SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 5781-5794
Open Access. Online & Print

DOI: 10.5281/zenodo.19642646

ENHANCING CLOUD DATA SECURITY WITH A ROBUST INTRUSION DETECTION SYSTEM BASED ON MACHINE LEARNING

Saman Khan[1]*, Dr. Syed Hauider Abbas[2]

[1]PhD Scholar Integral University: Department of Computer Science and Engineering Lucknow, India.
Email: mail2samankhan@gmail.com
[2] Faculty, Integral University: Department of Computer Science and Engineering Lucknow, India.
Email: shabbas@iul.ac.in

Received: 14/12/2025
Accepted: 10/03/2026

Corresponding Author: Saman Khan (mail2samankhan@gmail.com)

ABSTRACT

To improve a business enterprise's safety against cyberattacks, cyber safety must be implemented while imposing cloud computing. It's tough to have network intrusion detection without a false alarm. This comes from the dataset's disparity between redundant and beside the point functions and informative capabilities. This article suggests a novel hybrid intrusion detection machine based totally on gadgets getting to know. This has a look at makes use of the CICIDS 2017 dataset, which includes both benign and malicious network traffic, to simulate different real-global attack types, such as Brute pressure, DoS, net assaults, Infiltration, Botnet, Port scan, and DDoS. First, characteristic choice and normalization proficiency are used to be easy and prepare the data. We subsequent enforce and evaluate seven system learning techniques such as Navie Bayes, choice Tree, Logistic Regression, KNN, Random woodland, SVM and Multilayer Perceptron. Combining several machine gaining knowledge of techniques enables us to create a more correct and stronger IDS, in line with this proposed hybrid model is a good manner to beautify community safety because the proposed hybrid LR-MLP set of rules plays higher than present classifiers, according to the experimental results, with an accuracy of 94.4%.

KEYWORDS: Intrusion Detection System, Machine Learning, Hybrid Algorithm, Network Security, Cybersecurity, Anomaly Detection.

1. INTRODUCTION

One of the latest trends in the IT enterprise is cloud computing. The main benefit of cloud computing is that it offers get admission to without constraints of time and region. Cloud computing offers reduced costs, flexibility in coping with garage capacities, and resources for cellular and collaborative apps and offerings. Additionally, because of the fact cloud services are multi-sourced, clients can pick from a variety of company companies their wishes. utilizing cloud computing moreover reduces the need for maintenance and physical place for on-net website garage, similarly to capital prices and energy consumption.

A growing type of companies, banks, and governments are the usage of cloud computing services. sturdy safety features are vital because this alteration additionally made those structures prone to various varieties of intrusions through hackers and intruders. Several cloud provider vendors provide a spread of safety services as apps. As an example, bear in mind the Amazon internet services (AWS) internet site online, which offers offerings with various cut-off dates and validity based totally on period of carrier license.

continuous backup and updates are required because of the growing quantity of records, specifically medical records. Healthcare records are liable to hackers who can also additionally use it for illegal reasons, which includes political or financial gain. Prescription medication information, health records, and special affected individual records are all covered. To prevent undesirable, get proper access to banking, social media, and monetary statistics, net threats like fraud must be averted.[28]

Due to the fact affected individual and operational facts are personal, cybersecurity needs to be the primary difficulty for healthcare services while enforcing cloud computing technology. Current intrusion detection structures (IDS) are used as its method anomaly or signature detection. Operational and patient statistics may be taken if the detection method isn't always appropriate. Cybersecurity strategies beautify an employer's safety towards cyberattacks while enhancing inside the detection and defense closer to malicious attacks.

preceding studies finished in international hospitals have handiest targeted individual awareness of the importance of cybersecurity, collectively with the usage of strong passwords, putting off or detection of undesirable emails, statistics encryption, the non-public handling of credentials, restrained information access, and properly timed notification of safety breaches [2]. Even though the fitness area has intrusion detection systems in location, there can also nonetheless be weaknesses in its nearby infrastructure. The cloud but offers an extremely green and cozy approach to handling hospitals or every specific kind of enterprise company.[29].

Ultimately, all government IT sports could be achieved on cloud [30]. relying on the offerings provided through manner of the cloud service business enterprise, the cloud offers a spread of safety techniques. In some times, very non-nonprivate information cannot be saved at the cloud, however some operational records may be stored . facts hosted on inner servers must have the identical diploma of security as data saved within the cloud. Although it takes several try and optimize cloud aid protection, storing and sensitive clinical facts on the cloud develop recollect.[33]

To defend networks against cyberattacks, intrusion detection systems are essential. The reasons for those systems are designed to reveal community site visitors and perceive any uncommon and dangerous hobby. One main undertaking in community safety these days is identifying hidden attacks -people who aren't easily significant or are new. Detecting such assaults early is the first and most crucial step in securing any device.[34]

In this research we identifying seven common and dangerous styles of community attacks which are:

- **Brute Force Attack:** Trying many passwords and login attempts to gain access.
- **Heartbleed or Denial of Service (DoS):** Exploiting a system to crash and slow it down.
- **Web Attacks:** Targeting websites through code injections or unauthorized access.
- **Infiltration:** Gaining unauthorized access to internal systems.
- **Botnet:** Using a network or infected machines to perform attacks.
- **Port Scan:** Scanning ports to find system weakness.
- **Distributed Denial of Service (DDoS):** Flooding a server with traffic to make it unavailable.

To study and detect these attacks, we used the CICIDS -2017 dataset, which contains real -world examples of some average and malicious network traffic. From this dataset, we selected only the most important features using feature selection techniques. This helps improve the model's accuracy and reduces the amount of unnecessary data. The selected features were matched with specific machine learning algorithms to find the best results for each attack type.

We evaluated and compared various systems to get to know algorithms the usage of a tool named WEKA. After reading character algorithms, we went one step further by combining two or more supervised machines studying algorithms to create a hybrid model. This novel hybrid set of rules demonstrates how combining techniques could make IDS more clever and powerful while enhancing detection performance. The software of hybrid techniques additionally demonstrates how gadget mastering can be prolonged and changed to deal with extra complicated issues in network and cloud protection.

The relaxation of the paper is prepared as follows: In phase II, we talk about motivation and intention. Segment III elaborates on the associated work regarding IDS and associated research. The methodology and counseled hybrid set of rules are provided in segment IV and the materials and strategies hired on this observe are defined in segment V. In phase VI, the experimental outcomes are supplied. Lastly, phase VII includes the conclusion and destiny paintings.

2. MOTIVATION AND OBJECTIVE

Network protection is becoming a main problem and challenge in each day lives as the internet continues expanding. For network protection, intrusion detection techniques play a critical position in identifying any intrusion. techniques for detecting intrusions are hired now not only to pick out assaults however additionally to raise an alarm for approximately any abnormal interest occurring within a community. The conduct of the compromised community is uncommon in preferred and will suggest an attack or anomaly [1].

three strategies exist for identifying some invasion: (i) The signature-based technique, which detects recognized threats by using the use of their signatures without placing off a variety of false alarms. (ii) traditional structures and community performance are tested the usage of anomaly-based totally techniques, which recognize anomalies as departures from normal network hobby. This approach enables detecting novel or current assaults. (iii) Hybrid tactics integrate anomaly-primarily based and signature strategies. This method improves the charge of recognized intrusion detection whilst reducing the fake superb fee for hidden attacks [2].

