JCBI

Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 08 Issue 02
2025

Research Article
https://doi.org/10.56979/802/2025

Intelligent Cyber Security Framework for Threat Detection using Ensemble Learning Techniques

Talha Bin Tariq[1], Saima Noreen Khosa[2], Muhammad Zubair Hadi[3], Tanzeela Kiran[1], Maria Mansab[1], Urooj Akram[1], and Muhammad Faheem Mushtaq[1,*]

[1]Department of Artificial Intelligence, The Islamia University of Bahawalpur, 63100, Bahawalpur, Pakistan.
[2] Department of Information Technology, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan.
[3]Department of Computer Science, The Islamia University of Bahawalpur, 63100, Bahawalpur, Pakistan.
[*]Corresponding Author: Muhammad Faheem Mushtaq. Email: faheem.mushtaq@iub.edu.pk

Received: February 07, 2025 **Accepted:** February 28, 2025

Abstract: Cyber security is critical in today's fast-paced digital landscape. As AI-driven solutions become indispensable for safeguarding enterprises, the escalating volume and complexity of cyber threats frequently overwhelm conventional security measures, resulting in significant financial and reputational risks. To address this challenge, this study proposes an advanced cyber security framework based on an ensemble learning model that combines machine learning and deep learning algorithms. Using the HIKARI-2021 dataset (Kaggle), we evaluated and compared multiple classifiers, including Random Forest, Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, Logistic Regression, Multi-Layer Perceptron, and Convolutional Neural Network. By integrating these models through an ensemble approach, we leveraged their complementary strengths, achieving a notable 96.32% accuracy—a significant improvement over individual models. Beyond accuracy, the ensemble method enhances adaptability, enabling more dynamic and resilient security frameworks. Our findings highlight the efficacy of ensemble learning in cyber security, demonstrating its potential to fortify digital enterprises against evolving threats. This research not only advances practical solutions but also paves the way for future studies on AI-integrated cyber security, fostering innovation and robust digital infrastructure globally.

Keywords: Cyber Security; AI-driven Solutions; Ensemble Learning; Security Frameworks; Digital Infrastructure

1. Introduction

In today's fast-moving digital world, classical security methods have not kept pace with the changes which are initiated by modern technology and the ever-evolving nature of cyber threats [1]. Hence, cybersecurity is one of the greatest concerns of enterprises today [2]. The risk of cyberattacks increases as more companies, governments, and people use digital devices and the internet because more sensitive data is kept online, hackers and other cybercriminals target it. A gap between the evolving threats and static security measures could prove disastrous to the company's financial standing and reputation [3]. Nevertheless, many studies have focused on cybersecurity strategies based on machine learning and deep learning approaches [4]. These approaches have shown a certain similarity or effectiveness in aiding in detecting and modifying cyber threats [5]. By identifying and offering reliable solutions to digital firms, the objective is to strengthen businesses against evolving cyber threats and enhance predictive capabilities through the use of classifiers and the development of sophisticated, adaptable security frameworks that achieve secure digital infrastructure through AI-evolved means.

Older security systems depend on fixed rules and predefined patterns, and thus often fail against unexamined or new threats. AI and machine learning have the capability of learning and adapting to spotting abnormal patterns and recognizing possible cyber-attacks [6]. However, sometimes using any one model does not provide the best output. Ensemble learning makes use of multiple models that combine their advantages into protective solutions improving system security. Therefore, it is now suggested that ensemble learning method of combining several classifiers in a bid to gain optimum performance by leveraging their respective advantages very effective in further ensuring cybersecurity [7]. An ensemble-meta-classifier cognitive cybersecurity strategy utilizing preprocessed security indicators to create a robust dataset [8]. The study was effective against data incompleteness and diversity [9]. High-quality preprocessing provides potential bias. A lack of high-speed adaptability in model variance concerning the environment used makes model generalizability a bar. Since, cyber threats are evolving constantly therefore, the models should be updated continuously to maintain their effectiveness [10]. Cyberattacks are ever-growing complex and ever-harder to be detected [11]. Current methods at times warn too many false alerts or miss new types of threats. All such limitations highlight the challenges and the necessity of constant upgrades for models used concerning cybersecurity [12].

This research addresses the limitations of existing studies and proposes an ensemble learning model by combining deep learning and machine learning models. The models will include implementations that leverage the best aspects of multiple classifiers to raise the level of predictive accuracy along with developing advanced security frameworks capable of nurturing innovations in cybersecurity. The research attempts to build an innovative system in cybersecurity by ensemble various machine learning and deep learning models to improve the accuracy of detection of cyber threats or errors, designed to best fight the dynamically changing threat landscape.

