International Journal for
Electronic Crime Investigation

ISSN: 2522-3429 (Print)
ISSN: 2616-6003 (Online)

DOI: https://doi.org/10.54692/ijeci.2025.0902/255

Research Article Vol. 9 issue 2 Jul-Dec 2025

Securing 5G Network Infrastructure Against DDoS Threats Using ML-Based Anomaly Detection

Shahzaib Hassan[1], Alishba Tabassum[1], Lubna Nadeem[1], Yasar Amin[1], Tariq Mahmood[23]

[1]Department of Telecommunication Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan,

[2]Department of Information Science, University of Education, Lahore, Pakistan,

[3]Artificial Intelligence and Data Analytics (AIDA) Lab, CCIS Prince Sultan University, Riyadh, 11586, Kingdom of Saudi Arabia

Corresponding Author: lubna.nadeem@uettaxila.edu.pk

Received: Sep 18,2025; **Accepted:** Sep 30,2025; **Published:** Oct 17,2025

ABSTRACT

Today, millions of people and devices use the Internet to carry out daily activities, but the growing reliance on the Internet comes with major security concerns. Older security systems and traditional detection techniques are out of date because attackers continue to find new and smarter ways of penetrating networks. They are just not precise enough to stay in the race. This research discusses how that gap can be filled by machine learning (ML). Although in cybersecurity, ML has demonstrated potential, accuracy remains reliant on the selection of the appropriate models and the concentration on the most important parts of the data. Although ML has already shown its potential, our work aims at refining the approach to increase detection accuracy. The most promising among the techniques tested was the Random Forest (RF) algorithm, which had an impressive accuracy rate of 99.84%. This clearly indicates that our proposed system is far better than the previous methods, showing its capability to detect malicious activities.

Keywords: 5G networks, security threat, distributed denial of service (DDoS), machine learning

1. INTRODUCTION

Internet networks continue to expand globally due to technological advances that include smartphones, computers, communication systems, and IoT devices [1]. Research indicates that there exist more than 5 billion smart devices worldwide, while 3 billion users actively use the internet [2]. The widespread use of Internet networks produces enormous amounts of data second by second, which presents major challenges in protecting information against cyber threats [3]. Computer systems and their networks are dependent on cybersecurity to protect them from unauthorized access [4]. Data protection, along with privacy assurance, functions as a fundamental structure that protects organizations and states as well as individual users. Data transmitted on the Internet remains exposed to hacking and manipulation attempts by cybercriminals [5]. The 2017 cyberattack damages reached $5 billion, and analysts predict that this amount will increase to $6 trillion yearly starting in 2021 [6]. Distributed denial-of-service (DDoS) attacks represent one of the most common cybersecurity threats that cause servers and networks to crash when flooded with excessive data packets [7]. The number of distributed denial-of-service at tacks has increased significantly in recent times. A major DDoS attack on Amazon Web Services' (AWS) Amazon Simple Storage Service (S3) and other platforms generated a severe service disruption in February 2020, which lasted approximately eight hours [8]. The recorded attack, which stood out as one of the largest, reached its peak performance level at 2.3 terabytes per second. The research reported by Security Week shows that DDoS attacks occur 28,700 times a day on the Internet [9]. The rising need for strong cybersecurity systems able to detect cyber-attacks effectively drives the current market demand. The goal of cybersecurity professionals is to create IDS (Intrusion Detection Systems) that detect known threats and new attacks without producing false alerts [10]. Modern cyber-attacks especially DDoS attacks require intelligent detection methods because multiple existing intrusion detection approaches exist. Modern intrusion attempts have rendered traditional IDS solutions ineffective, according to research [11]. Artificial intelligence techniques have become necessary for cybersecurity practice and have achieved great success in all fields. The practice has proven successful in all areas since it became mandatory. The capability of big data exploration 1, through hidden models reaches tremendous heights because of their ability to discover patterns in data. Through ML techniques, organizations can detect and monitor network-based attacks [12]. Various studies used different ML techniques for intrusion detection. Some deficiencies remain in this approach, including the determination process. The re searchers have chosen basic and effective features to enhance the performance of ML techniques [13].

CONTRIBUTION

The main contributions are as follows.

- To build an innovative detection framework that will detect DDoS attacks accurately.
- Select important features that would boost the accuracy and efficiency of DDoS attack detection.
- Testing and comparing different machine learning models to see which could detect threats with the highest accuracy.
- Comparison of different machine learning models to see which one could detect threats more accurately.
- Selecting the optimal model, which can then be tested using open-source datasets through standard performance measurements.