With the aid of focusing on and keeping the important thing functions, characteristic selection techniques decorate IDS's universal performance. A powerful hybrid exemplary has been advanced the usage of the WEKA tool [3] [4]. Using a hybrid set of rules, special decided on attributes, and category algorithms the use of the WEKA device, a green and robust intrusion detection framework is advanced [5].

3. RELATED WORK

One important area that has been thoroughly researched by researchers is intrusion detection systems (IDS). IDS approaches can be broadly divided into three categories: hybrid, signature-based, and anomaly-based. The first step in handling massive volumes of IDS data is dataset preprocessing. Features selection is the primary preprocessing step. Following feature selection, machine learning techniques are used to categorize intruders' typical and unusual behavior. The IDS categories are highlighted in the text that follows, finally, the hybrid ML approaches that were applied in the IDS for this study.[35]

IDS Concept's and Related work

An intrusion detection system (IDS) is used to identify anomalies and malicious intrusions in cloud computing by monitoring every incoming and outgoing request to identify any unusual or typical activities. Since all services are available online and are susceptible to cyberattacks, infiltration and intrusions are the primary issues with networks and cloud computing. IDS software needs to be highly effective to use, scalable, dynamic, and self-adaptive. There are three steps in the intrusion detection procedure [31].

- Gathering data from the systems under observation.
- To determine whether it is intrusion or normal.
- Notifying the administrator of the opportunity of intrusion into the gadget in subject to be able to take the essential precautions to preserve the device relaxed.

Table 1: Overview of the Intrusion Detection Techniques.

Anomaly based IDS	
Characteristics	**Restrictions**
This method can identify unknown attacks. low false alarms rate for unidentified attackers. DS gathers information about network behavior for performance metrics and testing.	This method takes a lot of time. To accurately detect threats, it needs a lot of features.
Signature based IDS	
Characteristics	**Restrictions**

high detection rate	A very high false alarm rate for unidentified attackers. Unable to identify new attacks or variations of known attacks
Hybrid IDS Techniques	
Characteristics	**Restrictions**
it's far a mixture of or extra strategies. Detection accuracy is very high.	Computational cost is very high.

Table 2 summarizes preceding research that aimed to provide an in-depth study of the supervised device gaining knowledge of techniques implemented to intrusion detection. This research used trendy intrusion detection to datasets to teach and look at diverse devices, mastering tactics. Principally based on the network strains of popular datasets or facts furnished by using the controller, we can both extract new functions and use a subset of these fashionable datasets. Using recursive characteristic elimination techniques, we selected a set of features from the CICIDS-2017 dataset. We additionally took into attention supervised device learning strategies consisting of Naive Bayes (NB), decision Tree (DT), Logistic Regression (LR), ok-Nearest Neighbor (KNN), Random wooded area (RF), aid Vector device (SVM), and Multilayer Perceptron (MLP). For performance evaluation, we compute accuracy, precision, keep in mind, and F-degree using 80% schooling records and 20% checking out facts.

Table 2: Previous Machine Learning-Based Intrusion Detection Systems (IDS).

Author	Algorithms	Dataset	Work Done
Belagavi et al. [6]	LR, SVM, RF, and Gaussian NB	NSL-KDD	Develop a machine learning method for IDS.
Yen et al [25]	ANN, RF, SVM, RNN-IDS	NSL-KDD	provide a DL approach for intrusion detection based on recurrent neural networks
Khaled et al [7]	RBM along with the DBNs	KDDCUP99	Implemented a deep belief network along with a restricted Boltzmann machine (RBM).
Zho et al [8]	PNN-equipped DBNs	KDDCUP99	Developed a DBN and PNN-based intrusion detection technique.
Roy et al [9]	SVM and DNN	KDDCUP99	An analysis comparing DNN and SVM
XU et al [10]	LSTM and GRU	NSL-KDD	An RNN with GRU and an MLP soft max module are included in the developed IDS.
WU et al [11]	CNN, RF, RFT, MLP, SVM, RNN, J48, NB, and NBT	NSL-KDD	Proposed an IDS model that utilizes CNN.
Wang et al [12]	JSMA, FGSM	NSL-KDD	Analyzed the attack algorithms' effectiveness against DNN IDS using the NSL-KDD dataset.
Shone et al [13]	S-NDAE and DBN	KDD99	Created NDAE to learn features without supervision.
Ly et al [14]	RF, k-NN, BPNN, NB, C4.5	KDD99	Mature KNN algorithm
Lata et al [27]	AdaBoost, RUS Boost, Logit Boost, BT	NSL-KDD	Evaluated the effectiveness of the IDS using several classification methods.
Kamrudin et al [15]	DT, MLP, SVM, and NB	KDD99	IDS was created by employing ensemble classification methods.
Jia et al [16]	NDNN	KDD99	Suggested a novel deep neural network model as the basis for an IDS.
Gogoi et al [17]	Hybrid multi-level intrusion detection technique	KDD99 and NSL-KDD	Introduced a MLHI detection-method.
Geo et al [18]	J48, MLP, SVM, RF, RT, NB, NBT	Kyoto 2006, KDD Cup 99, and NSLKDD	Developed a fuzziness-based semi-supervised learning system.
Ambu Saidi et al. [19]	MIFS + LSSVM, LSSVM + FMIFS	KDD99	An algorithm based on mutual information was presented.

4. METHODOLOGY

A. Proposed Framework

The goal of this work is to offer a hybrid intrusion detection model that analyzes the accrued dataset the use of gadget gaining knowledge of algorithms. the important emphasis can be given the CICIDS-2017 Dataset, function selection, normalization, gadget learning techniques, and facts pre-processing. The use of predefined metrics, the performance of seven special gadgets mastering algorithms can be evaluated. After then, or more algorithms are blended to create a brand-new hybrid set of rules, and its performance is evaluated and compared with that of present devices gaining knowledge of

algorithms.

The best four individual learning algorithms Random Forest (RF), SVM, Logistic Regression (LR), and MLP were combined to construct the six hybrid classifiers shown in Figure 1. Classifiers from Proposed Module that produce satisfactory outcomes are considered as potential algorithms for the development of hybrids. SVM-LR, SVM-MLP, LR-MLP, RF-SVM, RF-LR, and RF-MLP are the six hybrid classifiers that were developed. It uses the same CICIDS-2017 dataset that was utilized in Proposed Module. This module's tests were all carried out in WEKA.