The HIKARI-2021 dataset from Kaggle will be used, which encompasses cybersecurity incidents and their parameters, such as flow time, source and destination IP addresses, and labeled identifications for various categories of cyberattacks. To train on this dataset, a variety of classifiers were used, including random forest, decision tree, Gaussian Naive Bayes, K-Nearest neighbors, logistic regression, multi-layer perceptron, and convolutional neural networks. The performance of each model was evaluated using widely recognized criteria like recall, accuracy, and precision. This study used an ensemble learning approach after carefully weighing the benefits and drawbacks of these standalone models. This approach produced a combination of predictions based on the strengths of classifiers, increasing the robustness of the authentication process and achieving an overall accuracy gain of 96.32%. This research aims to close the limitations of single models and improve their predictive powers. The contributions of this research are as follows:

- Proposed a reliable and efficient intrusion detection system using ensemble learning methods based on machine learning and deep learning models.
- Developed an innovative approach by addressing class imbalance and preprocessing the HIKARI-2021 dataset. This involves converting categorical variables into numerical form using label encoding and one-hot encoding techniques to ensure data quality and balance, thereby enhancing the accuracy of the classification models.
- Implemented and evaluate the multiple ensemble models, such as Random Forest, Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, Logistic Regression, Multi-Layer Perceptron, and Convolutional Neural Network, demonstrating the effectiveness of ensemble techniques in achieving high accuracy and robustness.
- The performance evaluation is conducted using evaluation parameters such as recall, F1-score, accuracy, and precision, highlighting the efficacy of the ensemble models in enhancing classification accuracy and reliability.

The rest of the paper is arranged as follows: Section 2 explains the earlier literature relevant to the proposed methodology; Section 3 describes the methodology, including the dataset and the applications of the deep learning and machine learning-based models; Section 4 contains the results and discussion; and Section 5 closes the study.

2. Related Work

Many studies have focused on improving fraud detection systems, but most face three key limitations: reliance on outdated and limited data sets, poor detection rates with high false alarms, and inability to adapt to evolving cyber threats. The proposed ensemble method directly addresses these gaps by combining machine-learning and deep-learning approaches. This combined strategy not only overcomes the inherent limitations of individual models but also establishes a more effective security framework against modern cybercrime.

The Ensemble Classifier Algorithm with a Stacking Process (ECASP) for bot detection demonstrates how even advanced machine learning approaches face fundamental challenges - while achieving 94.08% accuracy through optimal feature selection [10], its dependence on predetermined features creates vulnerabilities against evolving botnet structures that demand continuous updates. Similarly, the XGBoost-based intrusion detection framework achieves 92.86% accuracy on the CICIDS2017 dataset [11], yet this success remains constrained by the dataset's limited coverage of emerging real-world threats, ultimately restricting its effectiveness against novel attack patterns. These examples reveal how current solutions struggle with adaptability and generalization - limitations our ensemble method specifically addresses through dynamic feature adaptation and comprehensive threat coverage in the HIKARI-2021 dataset.

Furthermore, the machine-learning-based detections of network attacks is explored using three datasets, namely, UNSW-NB15, NSL-KDD, and BoT-IoT [12]. Feature selection is done using information gain and Pearson correlation, followed by classification with various models. Stacked LightGBM and random forest achieve the highest prediction accuracy across all datasets; however, its dependence on some predefined datasets may not reflect emerging zero-day attacks and evolving cyber threats, which may then possibly cripple the real-world adaptability. Another study analyzes the application of ensemble machine learning techniques in network intrusion detection using the NSL-KDD dataset [13]. It applied voting, bagging, and boosting methods specifically used in the evaluation of the voting classifier, random forest, and AdaBoost algorithms for boosting the efficiency of anomaly detection. The study is largely limited in its reliance on the NSL-KDD dataset, which may not adequately represent modern-day cyber threats, further constricting the generalizability of the proposed ensemble methods in real-world attacks.

Cyber Threat Intelligence (CTI)-based method for malicious URL detection using two-stage ensemble learning was proposed [14]. For better detection accuracy, CTI features are extracted from Google searches and data. Reclassification is done using a Random Forest (RF) algorithm, followed by the execution of a Multilayer Perceptron (MLP) for final decision-making. The model provides an overall 7.8% higher accuracy with a 6.7% lower false-positive from the level of traditional URL-based detection techniques. However, it relies on external sources, such as Google searches and Whois data, where delays within feature extraction may limit the real-time detection performance. This study [15], presents an ensemble learning-based approach for the detection of Cross-Site Scripting (XSS) attacks using a combination of multiple Bayesian networks, to include domain knowledge and threat intelligence to include. An analysis approach that ranks nodes according to their effect on the output will allow for better model interpretation for users. Head-to-head comparisons prove the model works on a real dataset and that it does particularly well in detecting large-scale attacks. The method thus depends on precise domain knowledge and threat intelligence, which limits adaptability to completely new or evolving attack patterns.

The HIKARI-2021 dataset stands out as an excellent benchmark for network intrusion detection systems by addressing critical gaps in existing datasets. Where NSL-KDD achieves 89% accuracy and CICIDS2017 reaches 92.86%, our model attains 96.32% on HIKARI - a 4-7% improvement in detecting modern attacks. Unlike traditional datasets, HIKARI combines both real and synthetic encrypted attack traffic including DDoS, port-scanning, and malware, providing superior coverage versus BoT-IoT's narrower IoT focus [16]. This unique composition directly tackles the limitations of previous approaches by offering a solid foundation for developing and testing ML and DL models [17] that can handle both known and emerging threats. Where traditional signature-based detection fails against novel attacks and anomaly detection methods struggle with false alarms [18], our ensemble approach leverages HIKARI's robust coverage to overcome these persistent challenges. By incorporating diverse machine learning models (random forest, decision tree, Gaussian Naive Bayes) [19] and deep learning architectures (CNNs, MLPs) [20], we demonstrate how HIKARI enables more effective intrusion detection compared to conventional datasets.