2. PAPER ORGANIZATION

Section III discusses the Literature Survey, and Section IV depicts the Gaps in research and the Motivation of our work. Then Section V is about Proposed Methodology. Section VI explains Key Performance Indicators. Section VII provides details about Dataset. Then Section VIII is about our main research contributions. Section IX explains in detail the proposed system model. Further, Section X demonstrates the Workflow of our research. Section XI is related to simulation results and discussion. Finally, Section XII is on

Conclusion and Future Recommendations.

3. LITERATURE SURVEY

The rapid technological advancements and widespread Inter net of Things (IoT) devices have created a situation where people increasingly depend on internet networks, so robust security must protect user privacy and data. Relevant research shows that artificial intelligence presents itself as an efficient method for handling cybersecurity threats. Research through multiple studies has investigated the usage of Intrusion Detection Systems (IDS) that employ both Machine Learning (ML) and Deep Learning (DL) methods. This section discusses multiple research approaches that use ML and DL methods to identify cyber-attacks. The research team of Bindra and Sood evaluated six ML techniques including Logistic Regression (LR), K-Nearest Neighbors (KNN), Random Forest (RF), Naïve Bayes (NB), Linear Support Vector Machine (SVM) along with Linear Discriminant Analysis (LDA) to identify the best method for DDoS attack detection [14]. RF demonstrated the best accuracy rate of 96.5% according to tests conducted on the CIC IDS dataset which outranked all other analysis methods. Chavan et al. evaluated DDoS attack detection by analyzing KNN, SVM, Decision Tree (DT), and LR as four ML techniques in their work [15]. The LR model showed superior accuracy results by reaching 90.4%. Combined methods used in decision-making produce better accuracy levels than single classification systems. Das, Saikat along with coauthors developed an ensemble model which unites Multilayer Perceptron (MLP) with SVM and KNN and DT as base ML techniques [16]. The researchers performed experiments on the NSL-KDD dataset, which demonstrated that their ensemble classifier achieved superior results than standalone models in the study. The Auto Encoding (AE) method proposed by Kasim serves both to reduce features and boost traffic classification efficiency [17]. Kasim proposed the application of Auto-Encoding (AE) for both feature selection and dimension reduction to achieve traffic classification. AE methods help reduce dimensions for effective traffic classification according to [18]. The researchers conducted performance tests using both CICIDS2017 and NSL-KDD data sets. The model achieved successful classification results in testing datasets according to the performance studies. Bhardwaj et al. [19] presented a method that merges well-stacked sparse AE. The approach uses Deep Neural Networks together with Deep Neural Network (DNN) feature learning to detect possible DDoS attacks. Highly efficient DL techniques discovered big data thanks to their exceptional discovery capabilities. Multiple groups have made systematic attempts to utilize this topic for cybersecurity research. Al-Emadi et al. [20] analyzed how DL techniques function within CNN and RNN systems for network intrusion detection. Table 1 shows the comparison of the proposed work with existing works.

4. RESEARCH GAPS &MOTIVATION

As 5G networks develop, telecommunications have reached new levels of speed, responsibility, and connectivity. Such improvements are used to facilitate critical services and extensive device communication. This improvement also entails an increase in security threats, especially DDoS attacks that can hamper vital operations. Creating secure 5G systems is more relevant than ever, and increases the necessity for smarter, adaptive security solutions that will ensure reliable and uninterrupted connectivity. The existing DDoS detection techniques can't effectively deal with the scale and complexity of a 5Genvironment. They are usually not flexible enough to cope with rapidly changing patterns of traffic and advanced threats. To fill this gap, our research suggests using a machine learning–based detection framework. It focuses on feature se lection to improve detection accuracy, tries different models, and determines the most successful approach to employment based on open datasets and conventional evaluation metrics.

5. PROPOSED METHODOLOGY

The proposed work develops an effective network intrusion detection system by combining ML

algorithms with feature selection strategies. The study conducts performance assessments on the intrusion detection of four ML methods RF, KNN, SVM, and DT. The system utilizes feature selection techniques to detect important features.

Table 1. Comparison of Proposed work with Existing works

Study	Year	Model	Dataset	Best ML Accuracy
[18]	2021	CNN, LSTM, and CLSTMNet	NSL-KDD	CLSTMNet (99.28%)
[14]	2020	CNN and RNN	NSL-KDD	CNN (97.01%)
[10]	2019	RF, LR, NB, KNN, Linear SVM, and LDA	CIC IDS	RF -96.50%
[15]	2020	AE+ SVM	CICIDS2017 & NSL-KDD	AE+ SVM (96.36%)
[6]	2020	CNN	NSL-KDD	CNN -99.30%
[5]	2019	Ensemble model, MLP, SVM, KNN, and DT	NSL-KDD	Ensemble model -99.10%
Proposed Work	2025	RF, KNN, SVM, and DT	NSL-KDD	RF -99.84%