Proposed Hybrid Module

Input: CICIDS-2017 dataset used

Output- Accuracy, precision, recall, F-measure

Step1: Define the problem in detail.
Step 2: Select the CICIDS-2017 data set and import it into the software.
Step 3: Apply data preprocessing on the selected dataset.
Step 4: Apply recursive feature elimination techniques.
Step 5: Standardize the data using normalization techniques.
Step 6: Select from one :
Random Forest (RF), Support Vector Machine (SVM), K-Nearest
Neighbor (KNN), Decision Tree (DT), Logistic Regression (LR),
Navie Bayes (NB), or Multilayer Perceptron (MLP).
Step 7: Apply machine learning techniques on a few selected features.
Step 8: Evaluate the defined metrics.
Step 9: If the outcomes obtained are satisfactory, consider the algorithms for more processing.
Step 10: Repeat steps through 6 to 9. For further machine learning algorithms.
Step 11: Analyzing and evaluating machine learning algorithm results
according to performance metrics.
Step 12: Go to step 5
Step 13: Create new hybrid algorithms by combining two or more machine learning
Algorithms from Step 6 and evaluate the outcome.
Step 14: Equivalence the outcomes.
Step 15: Compare and evaluate the best outcome.
Step 16: Algorithms End.

B. Pre-Processing

The process of preparing data in a specific format or manner that is suitable for implementing the intended machine learning algorithms into effect is known as data pre-processing. Preprocessing data is an important step in machine learning. Pre-processing means collecting data from various sources, converting it into a single, appropriate format, cleaning the data, normalizing it, and dividing it before applying various machine learning algorithms [26]. Sometimes, data is not present in one place.

C. Feature Selection

In machine learning, feature selection includes pre-processing a clean dataset, eliminating irrelevant characteristics, and selecting a unique subset for the construction of models. Just 22 out of the 79 features in the CICIDS-2017 dataset provide significant help to attack classification. A more effective machine learning model is produced by reducing the size of the dataset with highly related features using the Recursive Feature Elimination approach. This method aids in reducing the dataset's size.[20]

D. Normalization

When working with attributes on different scales, normalization is necessary to make sure that the efficiency of a significant or equally essential attribute is not impacted by the values of other attributes on a larger scale. This is because machine learning processes may result in inadequate data models when there are numerous features, yet their values fall on different scales. Consequently, they are normalized to put all the attributes on an identical scale.

Fig 1. Proposed Hybrid Module.

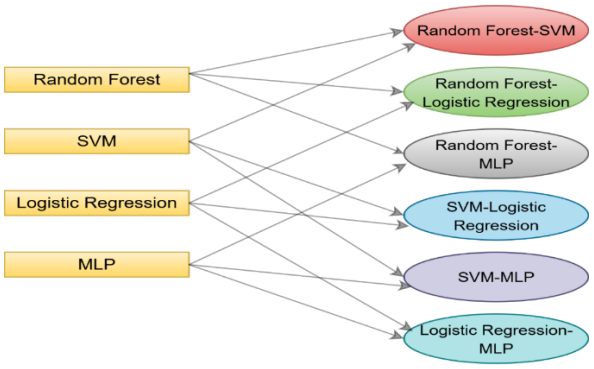

Fig 2. Hybrid Classifier.

E. Classification Algorithms

Seven machine learning methods were used in this work, each with a detailed description, and several analyses were carried out using the machine learning tool WEKA.

- **Navie Bayes (NB)**

A probabilistic classifier, the Navie Bayes algorithm is based on the Bayes theorem and assumes that functions and predictors are independent, meaning that one feature has no effect on the others. It is quick and easy to use because it classifies data using density-based assumptions. However, it has several disadvantages because each characteristic is independent.[20]

$$Posterior = \frac{Likelihood * Prior}{Marginal} \quad (1)$$

(Where A stands for a hypothesis and B for evidence

Three categories of Naive Bayes algorithms exist:

i. **Bernoulli Naïve Bayes-** Bernoulli Naïve Bayes uses Boolean or binary output features.

ii. **Gaussian Naïve Bayes-** Gaussian Naïve Bayes follows a continuous Gaussian distribution

iii. **Multi-nominal Naïve Bayes-** It uses numeric data with discrete features.

- **Random Forest (RF)**

One versatile and simple machine learning technique that is a member of the ensemble family of algorithms is the Random Forest algorithm. Prediction accuracy is increased by generating many weak learners and combining them with decision making. A random variable subset from the candidate's predictor variables is used to create each new instance, which the algorithm then classifies using a "forest" of random decision trees. This method can be applied to regression and classification with hyperparameters such as bagging classifiers or decision trees. On large datasets, Random Forest produces more accurate results, and by establishing a random threshold for overall features, numerous random trees can be produced. The over-fitting problem is likewise addressed by this algorithmic method.

- **KNN, or K-Nearest Neighbors**

This method of categorization is lazy because it doesn't build a model before looking at a test instance. Once a test example is reached, the learner shortens information and saves all training instances to forecast a new class label or numeric value. KNN is predicated on the notion of the nearest neighbor's decision rule, which holds that a collection of historical data determines the class of an item that has never been seen before [21].

The Minkowski distance of order 1 is another name for the Manhattan distance d_M between the tuples x1 and x2. The following illustrates this distance:

$$d_M(x_1, x_2) = \sum_{i=1}^{n} |x_{1i} - x_{2i}| \quad (2)$$

To calculate the Euclidean distance d_E between two tuples, x1 and x2:

$$d_E(x_1, x_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i}^2} \quad (3)$$

- **Support Vector Machine (SVM)**

A training algorithm is referred to as SVM. This approach uses statistical learning to train classification and regression rules from data. In addition to indirectly addressing the primary issues without providing a solution for more complex ones, it somewhat overcomes the bias-variance trade-off dilemma. There are two main approaches to implementing this algorithmic technique. In the first approach, statistical programming is used. However, the second approach uses kernel functions. This method, which makes use of kernel functions, further separates the data into two classes, P and N. Yi = +1 is the positive situation, represented by class P, and Yi = -1 is the negative scenario, represented by class N [22].

For linearly separable training data, there is a pair (w, b) such that

$W^T X_i + b \geq 1 for all X_1 \in P, and$

$W^T X_i + b \leq -1 for all X_1 \in P$ (4)

In the equations, a weight vector W and bias b are used to provide the prediction rule.

$f = sign(W.X) + b)$ (5)

- **Decision Tree (DT)**

This is a decision-making tool that shows potential outcomes, such as utility, costs, and consequences using a tree model. Domains display characteristics, branches indicate yields, and classes display classes in this conditional control statement algorithm that shows rules. To demonstrate unique classification based on most facts, decision-makers can select the best option and proceed from the root to the episode. Many scientists in the field use the decision tree extensively [23].

- **LR or logistic regression**

Weighted features are extracted from input data using the logistic regression model, which then combines the features linearly, multiplies them by a weight, and summarizes the results. It makes predictions about what will happen at an event using forecast variables, which can be either numerical or categorical. To provide predictions, this method incorporates data into a logistic function [24].
Here is a statement of the Logistic Regression theory:

$h_\theta(x)=g(\theta^T x)$ ***(6)***

- **MLP, or multilayer perceptron's**

A subfield of computer science called ANN—also referred to as neural networks or multi-layer perceptions—uses simple biological brain models to solve computational issues like machine learning predictive modeling. The objective is to improve predictivemodeling in machine learning by developing robust methods and data structures that can depict complex scenarios instead of creating realistic brain models.