The efficiency of the IoT Intrusion Detection System is improved with the aid of supervised machine learning and ensemble classifiers [21]. Different models including random forest, decision tree, logistic regression, and k-nearest neighbor were trained on the TON-IoT dataset [22]. They were combined using both stacking and voting methods. The ensemble classifiers outperformed the individual classifiers with over 89.63% accuracy, thus improving IDS reliability and reducing classification errors. However, research focused only on binary classification between normal and abnormal traffic would probably restrict their ability to detect or discriminate between specific attack types and so limit their applicability in the real-world multi-class IoT threat.

Recent studies [23] reveal both the promise and limitations of machine learning in cyber threat detection. While models like random forest, decision tree, support vector machine, and k-nearest neighbors show strong classification potential, and deep learning approaches (CNNs, RNNs) excel at complex pattern recognition, significant challenges remain in achieving robust real-world performance. Ensemble methods like bagging and boosting [24] have improved accuracy by combining classifiers, yet they still face persistent gaps in handling sophisticated attack patterns - particularly when trained on conventional datasets. This explains the growing emphasis on hyperparameter tuning and grid search [25] to optimize performance. The integration of hybrid datasets like HIKARI-2021 [26] marks a crucial advancement, as our work demonstrates by combining these approaches to overcome the adaptability limitations of prior ensemble methods while maintaining their accuracy benefits.

3. Proposed Methodology

An ensemble learning-based approach to improve intrusion detection systems has been proposed. The framework takes on several steps as follows: firstly, the HIKARI-2021 dataset was taken from Kaggle. The preprocessing steps are applied to ensure the quality and balance of the dataset. Next, the dataset was split into two parts, 80% for training and 20% for testing. Various classifiers like Random Forest, Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, Logistic Regression, Multi-Layer Perceptron, and Convolutional Neural Network were used on the training data. The performance of individual classifiers was compared and evaluated on the basis of recall, F1-score, accuracy, and precision. An ensemble model is then used to combine the strengths of these individual classifiers and obtain the highest accuracy. Hyperparameter tuning was performed to optimize the performance of each model. The proposed ensemble method yielded marked improvements in detection accuracy, thus confirming that combining ML and DL methods provides robust intrusion detection. This approach helps tackle the disadvantages of single models and improves the prediction abilities, helping develop an advanced security framework and also reinforcing organizations against ever-evolving cyber threats.

3.1 Dataset HIKARI-2021

The HIKARI-2021 dataset provides a comprehensive range of options for the evaluation of a network intrusion detection system. It contains about 1.2 million records, out of which 80% is training data (960,000 records), while the remaining 20% forms the testing data (240,000 records). It is composed of both synthetic and real encrypted attack traffic generated by malware, port scanning, and DDoS attacks. The variety of this dataset alone offers some degree of attack vectors for the training and assessment of deep learning and machine learning classification models. Integrating true and synthetic data allows a reasonably good simulation of the real world via the HIKARI-2021 dataset by the researchers for developing different intrusion detection systems with efficacy and adaptability.

3.2 Data Preprocessing

A key factor for consideration during data analysis is preprocessing. It guarantees data transformation from its raw unprocessed state to a format suited for deep learning and machine learning techniques. Several indispensable steps in the preprocessing phase that are included to boost model accuracy.

A. Removing Null Values:

Machine learning and deep learning algorithms expect categorical entities such as protocol types or service names contained in network security datasets like the HIKARI-2021 dataset to be assigned a numeric representation. Categorical conversion is performed using different ways including label encoding and one-hot encoding. One-hot encoding creates n binary columns, yes or no, where n is the number of categories, whereas label encoding assigns different numbers for different categories. The choice of which method to use depends on the algorithm involved and the type of data. Hence, categorical data should be treated properly, so that the models are reliable and functional.