Table 2. NSL-KDD Dataset Features

Index	Feature	Index	Feature
1	src_bytes	22	dst_host_same_src_port_rate
2	dst_host_srv_count	23	same_srv_rate
3	num_access_files	24	dst_host_count
4	logged_in	25	dst_bytes
5	serror_rate	26	dst_host_srv_serror_rate
6	su_attempted	27	srv_rerror_rate
7	num_access_files	28	num_file_creations
8	root_shell	29	num_compromised
9	is_host_login	30	protocol_type
10	count	31	num_shells
11	duration	32	diff_srv_rate
12	srv_serror_rate	33	dst_host_srv_sror_rate
13	num_root	34	srv_count
14	land	35	service
15	dst_host_diff_srv_rate	36	urgent
16	wrong_fragment	37	hot
17	is_guest_login	38	dst_host_srv_count
18	serror_rate	39	flag
19	num_failed_logins	40	class
20	rerror_rate	41	num_outbound_cmds
21	dst_host_same_srv_rate	42	srv_diff_host_rate

5.1. DATASET

The research utilized NSL-KDD dataset because it represents a clean and refined version. The KDD data set received the traffic information from which it was built. The NSL KDD data set contains 148,517 structured samples with 42 characteristics (Table 2 provides their list)

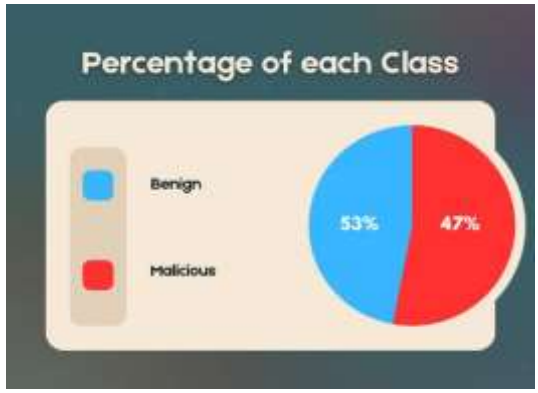

Figure 1. Pie Chart of Classification

5.2. TRADITIONAL METHOD USED FORANOMALY DETECTION

5.2.1 Decision Tree (DT)

Decision Trees represent one of the most applied non parametric supervised learning methods, which functions for both classification and regression tasks. The system arranges itself into a branching pattern that extends from the root node through decisions based on established rules.

5.2.2. Random Forest (RF)

The supervised learning method Random Forest creates various decision trees as an ensemble approach to maximize both regression and classification success rates. Combining multiple trees into one model improves overall model performance.

5.2.3. Support Vector Machine (SVM)

SVM operates as a supervised machine learning model that mainly provides classification functions. SVM operates through identifying the optimum division (or hyperplane) that distinguifies different classification categories in a data set.

5.2.4. K-Nearest Neighbors (KNN)

The simple KNN algorithm serves as a solution for classification and regression problems.

6. PROPOSED MODEL

This model diagram depicts the DDoS attack structure within IoT-enabled 5G networks. The architecture implements several connected layers beginning with the IoT gateway and extending through the backhaul network as well as operator network cloud and external network services that face DDoS attack risks which lead to operational disruptions.

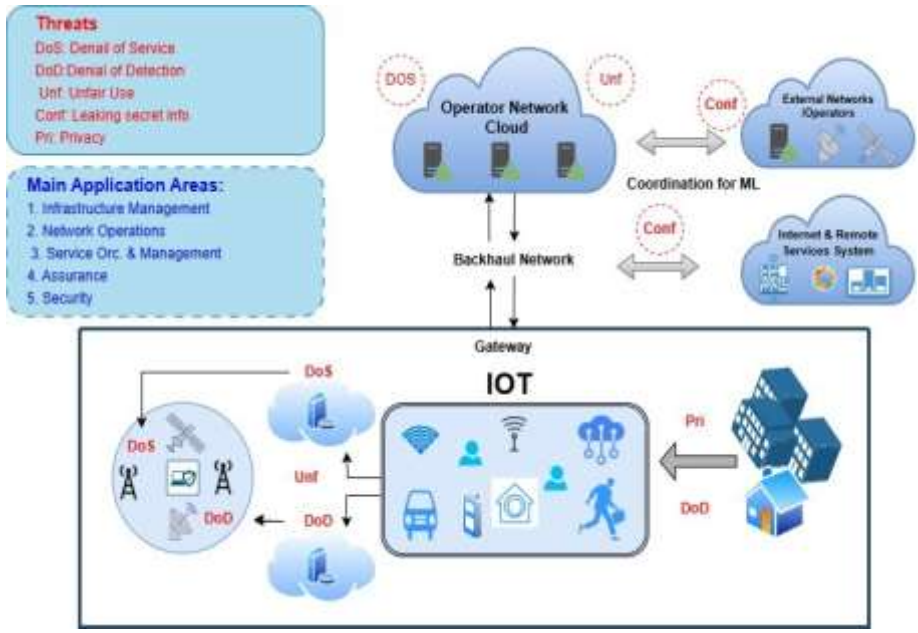

Figure 2. System Model

6.1. DDOS ATTACK THREATS IN IOT-5G NETWORKS

A Distributed Denial of Service attack uses network re- sources to an overwhelming point thus preventing legitimate users from accessing services. The model demonstrates the essential locations which DDoS attacks launch from and spread between.