The strength of neural networks is their capacity to align anticipated output variables with training data representations. They are universal approximation algorithms that acquire mapping functions mathematically. Their hierarchical structure, which can identify and combine traits of various sizes and resolutions to produce bigger features, is what gives them their predictive potential. Neural networks can make precise predictions thanks to their multi-layeredstructure.

- **Performance evaluation**

The following criteria are applied to all classifiers: F Measure, Accuracy, Precision, and Recall. FP and True TP, TN, and FN have all been used to calculate performance. All the above values are computed using confusion matrices

❖ The following formula is used to determine accuracy, which is the precise forecasting of both positive and negative occurrences:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (8)$$

❖ Precision determines the probability of a successful positive prediction.

$$Precision = TP/(TP + FP) \quad (9)$$

❖ The recall calculates how many accurate classifications suffer by the number of entries that are missed.

$$Recall = TP/(TP + FN) \quad (10)$$

❖ This is used to determine effectiveness.

$$F - Measure = 2 * ((Precision * Recall)/Precision + Recall)) \quad (11)$$

5. MATERIALS AND METHODS

This section describes the different hardware and software utilized and provides a concise of the CICIDS-2017 datasets, which are used for intrusion detection.

Hardware Used

- CPU: Intel Core i7 2.80 GHz
- RAM: 16 GB
- HDD: 1 TB

Software Used

Using machine learning algorithms, we have examined and analyzed the CICIDS-2017 dataset using the WEKA Software tool. Waikato Environment for Knowledge Learning, or WEKA, is an open-source graphical user interface tool. The WEKA software program was created by the University of Waikato in New Zealand and can identify data from a large amount of information gathered from many domains. It may be applied to a wide range of data mining and machine learning tasks, including preprocessing, clustering classification, regression, feature selection, and visualization regression [32].

CICIDS-2017 Dataset

The Canadian Institute for Cybersecurity released CICIDS-2017 in 2017 as an intrusion detection assessment dataset, containing both benign and malicious traffic recorded in PCAP files. The dataset is identical to real-world data and includes data flows labelled by timestamp and types of assaults in CSV files.

CICIDS-2017 Attacks

Over 2.8 million flows are included in the CICIDS 2017 dataset. This dataset also tries to cover a varied and up-to-date range of assaults that may be found in today's networks. This dataset includes eight attacks that reflect several attack categories: web, brute force, DoS, DDoS, infiltration, heart bleed, botnet, and port scan.

Details of CICIDS-2017 Dataset

The statistics changed into accumulated for a complete of 5 days. It began on Monday, July 3, 2017, at 9 a.m. and completed on Friday, July 7, 2017, at five p.m. during this time, several assaults were carried out. Monday's site visitors incorporate the standard traffic of the day. The assaults have been finished inside the morning and afternoon on Tuesday, Wednesday, Thursday, and Friday, respectively. Brute force FTP, Brute stress SSH, DoS, Heartbleed, net assault, Infiltration, Botnet, and DDoS are several of the attacks which have been deployed.

Table 4 provides a list of various attacks, the magnitude of each of the 14 attacks, benign data, and the day each attack was carried out.

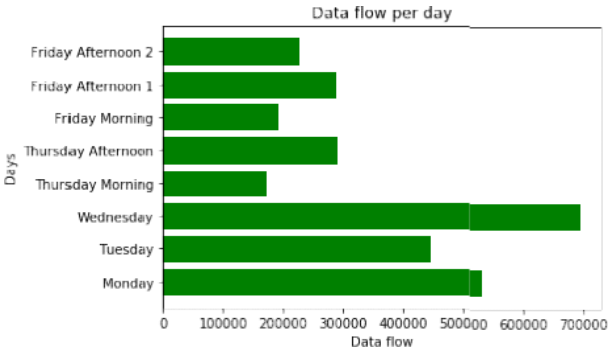

Fig 3. Data Flow per Day.

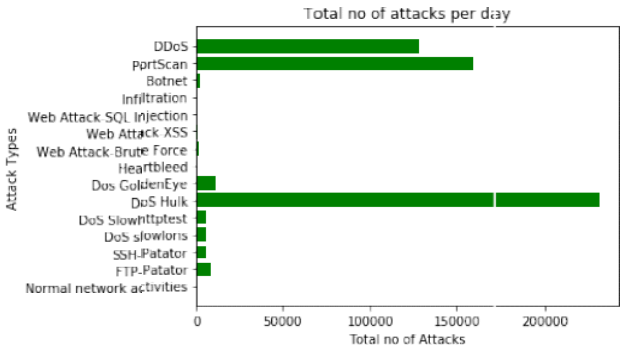

Fig 4. Total Number of Attacks.

Table 4: Total Number of Flows and Attacks per Day.

Day	Total Flows	No of Attacks	Attacks Type
Monday	529918	0	Normal network activities
Tuesday	445909	7938	FTP-Patator
		5897	SSH-Patator
Wednesday	692703	5796	DoS Slowloris
		5499	DoS Slowhttptest
		231073	DoS Hulk
		10293	Dos GoldenEye
		11	Heartbleed
Thursday Morning	170,366	1507	Web Attack - Brute Force
		652	Web Attack - XSS
		21	Web Attack - SQL Injection
Thursday Afternoon	288602	36	Infiltration
Friday Morning	191033	1966	Botnet
Friday Afternoon	286467	158930	PortScan
Friday Afternoon 2	225745	128027	DDoS
Total	**2830743**	**557646**	**19.70%**

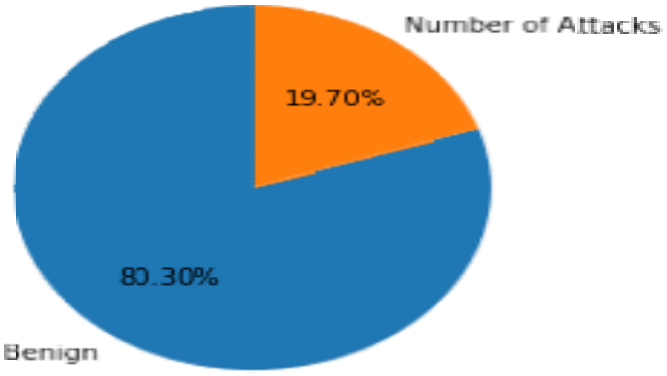

Fig 5: Data Distribution of Data Flow and Total Attacks.

Table 5: Features of CICIDS-2017 Dataset.