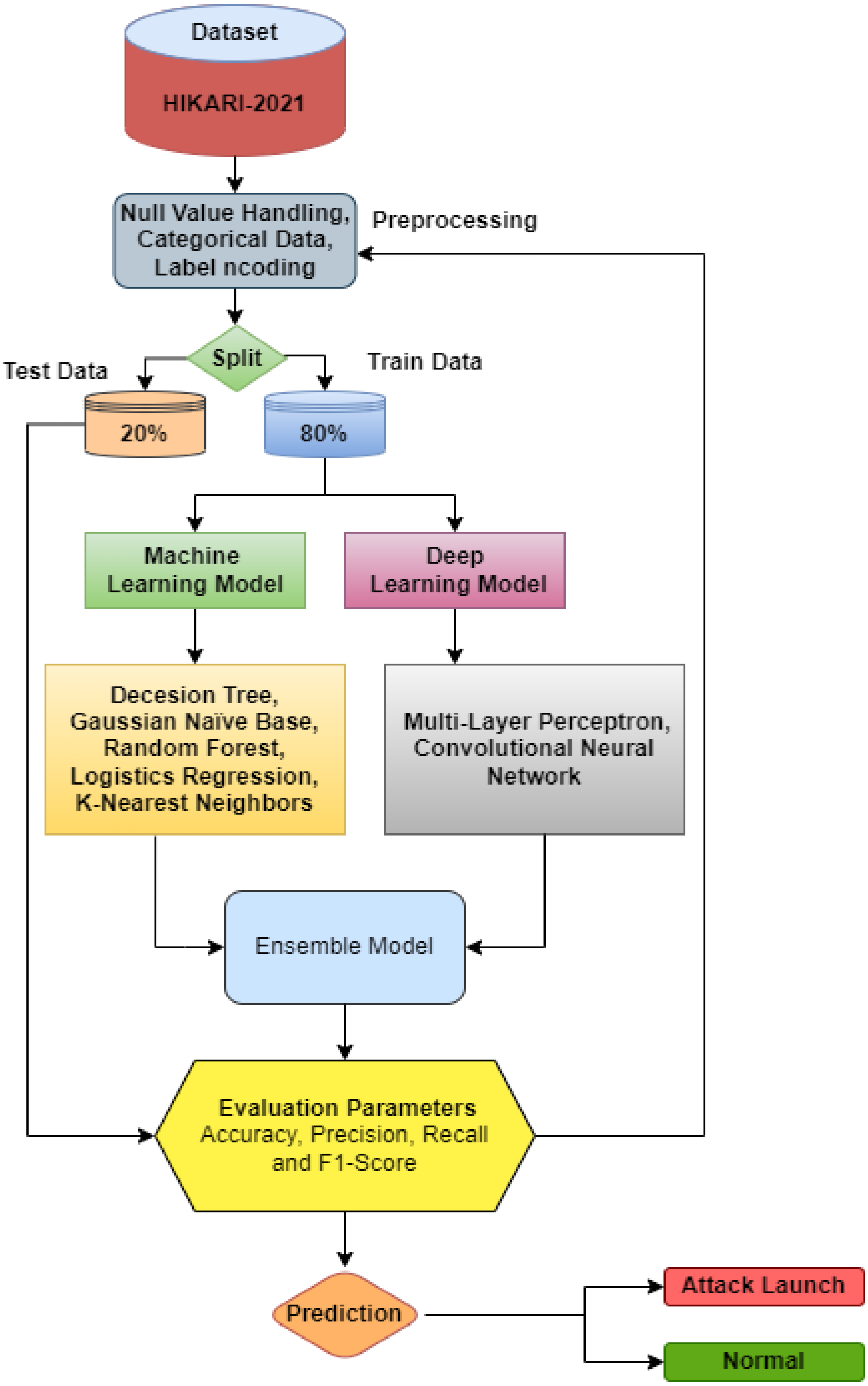

Dataset
HIKARI-2021
Null Value Handling, Categorical Data, Label ncoding
Preprocessing
Split
Test Data
Train Data
20%
80%
Machine Learning Model
Deep Learning Model
Decesion Tree, Gaussian Naïve Base, Random Forest, Logistics Regression, K-Nearest Neighbors
Multi-Layer Perceptron, Convolutional Neural Network
Ensemble Model
Evaluation Parameters Accuracy, Precision, Recall and F1-Score
Prediction
Attack Launch
Normal

Figure 1. Intelligent Cyber Security Framework using Ensemble Learning Technique

B. Handling Categorical Data

This study obtained the HIKARI-2021 dataset from Kaggle. There are about 2 million records in this dataset. There were certain preprocessing steps carried out where categorical variables were transformed into numerical form employing label encoding and one-hot encoding techniques. The transformation enabled compatibility with the machine learning and deep learning algorithms that led to an efficient and accurate training and testing of the models.