1) **IoT Devices & Gateway:** Unsecure IoT devices allow attackers to create botnets which send large amounts of traffic to IoT gateways thus causing service degradation.
2) **Backhaul Network:** The traffic overload traverses through the operator network cloud until it reaches its maximum processing capacity and bandwidth thresh- old.
3) **Operator Network Cloud:** The cloud infrastructure in Operator Network Cloud suffers from network congestion along with service unavailability triggered by malicious requests.
4) **External Networks & Remote Services:** Cyber at- tackers can strike external networks and remote ser- vices to cause extensive service breakdowns among linked systems.

6.2. DDOS ATTACK PROPAGATION & IMPACT

The propagation of DDoS attacks happens through com- promised IoT nodes that serve as botnet members to send excessive request floods. The impact includes:

1) Excessive traffic on gateway servers as well as cloud systems result in both longer communication response times and data transmission failures.
2) The excessive request flow during DDoS attacks uses up all available CPU power together with memory capacity and network bandwidth thus blocking access to services.

3) Iot attacks against critical infrastructure result in permanent breakdowns of medical care delivery and residential automation systems and industrial automation networks.

6.3. WORKFLOW

1) **Dataset Selection & Preparation:** We used the NSL- KDD dataset, which is well-suited for detecting DDoS attacks due to its labeled network traffic data.
2) **Feature Selection:** From the dataset, we extracted the most relevant features that significantly contribute to DDoS detection.
3) **Algorithm Selection:** We plan to use machine learning algorithms like DT, SVM, RF, and KNN. These algorithms are selected based on their proven effectiveness in detecting DDoS attacks.
4) **Model Training & Validation:** We split the dataset into training and testing sets. We trained the models using Python libraries like Scikit-learn, ensuring their reliability through cross-validation techniques.
5) **Evaluation Metrics:** To evaluate model performance, we used metrics like accuracy, precision, recall, F1- score, and specificity. A comparison of these results helped us select the best algorithm for real-time application.

6.4. KEY PERFORMANCE INDICATORS

Five quality measures evaluated the performance of ML techniques for evaluation purposes. The evaluation of ML techniques depends on five measures including Accuracy, Precision, Sensitivity, Specificity and F1-Score. The classification value of '1' represents positive outcomes in this analysis. While benign samples are considered negative and represented by '0'. All performance measure formulas are represented below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times (\frac{Precision \times Recall}{Precision + Recall})$$

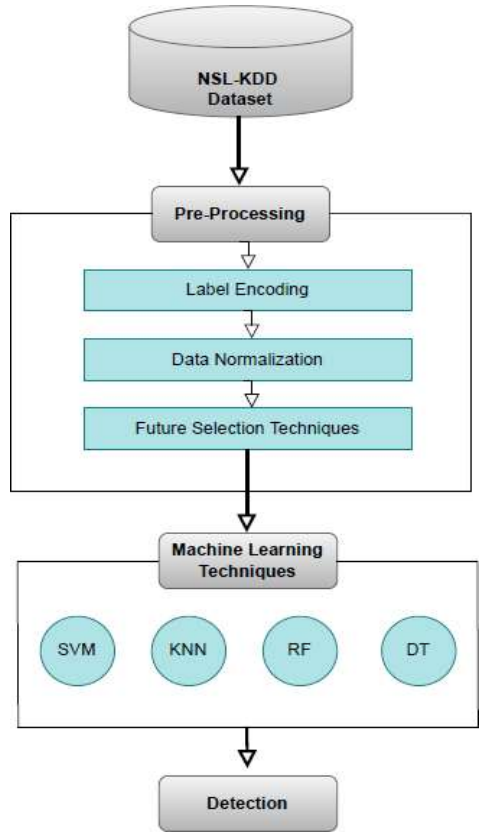

Figure 3. System WorkFlow

TP	FN
FP	TN

Figure 4. Parameters Table

- TP refers to cases where malicious samples receive proper detection as malicious.
- The detection of benign samples as benign falls under True Negatives category.
- The system identifies benign samples incorrectly as malicious through its FP output.
- The detection system marks malicious samples as benign when they are already identified as harmful.