Feature No.	Feature	Feature No.	Feature	Feature No.	Feature
1.	Destination Port	28.	Bwd IAT Std	55.	AvgBwd Segment Size
2.	Flow Duration	29.	Bwd IAT Max	56.	Fwd Header Length
3.	Total Fwd Packets	30.	Bwd IAT Min	57.	FwdAvg Bytes/Bulk
4.	Total Backward Packets	31.	Fwd PSH Flags	58.	FwdAvg Packets/Bulk
5.	Total Length of Fwd Packets	32.	Bwd PSH Flags	59.	FwdAvg Bulk Rate
6.	Total Length of Bwd Packets	33.	Fwd URG Flags	60.	BwdAvg Bytes/Bulk
7.	Fwd Packet Length Max	34.	Bwd URG Flags	61.	BwdAvg Packets/Bulk
8.	Fwd Packet Length Min	35.	Fwd Header Len	62.	BwdAvg Bulk Rate
9.	Fwd Packet Length Mean	36.	Bwd Header Length	63.	SubflowFwd Packets
10.	Fwd Packet Length Std	37.	Fwd Packets/s	64.	SubflowFwd Bytes
11.	Bwd Packet Length Max	38.	Bwd Packets/s	65.	SubflowBwd Packets
12.	Bwd Packet Length Min	39.	Min Packet Length	66.	SubflowBwd Bytes
13.	Bwd Packet Length Mean	40.	Max Packet Length	67.	Init_Win_bytes_forward
14.	Bwd Packet Length Std	41.	Packet Length Mean	68.	Init_Win_bytes_backward
15.	Flow Bytes/s	42.	Packet Length Std	69.	act_data_pkt_fwd
16.	Flow Packets/s	43.	Packet Length Variance	70.	min_seg_size_forward
17.	Flow IAT Mean	44.	FIN Flag Count	71.	Active Mean
18.	Flow IAT Std	45.	SYN Flag Count	72.	Active Std
19.	Flow IAT Max	46.	RST Flag Count	73.	Active Max
20.	Flow IAT Min	47.	PSH Flag Count	74.	Active Min
21.	Fwd IAT Total	48.	ACK Flag Count	75.	Idle Mean
22.	Fwd IAT Mean	49.	URG Flag Count	76.	Idle Std
23.	Fwd IAT Std	50.	CWE Flag Count	77.	Idle Max
24.	Fwd IAT Max	51.	ECE Flag Count	78.	Idle Min
25.	Fwd IAT Min	52.	Down/Up Ratio	79.	Label
26.	Bwd IAT Total	53.	Average Packet Size		
27.	Bwd IAT Mean	54.	AvgFwd Segment Size		

6. EXPERIMENTAL RESULTS

each benign and assault facts facts for community links are protected in the CICIDS-2017 dataset; every linked data is made from 79 wonderful homes (desk 5). The CICIDS-2017 dataset became used in the following assessments with 10-fold go-validation to assess gadget performance. For function choice, the caution method uses the Recursive function removal method (RFE), focusing at the most vital and related feature.

Accuracy, precision, don't forget, and F-degree are calculated the usage of the WEKA software tool on CICIDS-2017 by means of NB, DT, LR, KNN, RF, SVM, and MLP with feature selection.22 crucial functions for Module I are listed in desk 6. Those 22 functions are the most critical for Module I. Moreover, these traits will be used to distinguish malicious and accurate information.

As will be distinct underneath, we've divided the experimental settings in Fig. 1 into four extraordinary stages.

- This test makes use of CSV records from the CICIDS-2017 dataset for gadgets to get to know. reproduction functions within the dataset need to be removed because of this. After that, the relabeling procedure might be accomplished .Following that, gadget getting to know CSV statistics is separated into two organizations: eighty% for training statistics and 20% for testing facts.
- capabilities are decided on from the schooling facts the usage of the Recursive feature removal method.
- The subset capabilities are then categorized the usage of Classifiers inclusive of Navie Bayes (NB), decision Tree (DT), Logistic Regression (LR), k-Nearest Neighbor (KNN), Random woodland (RF), aid Vector device (SVM), or Multilayer Perceptron (MLP). The analysis considers the characteristics of accuracy, precision, keep in mind, and F-degree.
- Following that, hybrid class algorithms which combine two or more techniques from step 3 have been used to calculate the four parameters of accuracy, precision, don't forget, and F-degree.
- In end, examine and contrast the output of the recommended hybrid algorithm in Module I (the pinnacle-appearing hybrid set of rules of step four) with seven distinct step 3 category algorithms. All checking out and getting to know procedures are executed using ten-fold go-validation.

The CICIDS-2017 dataset analyzed on this work the use of a ramification of systems to get to know algorithms to discover a simple and dependable

classifier. As a result, we have selected seven specific category algorithms from the WEKA-tool the use of a function choice algorithm: NB, DT, SVM, RF, KNN, Logistic Regression (LR), and MLP. The performance becomes then as compared using the F-degree, consider, accuracy, and precision. table 7 and Figs. 6, 7, eight, and nine display the results of seven device learning fashions at the CICIDS-2017 dataset.

Table 6: Predefined Selected Feature Using Recursive Feature Elimination algorithm.

S.No	Attribute No.	Name of Attribute
1	41	Packet Length Std
2	13	Total Length of Bwd Packets
3	65	SubflowBwd Bytes
4	8	Destination Port
5	42	Packet Length Variance
6	20	Bwd Packet Length Mean
7	54	Avg Bwd Segment Size
8	18	Bwd Packet Length Max
9	67	Init_Win_bytes_backward
10	12	Total Length of Fwd Packets
11	63	SubflowFwd Bytes
12	66	Init_Win_bytes_forward
13	52	Average Packet Size
14	40	Packet Length Mean
15	39	Max Packet Length
16	14	Fwd Packet Length Max
17	22	Flow IAT Max
18	36	Bwd Header Length
19	9	Flow Duration
20	26	Fwd IAT Max
21	55	Fwd Header Length
22	24	Fwd IAT Total

Table 7. Comparing the Performance of Seven Models on the CICIDS-2017 Data-set

Algorithms	Accuracy	Precision	Recall	F-measure
Naive Bayes	0.525	0.913	0.537	0.681
Random Forest	0.873	0.921	0.934	0.927
KNN	0.847	0.924	0.906	0.929
SVM	0.896	0.899	0.945	0.946
Decision Tree	0.832	0.936	0.894	0.914
Logistic Regression	0.897	0.899	0.955	0.946
MLP	0.913	0.915	0.958	0.954

Figure 6 (Table 7) compares the **Accuracy** of various classification models, including SVM, Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree , Logistic Regression, Navie Bayes. By outperforming other classifiers, the MLP achieves the maximum accuracy rate of around **91.30%,** according to the results of analysis.