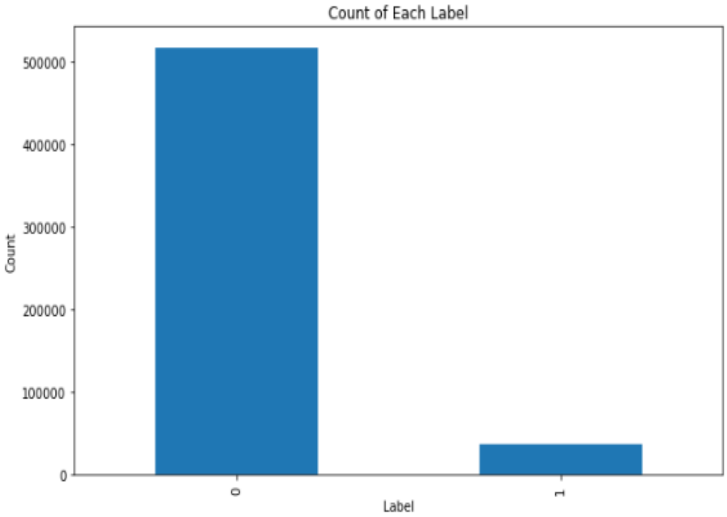


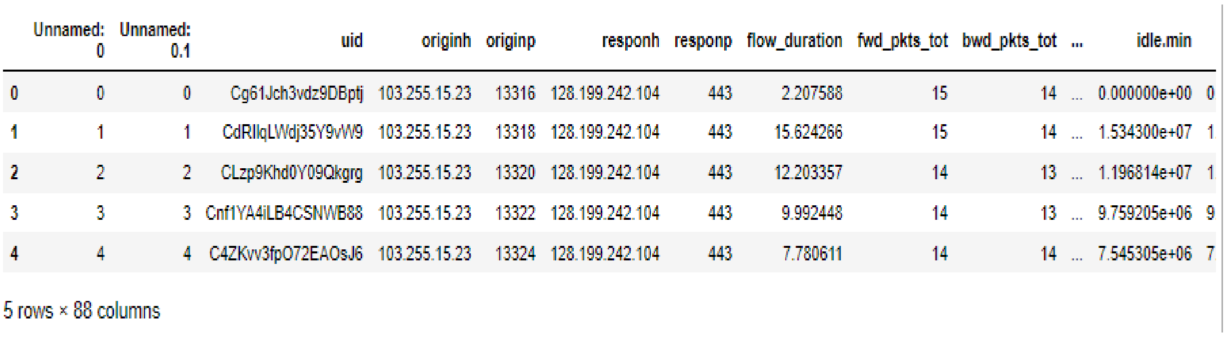
Figure 2. Label Count

C. Label Encoding

Label encoding is an important first preprocessing step for converting categorical data into a proper numerical format for machine learning and deep learning models. In the HIKARI-2021 dataset, the labels for detection, 'normal' and 'attack', are transformed into label 0 and label 1 respectively. This attribute eventually helps in improving the performance and effectiveness of the intrusion detection system by enabling the algorithms to comprehend and learn from data during model training.

3.3 Train-Test Split

Essentially, a train-test split is necessary to prepare the HIKARI-2021 dataset for analysis and modeling tasks in this study. The dataset is divided into 2 subsets: a training dataset and a testing dataset. 80% of the records, roughly 960,000, constitute the training dataset-a huge mass, whereas the rest, about 240,000, are the testing dataset. Training data is then fed into machine learning and deep learning models to learn data patterns and characteristics. After that, the trained models are evaluated using the testing data so as to determine their capability to generalize from novel data. This split is important in ascertaining the models' effectiveness in detecting intrusions and their corresponding accuracy and reliability in real situations. This train-test split thus intends to build robust and trustworthy models useful in intrusion detection to deal with the mainly changing and diverse cyber threats.



	Unnamed: 0	Unnamed: 0.1	uid	originh	originp	responh	responp	flow_duration	fwd_pkts_tot	bwd_pkts_tot	...	idle.min	
0	0	0	Cg61Jch3vdz9DBptj	103.255.15.23	13316	128.199.242.104	443	2.207588	15	14	...	0.000000e+00	0
1	1	1	CdRIlqLWdj35Y9vW9	103.255.15.23	13318	128.199.242.104	443	15.624266	15	14	...	1.534300e+07	1
2	2	2	CLzp9Khd0Y09Qkgrg	103.255.15.23	13320	128.199.242.104	443	12.203357	14	13	...	1.196814e+07	1
3	3	3	Cnf1YA4iLB4CSNWB88	103.255.15.23	13322	128.199.242.104	443	9.992448	14	13	...	9.759205e+06	9
4	4	4	C4ZKvv3fpO72EAOsJ6	103.255.15.23	13324	128.199.242.104	443	7.780611	14	14	...	7.545305e+06	7

5 rows × 88 columns

Figure 3. HIKARI-2021 Data-Set

3.4 Machine Learning Models

Machine learning is a group of artificial intelligence that includes the development of algorithms and techniques that help machines learn from data and improve performance over time. Unlike traditional programming, where a user needs to provide explicit instructions, machine learning algorithms extract patterns from data and then use them to predict an outcome, classify data or improve some process. It is the nature of adaptive learning that enables these machines to improve with time and increase their accuracy and efficiency on various tasks [27].

Random Forest is an ensemble learning technique used to increase classification accuracy and mixes many decision trees. The mean (for regression) or mode (for classification) of the predictions made by each individual tree is the final prediction [28].

$$Q_{lef\ t}(\theta) = \{(x,\ y) \mid x \leq t_m\}$$
$$Q_{right}(\theta) = Q_m - Q_{lef\ t}(\theta) \qquad (1)$$

A decision tree is a model that resembles a tree and bases choices on feature values splitting data into subsets using conditions on these features. Formula: Decision trees use the Gini impurity or entropy to split nodes [29].

$$Gini = 1 - \sum_{i=1}^{c} (p_i)^2 \qquad (2)$$

Gaussian Naive Applying Bayes' theorem under the presumption of feature independence is the foundation of the Bayes probabilistic classifier [27]. The following formula is used to determine the likelihood of a class given features:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}) \qquad (3)$$

KNN uses the majority class of its k closest neighbors in the feature space to classify a data item [28]. Euclidean distance is frequently used to determine the separation between data points.

$$d(p,q) = \sqrt{\sum_{i=1}^{n} \ (p_i - q_i)2} \qquad (4)$$

A logistic function is used in logistic regression to represent the likelihood of a binary outcome [29]. The logistic function is calculated as follows:

$$z = (\sum_{i=1}^{n} w_i . x_i) + b \qquad (5)$$

3.5 Deep Learning

The term deep refers to a subtype of machine learning that uses multi-layered neural networks. By simulating the way the human brain processes information, these neural networks enable computers to learn and make judgments with little assistance from humans. Deep learning is especially good at complicated tasks like picture and audio recognition, natural language processing, and autonomous driving because it can identify patterns in large volumes of data. It is a potent artificial intelligence tool because of its capacity to learn from experience and data [30]. MLP is a feedforward artificial neural network with several neuronal layers of neurons with activation functions [31]. CNN are deep learning algorithms that work very well for visual data analysis. They use convolutional layers to extract features [32].

3.6 Ensemble Technique

In this research, the ensemble technique to improve classification accuracy beyond that of all individual algorithms. Predictions from multiple base models were combined such as K-Nearest Neighbors, Gaussian Naive Bayes, Random Forest, Decision Tree, Logistic Regression, Multi-layer Perceptron, and Convolutional Neural Networks. Taking advantage of these different models, the research developed a meta-model that combines their predictions in an optimal manner, resulting in a stronger penetration detection strategy over the method used by individual algorithms. This proved to greatly ameliorate overall performance while offering valuable insights into how to improve network security applications.