7. RESULTS AND DISCUSSION

7.1 PERFORMANCE EVALUATION

8. The four Machine Learning (ML) techniques (RF, DT, KNN, and SVM) exist within Scikit Learn library which functions as a robust Python library for implementing ML development. One of the most potent libraries for building and implementing ML techniques are Scikit learn.

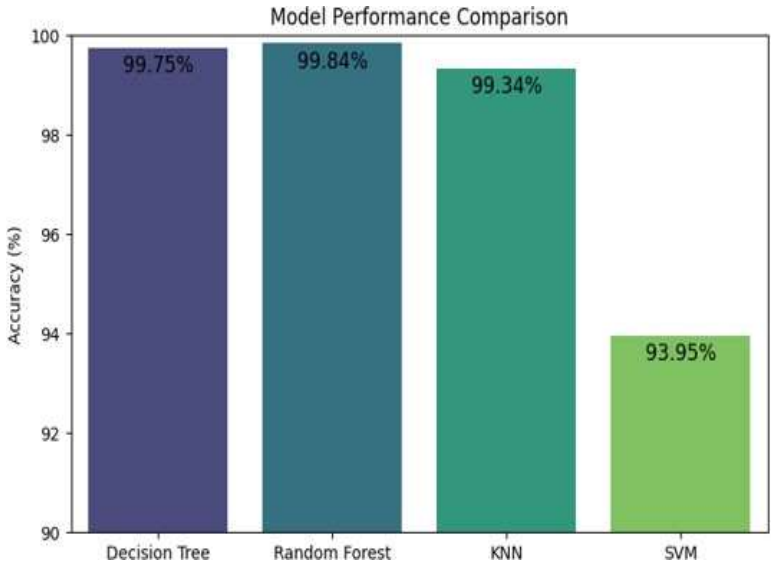

Figure 5. Accuracy Result

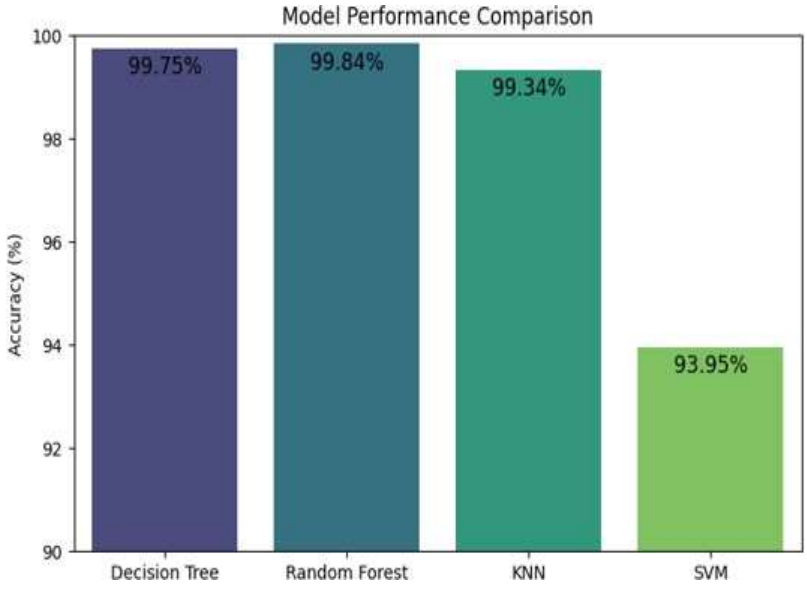

Figure 6. Other Parameters Result

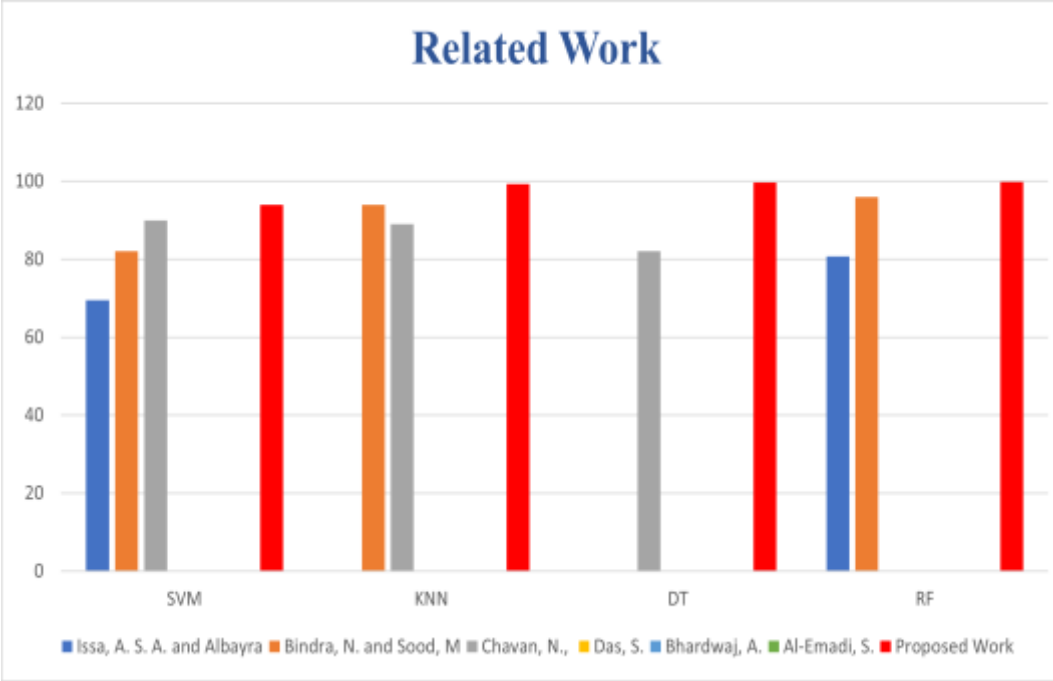

Figure 7. Related Work

Figure 7 depicts in detail the comparison of different ML techniques with the proposed work. It is observed that the accuracy of the proposed technique is better than all the existing techniques in literature.

The performance metrics of the four ML techniques are presented in Table 3.

Table 3. Results

Technique	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Support Vector Machine	93.95%	94.50%	94.50%	94.50%
Random Forest	99.84%	99.90%	99.90%	99.90%
K-Nearest Neighbors	99.34%	99.00%	99.00%	99.00%
Decision Tree	99.75%	99.90%	99.90%	99.90%

The research dataset split its content into 80% training material and 20% testing material. A comparison of the four techniques occurred on just twenty-five available features. The selected dataset features underwent feature selection before usage in this study. The RF method demonstrated in Figure 5 the best accuracy performance among all techniques. 99.84%, superior to the rest of the techniques. The KNN technique obtained accuracy results of 99.34% while the second-best accuracy rate went to the DT technique with 99.75%. The accuracy rate for SVM technique came in as the lowest at 93.95% but the RF technique led with 99.84%. Both the DT technique and KNN technique performed with respective accuracies of 99.75% and 99.34%.