Figure 7 (Table 7) compares the **Precision** of various classification models, including Support Vector Machine, Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Random Forest (RF), DT, LR, NB, The analysis reveals that the MLP outperforms the other classifiers, with the highest Precision rate of about **91.50%.** The range of precision attained by the other classifiers was **52.50% to 89.60%.**

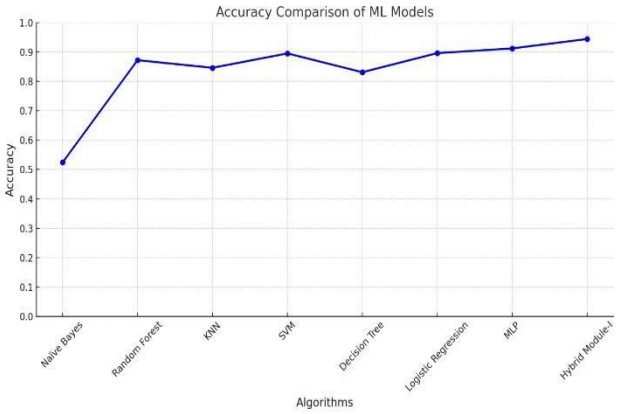

Fig 6. Seven machine studying fashions' accuracy the usage of the CICIDS-2017 dataset.

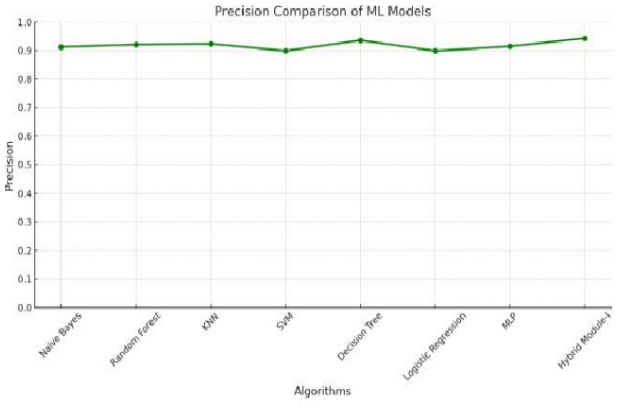

Fig 7. Seven machine learning models' Precision using the CICIDS-2017 dataset

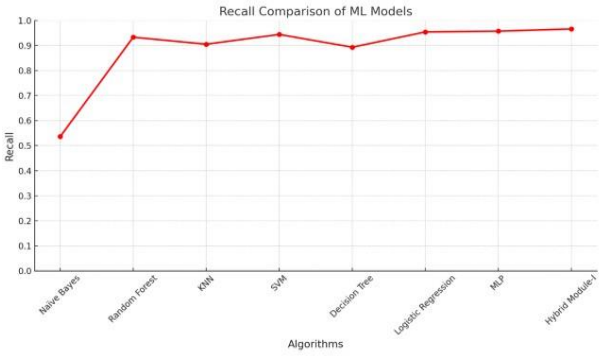

Fig 8 Seven machine learning models' Recall using the CICIDS-2017 data-set

Figure 8 (Table 7) suggests an evaluation of numerous fashions for type based totally on don't forget, along with Navie Bayes (NB), selection Tree, Logistic Regression , k-Nearest Neighbor (KNN),

Random woodland (RF), assist Vector machine (SVM), or Multilayer Perceptron (MLP). Analytical effects display that the MLP outperforms different classifiers, accomplishing the most remembered rate of ninety-five.

Figure 9 (Table 7). Demonstrates F-measures primarily based comparison of various class fashions, such as Navie Bayes (NB), choice Tree (DT), Logistic Regression (LR), ok-Nearest Neighbor (KNN), Random Forest (RF), help Vector gadget (SVM), or Multilayer Perceptron (MLP). The analytical results show that the MLP outperforms the opposite classifiers and achieves the very best F-degree rate of ninety-five.40%.

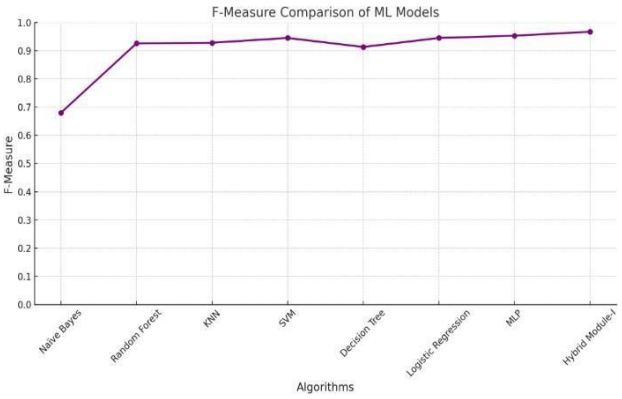

Fig 9. Seven machine learning models' F-measure using the CICIDS-2017 dataset.

Table 8. Performance Comparison of Seven Machine Learning Models with Hybrid Module on the CICIDS-2017 Data-set

ALGORITHMS	ACCURACY	PRECISION	RECALL	F-MEASURE
Naïve Bayes	0.525	0.913	0.537	0.681
Random Forest	0.873	0.921	0.934	0.927
KNN	0.847	0.924	0.906	0.929
SVM	0.896	0.899	0.945	0.946
Decision Tree	0.832	0.936	0.894	0.914
Logistic Regression	0.897	0.899	0.955	0.946
MLP	0.913	0.915	0.958	0.954
Proposed Approach (Hybrid Module) LR-MLP	**0.945**	**0.943**	**0.967**	**0.968**

Table 9: Comparing the Performance of Different Hybrid Algorithms.

ALGORITHMS	ACCURACY	PRECISION	RECALL	F-MEASURE
RF-SVM	0.883	0.932	0.942	0.936
RF-LR	0.893	0.886	0.949	0.949
RF-MLP	0.892	0.927	0.938	0.934
SVM-LR	0.898	0.899	0.947	0.947
SVM-MLP	0.907	0.901	0.946	0.948
Proposed Approach (Hybrid Module) LR-MLP	**0.945**	**0.943**	**0.967**	**0.968**

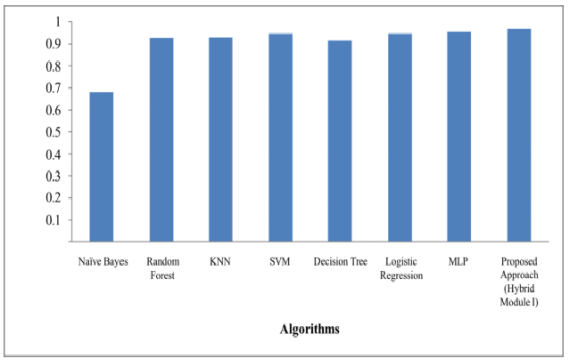

Fig 10. Comparing the Accuracy of seven machine learning models with the Hybrid Module on the CICIDS-2017 data-set

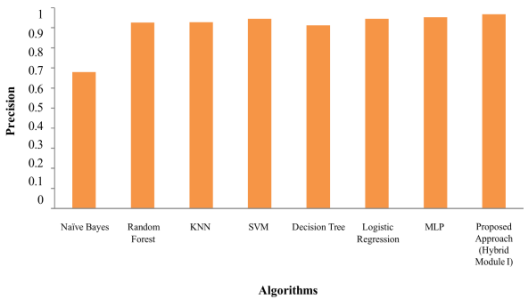

Fig 11. Comparing the Precision of seven machine learning models with the Hybrid Module on the CICIDS-2017 data-set

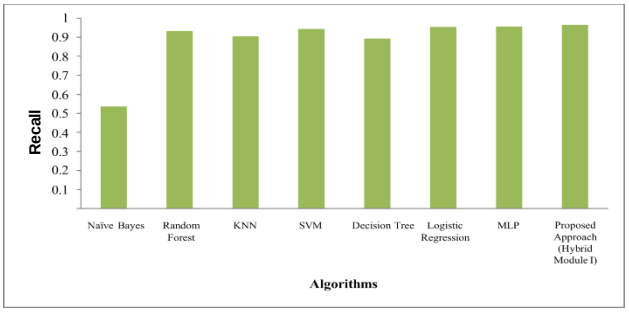

Fig 12. Comparing the Recall of seven machine learning models with the Hybrid Module on the CICIDS-2017 dataset.