3.7 Evaluation Parameters

The performance evaluation of intrusion detection algorithms encompasses various metrics that assign numerical values to their efficacy. These metrics capture varied dimensions of the class accuracy, robustness, and proficiency of an algorithm in detecting various cyber threats. There are a few standard parameters used in the evaluation of intrusion detection [33].

- The accuracy determines how accurate the model is overall by dividing the number of accurately predicted occurrences by the total number of cases. Where FN stands for False Negatives, FP for False Positives, TN for True Negatives, and TP for True Positives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}*100 \qquad (6)$$

- Precision shows the percentage of positive prediction by showing the percentage of true positive predictions among all positive predictions.

$$Precision = \frac{TP}{TP+FP}*100 \qquad (7)$$

- Recall evaluate how well the model can distinguish genuine positive examples from all other positive examples.

$$\mathrm{Re}\,call = \frac{TP}{TP+FN}*100 \qquad (8)$$

- F1-score is the one statistic that balances false positives and false negatives is the harmonic mean of accuracy and recall.

$$F1-score = 2*\frac{\mathrm{Pr}\,ecision*\mathrm{Re}\,call}{\mathrm{Pr}\,ecision+\mathrm{Re}\,call} \qquad (9)$$

A confusion matrix shows the numbers of true positive, true negative, false positive, and false negative predictions in order to characterize how well a classification model is doing.

4. Results and Discussion

This section details the outcomes of the use of several machine learning and deep learning algorithms on our dataset for intrusion detection, namely Random Forest, Decision Tree, Gaussian Naive Bayes, K nearest neighbor, Logistic Regression, Multi-layer perceptron, and Convolutional Neural Network. The performance of each algorithm was evaluated using key performance indicators such as accuracy, precision, recall, and F1 score. To further optimize the classification accuracy, this research applied an ensemble technique using the predictions produced by these base models as inputs into a meta-model. By capitalizing on the diverse strengths of each algorithm, performance increased significantly, making for a more effective and robust intrusion detection system. The findings show how much ensemble learning can actually improve security measures.

4.1 Performance Evaluation of Random Forest

The random forest classifier worked powerfully for intrusion detection dataset. Maximum accuracy was 88.15% representing large numbers of correctly classified instances shown in Figure 4. A precision score of 87.64% indicates that a large percentage of instances classified as positive were correctly classified. The recall stands at 88.15%, representing the strength of the model in detecting actual positives with near-zero false negatives. F1 score is 87.89%, with precision and recall balanced, indicating overall reliability and robustness of the classifier.

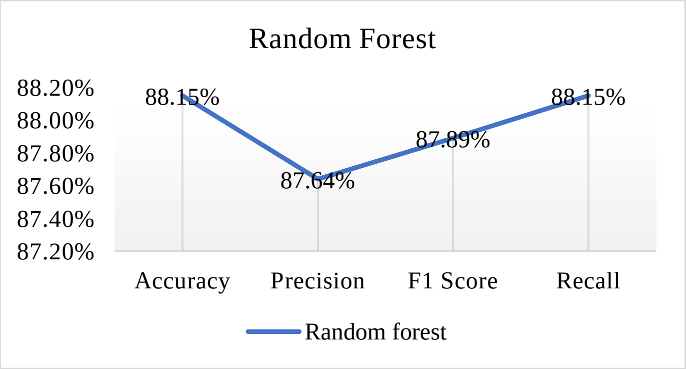


Figure 4. Random Forest performance measurement graph of different evaluation matrices

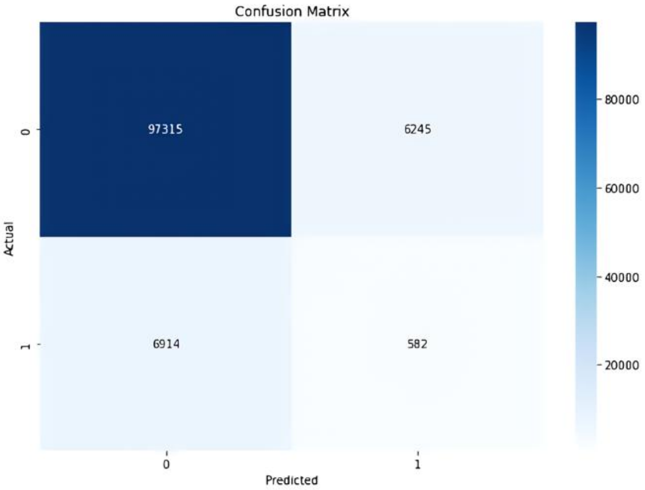


Figure 5. Confusion Matrix for Random Forest

4.2 Performance Evaluation of Decision Tree

The overall accuracy was 87.82%, indicating that a fair amount of cases were correctly classified shown in Figure 6. Positive score: The precision of the model is equal to 87.55%. The 87.82% vis-a-vis recall helps one to ascertain the cases of actual positives. Similarly, F1 scores showing 87.69% mean that the classifier was robust, thus maintaining a decent balance between accuracy and recall.