Figure 6 demonstrates the comparison of ML techniques with the other parameters like Precision, Recall F1 score. The confusion matrix of the four ML techniques also appears in Figure 8 to Figure 11. Each confusion matrix represents model's performance regarding the separation of various classes. The application of pre-processing techniques alongside feature selection methods creates necessary steps for implementing ML methods. A model demonstrates superior data preprocessing implementation. The performance accuracy would rise after conducting critical feature testing and pre- processing. Figure 8 illustrates that Random Forest (RF) model had correctly assigned most of the samples with few misclassifications. It has strong class prediction abilities of all classes. From figure 9 it can be observed that the Decision Tree (DT) model was good; however, it committed a bit of errors when compared with the Random Forest (RF) especially on differentiating similar threat types. From Figure 10, the K- Nearest Neighbors (KNN) model had difficulties with class boundaries and thus it performed worse than other models with more misclassifications. It is more sensitive to data imbalance. In Figure 11, the Support Vector Machine (SVM) had good performance though some of the classes were not well separated which was as a result of overlapping feature distributions.

The research utilized RF technique which produced superior results than other RF studies. The RF technique employed here exceeded previous applications of RF technique in other research studies. In addition, the KNN, SVM, and DT. The KNN along with SVM and DT techniques from this study achieved superior performance results compared to other KNN and SVM and DT techniques. An RF model which utilized feature selection methods achieved the recommendations. The recommended model based on feature selection techniques and RF classifications shows promising and reliable future performance.The performance of this model exceeded every measurement in this study. The pro- posed model study draws data from the mentioned research papers.