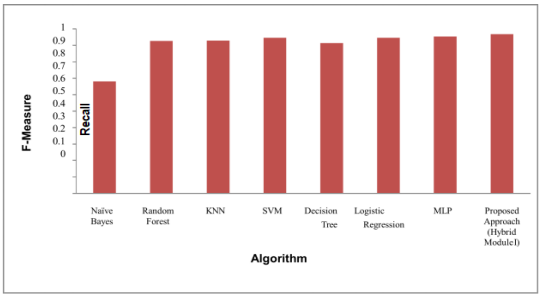

Fig 13. Comparing the F-measure of seven machine learning models with the Hybrid Module on the CICIDS-2017 data-set

Table 8 and Figures 10, 11, 12, and 13 use the CICIDS-2017 dataset to compare seven machine learning models with the Hybrid Module. Using the hybrid algorithm, we found that, in comparison to other machine learning methods, the Hybrid module with Feature Selection offers the highest accuracy (94.5%) (Fig. 10 and Table 8).In addition, we discovered that when comparing the precision (94.30%), F-measure (96.80%), and recall (96.70%) features with other machine learning techniques, the Hybrid module produces the greatest results (Table 8 and Fig 10, 11, 12, and 13).

7. CONCLUSION AND FUTURE WORK

Conclusion

Computer system pc gadget intrusion assaults are growing because of the big use of neighborhood networks and the internet. network protection solutions have developed because of the fast growth of laptop networks. It created the requirement for a device that relies on intrusion prevention technologies and can discover community threats. identifying such dangers enables fending off more attacks similarly to offering statistics on damage evaluation. With the usage of an intrusion detection gadget, those attempts are often recognized. because one-of-a-kind networks have specific safety problems, researchers have created intrusion detection structures for various contexts. The Intrusion Detection system's features are collected to study any potential security breaches and to examine information from diverse components of a network or computer. Intrusion detection and different safety technologies like firewalls, cryptography, and authentication have won huge reputation within the ultimate ten years. The literature evaluation analyzes the scope and boundaries of several gadgets to get to know techniques. It appears at techniques for characteristic choice, preprocessing, and normalization that considerably beautify device studying model overall performance. The evaluation evaluation, however, shows that those models require development. Consequently, a hybrid approach is proposed in the study.

To secure data in cloud computing and detect abnormalities or intrusions on the CICIDS2017 dataset, we present a novel hybrid intrusion detection method in this study. Based on machine learning methods, we have proposed a novel hybrid model. due to the large range of 9aaf3f374c58e8c9dcdd1ebf10256fa5 up to date attacks (Binary or Multiclass class), the CICIDS-2017 dataset becomes used. Firstly, we achieved statistics practice and function choice, which includes removing functions from the dataset which might be irrelevant, redundant, or unrelated to other attributes. ultimately, there are seven machines to get to know intrusion detection strategies which can be broadly utilized, and every have precise characteristics. using the WEKA gadget studying device, we examined seven machine learning strategies at the CICIDS 2017 dataset: Naive Bayes (NB), selection Tree (DT), Logistic Regression (LR), okay-Nearest Neighbor (KNN), Random woodland (RF), guide Vector machine (SVM), and Multilayer Perceptron (MLP).The evaluation of the proposed and present gadget learning algorithms demonstrates that the Proposed algorithm show the higher outcomes. The cautioned hybrid model plays higher and reaches ninety-four.50% Accuracy, in step with the results. Similarly, we located that the advised module produces the greatest consequences while compared to different system learning techniques in terms of Precision (94.30%), F-measure (96.70%), and keep in mind (ninety six.80%) parameters.

Future Work

Developing an efficient intrusion detection system (IDS) that can detect intrusions based on the investigation is the aim of this project. The following could be included in future work:

1. comparing various feature selection techniques while working with a large dataset.
2. To improve feature extraction and classification. A variety of deep learning and classification techniques can be used to improve the suggested model.
3. To improve the algorithms' efficiency by combining clustering and feature selection.
4. Explore using unsupervised learning to train models on security-related unlabeled datasets.

ACKNOWLEDGMENT

This work is acknowledged under Integral University manuscript No. IU/R&D/2026-MCN0004458

REFERENCES

1. W. Lee and S. J. Stolfo, A Framework for Constructing Features and Models for Intrusion Detection Systems, ACM Transactions on Information and System Security, vol. 3, no. 4. 2001.
2. Y. Xin et al., “Machine Learning and Deep Learning Methods for Cybersecurity,” IEEE Access, vol. 6, pp. 35365–35381, 2018, doi: 10.1109/ACCESS.2018.2836950.