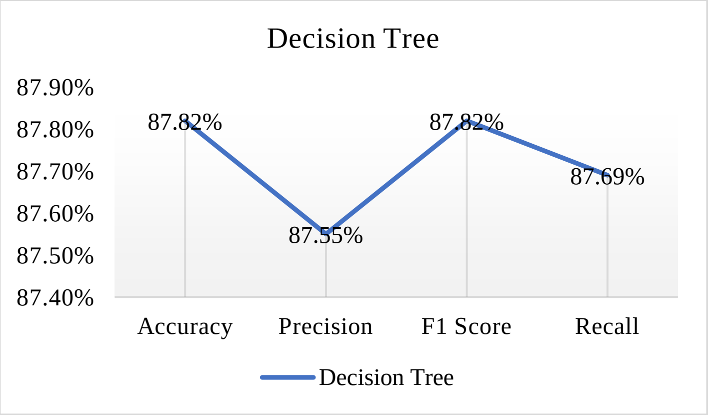


Figure 6. Performance Evaluation of Decision Tree

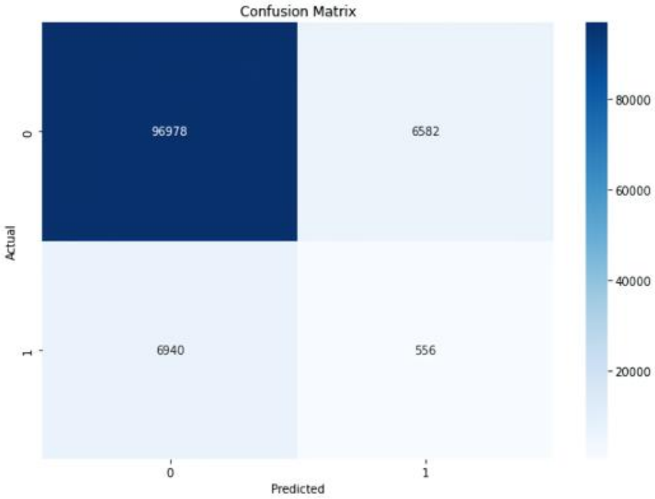


Figure 7. Confusion Matrix for Decision Tree

4.3 Performance Evaluation of Logistic Regression

The logistic regression model performed exceptionally well on our intrusion detection dataset, achieving an accuracy of 93.2%, thus establishing that a lot of instances were classified correctly shown in Figure 8. The precision of 86.96% denotes the correctness of predictions from the model, while the recall of 93.25% highlights how well the model is able to identify actual positives. The F1 score of 89.99% denotes a balance between precision and recall and confirms the overall strength and reliability of the model.

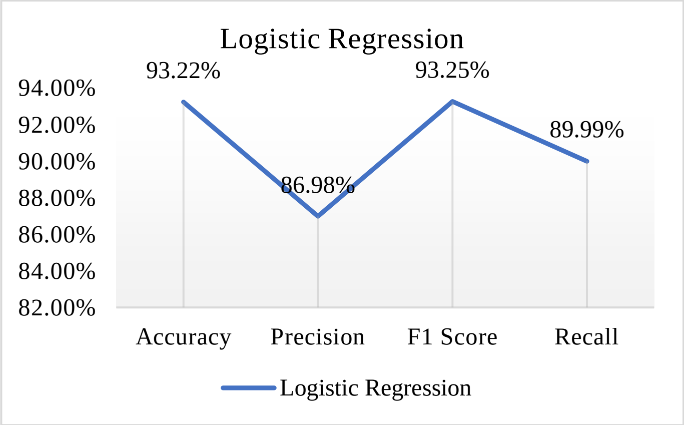


Figure 8. Performance Evaluation of Logic Regression

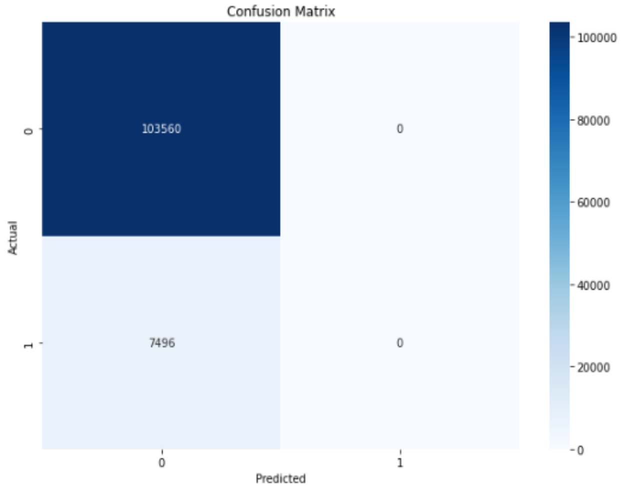


Figure 9. Confusion Matrix for Logic Regression

4.4 Performance Evaluation of Gaussian Naive Bayes

Compared to the other models, the Gaussian Naive Bayes model performed well on our intrusion detection dataset, reporting an accuracy of 72.21%, a precision of 94.38%, a recall of 72.21%, and an F1 score of 79.13% shown in Figure 10. These findings mean that while the model predicts positive outcomes very well in terms of accuracy, it detects actual positive outcomes only moderately well because of comparatively lower recall. The balanced F1 score points toward a possibility for enhancement of the detection capability by suggesting the trade-off between accuracy and recall. Understanding the proposed measures helps get to know the contribution offered by the Gaussian Naive Bayes model to the intrusion detection system while seeking to emphasize that the Gaussian Naive Bayes output like with other models might be combined via ensemble approaches to boost performance.