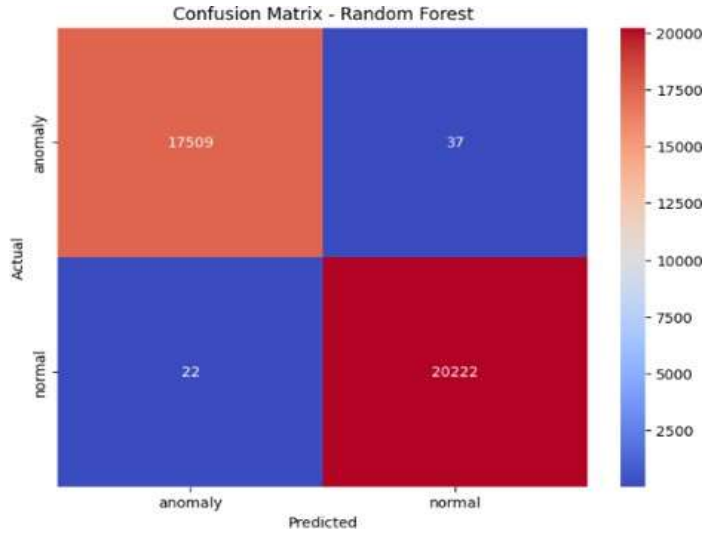

Figure 8. Confusion Matrix of RF

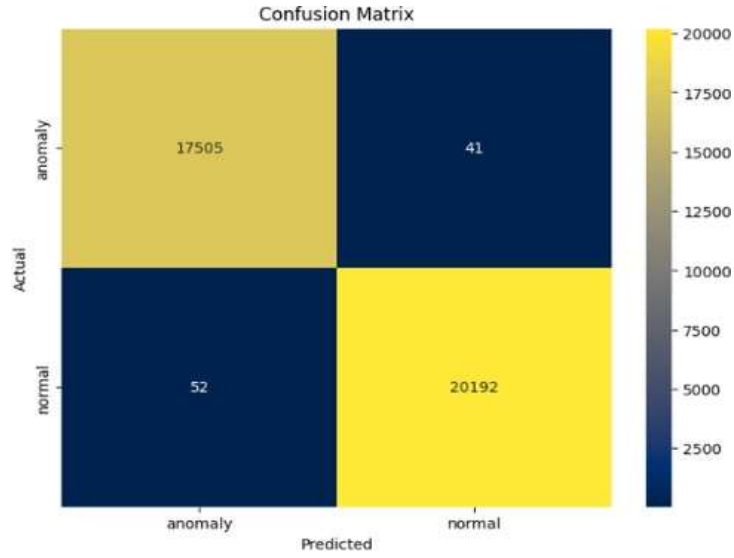

Figure 9. Confusion Matrix of DT

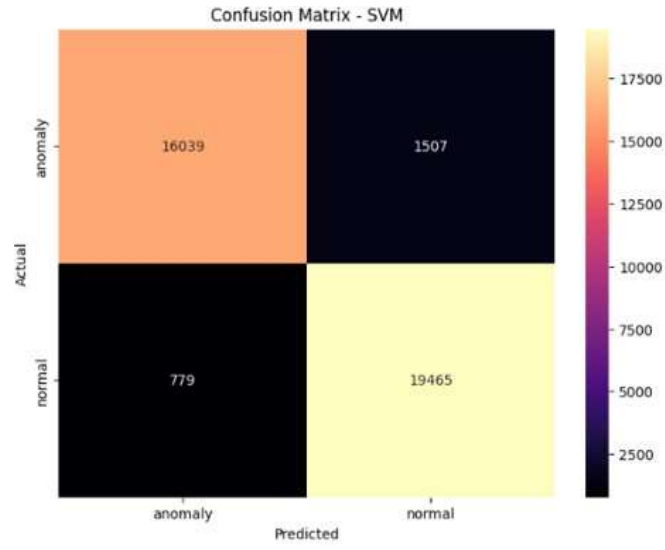

Figure 10. Confusion Matrix of SVM

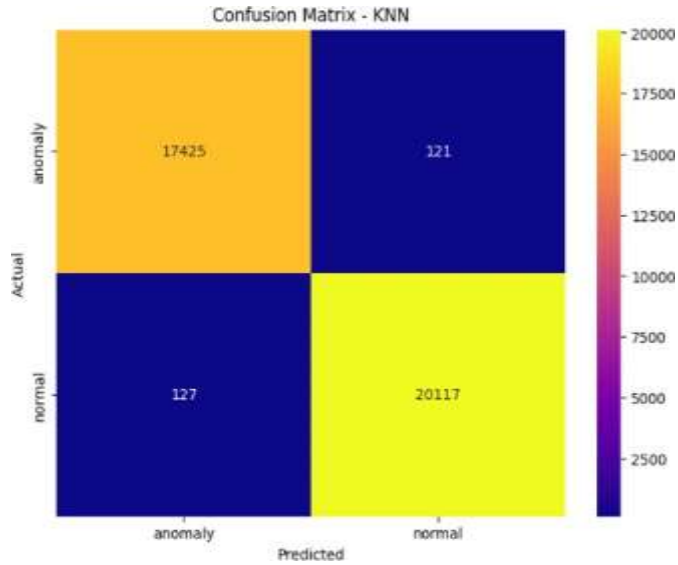

Figure 11. Confusion Matrix of DT

7.2. FUTURE RECOMMENDATIONS

Future work needs to employ different machine learning methods to enhance detection accuracy together with error reduction. Real-time detection technology protected by fast response systems serves as an essential measure to enhance 5G network protection. Deep learning techniques

combined with novel methods of choosing detection features would improve threats detection capabilities. New security systems which present the capability to broadcast threat information between different networks will result in better overall security responses.