3. D. Zhu, G. Premkumar, X. Zhang, and C.-H. Chu, "Data Mining for Network Intrusion Detection: A Comparison of Alternative Methods," Decis. Sci., vol. 32, no. 4, pp. 635–660, Dec. 2001, doi: 10.1111/j.1540-5915.2001.tb00975.x.
4. Z. Liu and Y. Lai, "A Data Mining Framework for Building Intrusion Detection Models Based on IPv6," 2009, pp. 608–618. doi: 10.1007/978-3-642-02617-1_62.
5. R. R. Bouckaert, E. Frank, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, "WEKA Manual for Version 3-7-2," 2002, [Online]. Available:http://netcologne.dl.sourceforge.net/project/weka/documentation/3.7.x/WekaManual-3-7-12.pdf
6. M. C. Belavagi and B. Muniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection," Procedia Comput. Sci., vol. 89, pp. 117–123, 2016, doi: 10.1016/j.procs.2016.06.016.
7. K. Alrawashdeh and C. Purdy, "Toward an Online Anomaly Intrusion Detection System Based on Deep Learning," in 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2016, pp. 195–200. doi: 10.1109/ICMLA.2016.0040.
8. G. Zhao, C. Zhang, and L. Zheng, "Intrusion Detection Using Deep Belief Network and Probabilistic Neural Network," in 22017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Jul. 2017, pp. 639–642. doi: 10.1109/CSE-EUC.2017.119.
9. S. S. Roy, A. Mallik, R. Gulati, M. S. Obaidat, and P. V. Krishna, "A Deep Learning Based Artificial Neural Network Approach for Intrusion Detection," Mathematics and Computing, pp. 44–53, 2017. doi: 10.1007/978-981-10-4642-1_5.
10. C. Xu, J. Shen, X. Du, and F. Zhang, "An Intrusion Detection System Using a Deep Neural Network with Gated Recurrent Units," IEEE Access, vol. 6, pp. 48697–48707, 2018, doi: 10.1109/ACCESS.2018.2867564.
11. K. Wu, Z. Chen, and W. Li, "A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks," IEEE Access, vol. 6, pp. 50850– 50859, 2018, doi: 10.1109/ACCESS.2018.2868993.
12. Z. Wang, "Deep Learning-Based Intrusion Detection with Adversaries," IEEE Access, vol. 6, pp. 38367–38384, 2018, doi: 10.1109/ACCESS.2018.2854599.
13. N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," IEEE Trans. Emerg. Top. Comput. Intell., vol. 2, no. 1, pp. 41–50, Feb. 2018, doi: 10.1109/TETCI.2017.2772792.
14. L. Li, Y. Yu, S. Bai, Y. Hou, and X. Chen, "An Effective Two-Step Intrusion Detection Approach Based on Binary Classification and NN," IEEE Access, vol. 6, pp. 12060–12073, 2018, doi: 10.1109/ACCESS.2017.2787719.
15. M. H. Kamarudin, C. Maple, T. Watson, and N. S. Safa, "A LogitBoost-Based Algorithm for Detecting Known and Unknown Web Attacks," IEEE Access, vol. 5, pp. 26190–26200, 2017, doi: 10.1109/ACCESS.2017.2766844.
16. Y. Jia, M. Wang, and Y. Wang, "Network intrusion detection algorithm based on deep neural network," IET Inf. Secur., vol. 13, no. 1, pp. 48–53, Jan. 2019, doi: 10.1049/iet-ifs.2018.5258.
17. P. Gogoi, D. K. Bhattacharyya, B. Borah, and J. K. Kalita, "MLH-IDS: A Multilevel Hybrid Intrusion Detection Method," Comput. J., vol. 57, no. 4, pp. 602– 623, Apr. 2014, doi: 10.1093/comjnl/bxt044.
18. Y. Gao, Y. Liu, Y. Jin, J. Chen, and H. Wu, "A Novel Semi-Supervised Learning Approach for Network Intrusion Detection on Cloud-Based Robotic System," IEEE Access, vol. 6, pp. 50927–50938, 2018, doi: 10.1109/ACCESS.2018.2868171.
19. M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm," IEEE Trans. Comput., vol. 65, no. 10, pp. 2986–2998, Oct. 2016, doi: 10.1109/TC.2016.2519914.
20. S. Wang, J. Tang, and H. Liu, "Feature Selection," in Encyclopedia of Machine Learning and Data Mining, Boston, MA: Springer US, 2016, pp. 1–9. doi: 10.1007/978-1-4899-7502-7_101-1.
21. H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, "A New K-Nearest Neighbors Classifier for Big Data Based on Efficient Data Pruning," Mathematics, vol. 8, no. 2, p. 286, Feb. 2020, doi: 10.3390/math8020286.
22. R. Vijayanand, D. Devaraj, and B. Kannapiran, "Intrusion detection system for wireless mesh network using multiple support vector machine classifiers with genetic-algorithm-based feature selection," Comput. Secur., vol. 77, pp. 304–314, Aug. 2018, doi: 10.1016/j.cose.2018.04.010.
23. E. Besharati, M. Naderan, and E. Namjoo, "LR-HIDS: logistic regression host-based intrusion detection system for cloud environments," J. Ambient Intell. Humaniz. Comput., vol. 10, no. 9, pp. 3669–3692, Sep. 2019, doi: 10.1007/s12652-018-1093-8.
24. M. S. Mok, S. Y. Sohn, and Y. H. Ju, "Random effects logistic regression model for anomaly detection," Expert Syst. Appl., vol. 37, no. 10, pp. 7162–7166, Oct. 2010, doi: 10.1016/j.eswa.2010.04.017.
25. C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," IEEE Access, vol. 5, pp. 21954– 21961, 2017, doi: 10.1109/ACCESS.2017.2762418.
26. F. Jiang et al., "Deep Learning Based Multi-Channel Intelligent Attack Detection for Data Security," IEEE Trans. Sustain. Comput., vol. 5, no. 2, pp. 204–212, Apr. 2020, doi: 10.1109/TSUSC.2018.2793284.
27. K. Kumari, "Linear regression analysis study," Journal of the Practice of Cardiovascular Sciences, 2018, doi: 10.4103/jpcs.jpcs.

28. Medical Data in Crosshairs: Why Is Healthcare an Ideal Target? 14 August 2021. Available online: https://www.trendmicro. com/vinfo/us/security/news/cyber-attacks/medical-data-in-the-crosshairs-why-is-healthcare-an-ideal-target (accessed on 15 May 2020).
29. Conaty-Buck, S. Cybersecurity and healthcare records. Am. Nurse Today 2017, 12, 62–64.
30. E GOVERNMENT, Cloud Computing Initiatives. 22 April 2021. Available online: https://www.bahrain.bh/ (accessed on 15 July 2021).
31. He, D.; Qiao, Q.; Gao, Y.; Zheng, J.; Chan, S.; Li, J.; Guizani, N. Intrusion detection based on stacked autoencoder for connected healthcare systems. IEEE Netw. 2019, 33, 64–69.
32. R. R. Bouckaert, E. Frank, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, "WEKA Manual for Version 3-7-2," 2002, [Online]. Available:http://netcologne.dl.sourceforge.net/project/weka/documentation/3.7.x/WekaManual-3-7-12.pdf.
33. Khan, S., & Abbas, S. H. (2024). A review of machine learning-based security in cloud computing. GIS Science Journal, 11(10), 321-340. ISSN No: 1869-9391.
34. Khan, S., & Abbas, S. H. (2025). Enhancing Cloud Data Security using a Hybrid Cryptographic Model: A Combination of Advanced Encryption Standard and Elliptic Curve Cryptography. Journal of Information Systems Engineering and Management, 10(34s), 1-13. e-ISSN: 2468-4376.
35. Srivastava, M. K., Akhtar, J., Ahmad, D., Ansari, H., Maurya, K. C., Khan, S., Faiz, M., Ali, S., Ranjan, R., & Abbas, S. H. (2025). The future of surgery: A guide to machine learning for surgeons. *Journal of Neonatal Surgery, 14*(13s), 37–45.
36. Abbas, S. H., Ranjan, R., Maurya, B., Warsi, A. H., & Khan, S. (2025). Evaluating healthcare providers' perceptions, expertise, and barriers regarding the adoption of AI in rehabilitation. *Cuestiones de Fisioterapia, 54*(3), 4423–4439.