 12–15 August 2018; pp. 106–113.