9. CONCLUSION

The detection of DDoS attacks in 5G networks becomes effective through the utilization of Machine Learning (ML) algorithms that include Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree (DT). Random Forest yielded the best accuracy rate at 99.84% among the applied Machine Learning algorithms which demonstrates the potential of ML for network protection. The combination of ML algorithms with feature selection methods delivers better detection results by converting traditional IDS shortcomings into effective solutions when it comes to highly complex 5G cyberattacks.

10. REFERENCES

[1] D. Dasgupta, Z. Akhtar, and S. Sen, "Machine learning in cybersecurity: A comprehensive survey," *The Journal of Defense Modeling and Simulation*, vol. 19, no. 1, pp. 57–106, 2022.

[2] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani, "A survey of machine and deep learning methods for Internet of Things (IoT) security," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1646–1685, 2020.

[3] A. Salih, S. T. Zeebaree, S. Ameen, A. Alkhyyat, and H. M. Shukur, "A survey on the role of artificial intelligence, machine learning and deep learning for cybersecurity attack detection," in *Proc. 7th Int. Eng. Conf. "Research & Innovation amid Global Pandemic" (IEC)*, 2021, pp. 61–66.

[4] S. R. Zeebaree, K. Jacksi, and R. R. Zebari, "Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 510–517, 2020.

[5] I. Avci and M. Koca, "Cybersecurity attack detection model using machine learning techniques," *Acta Polytechnica Hungarica*, vol. 20, no. 7, pp. 29–44, 2023.

[6] W. Tong, L. Lu, Z. Li, J. Lin, and X. Jin, "A survey on intrusion detection system for advanced metering infrastructure," in *Proc. 2016 6th Int. Conf. on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*, 2016, pp. 33–37.

[7] R. Cohen, "Cloud attack: Economic denial of sustainability (EDoS)," 2009. [Online].

[8] C. Cimpanu, "AWS said it mitigated a 2.3 Tbps DDoS attack, the largest ever," *ZDNet*, 2020. [Online].

[9] J. Reo, "Academic research reports nearly 30,000 DoS attacks per day," 2021. [Online].

[10] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine learning and deep learning approaches for cybersecurity: A review," *IEEE Access*, vol. 10, pp. 19 572–19 585, 2022.

[11] A. S. A. Issa and Z. Albayrak, "CLSTMNet: A deep learning model for intrusion detection," *Journal of Physics: Conference Series*, vol. 1973, no. 1, p. 012244, 2021.

[12] M. Abdullahi, Y. Baashar, H. Alhussian, A. Alwadain, N. Aziz, L. F. Capretz, and S. J. Abdulkadir, "Detecting cybersecurity attacks in Internet of Things using artificial intelligence methods: A systematic literature review," *Electronics*, vol. 11, no. 2, p. 198, 2022.

[13] A. A. Salih and M. B. Abdulrazaq, "Combining best features selection using three classifiers in intrusion detection system," in *Proc. 2019 Int. Conf. on Advanced Science and Engineering (ICOASE)*, 2019, pp. 94–99.

[14] Kaggle, "NSL-KDD Dataset," 2019. [Online]. Available: https://www.kaggle.com/datasets/hassan06/n

slkdd. Accessed: Dec. 16, 2019.

[15] N. Bindra and M. S. Sood, "Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset," *Automatic Control and Computer Sciences*, vol. 53, no. 5, pp. 419–428, 2019.

[16] N. Chavan, M. Kukreja, G. Jagwani, N. Nishad, and N. Deb, "DDoS attack detection and botnet prevention using machine learning," in *Proc. 2022 8th Int. Conf. on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, 2022, pp. 1159–1163.

[17] S. Das, A. M. Mahfouz, D. Venugopal, and S. Shiva, "DDoS intrusion detection through machine learning ensemble," in *Proc. 2019 IEEE 19th Int. Conf. on Software Quality, Reliability and Security Companion (QRS-C)*, 2019, pp. 471–477.

[18] Ö. Kasim, "An efficient and robust deep learning-based network anomaly detection against distributed denial of service attacks," *Computer Networks*, vol. 180, p. 107390, 2020.

[19] A. Bhardwaj, V. Mangat, and R. Vig, "Hyperband tuned deep neural network with well-posed stacked sparse autoencoder for detection of DDoS attacks in cloud," *IEEE Access*, vol. 8, pp. 181 916–181 929, 2020.

[20] S. Al-Emadi, A. Al-Mohannadi, and F. Al-Senaid, "Using deep learning techniques for network intrusion detection," in *Proc. 2020 IEEE Int. Conf. on Informatics, IoT, and Enabling Technologies (ICIoT)*, 2020, pp. 171–176.