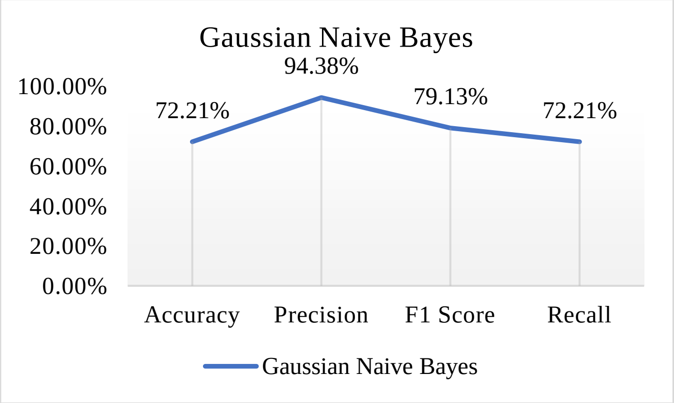


Figure 10. Performance Evaluation of Gaussian Naïve Bayes

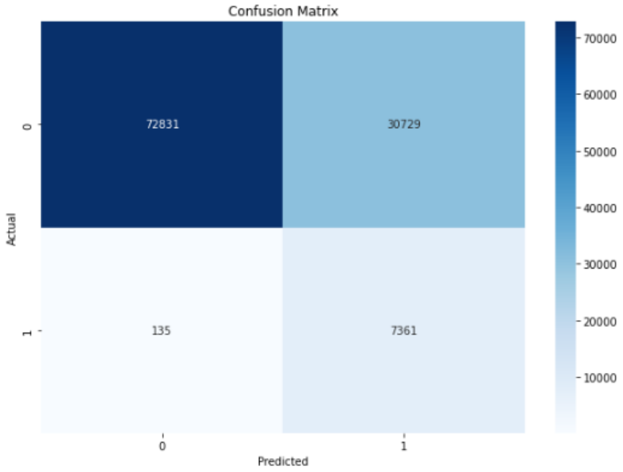


Figure 11. Confusion Matrix for Gaussian Naive Bayes

4.5 Performance Evaluation of Gradient Boosting

To the dataset used for intrusion detection, the Gradient Boosting Classifier performed satisfactorily, scoring not less than 93.35% accuracy, 91.14% precision, 93.35% recall, and an F1 score of 90.97% shown in Figure 12. These findings indicate that the model performed quite well overall, with a high accuracy meaning that most cases were predicted correctly. The precision value indicates how many of the total predicted positive cases were actual affirmative cases. The recall value states the model is good in detecting true positives, whereas the accuracy value proposes how accurately the model has predicted positive outcomes. The balance of the F1 score suggests that the model is maintenance-free and trustworthy with its coverage of classification tasks.

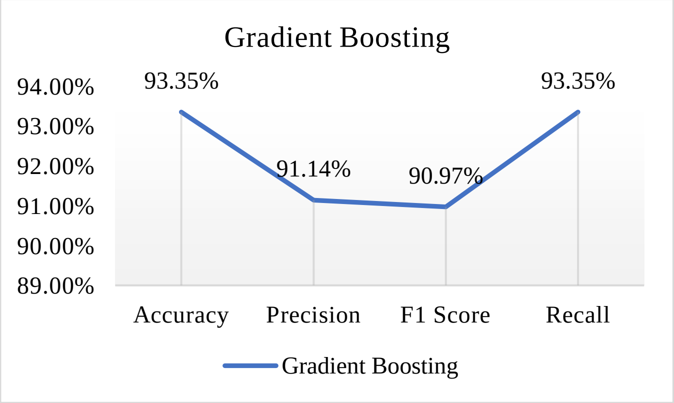


Figure 12. Performance Evaluation of Gradient Boosting

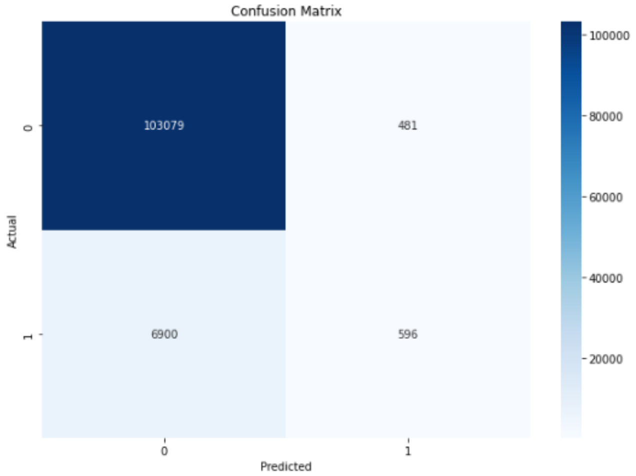


Figure 13. Confusion Matrix for Gradient Boosting Classifier

4.6 Performance Evaluation of K-Nearest Neighbors

K-Nearest Neighbors classifier demonstrated superior behavior while handling our intrusion detection dataset. The KNN classifier has an accuracy rate of 92.32%, indicating that the majority of the cases were correctly labeled shown in Figure 14. The accuracy of the model correctly forecasting the positive class is represented as 89.04%, while the recall at 92.32% is for true positive events. The strength and reliability of the model were highlighted by an F1 score of 90.28%, emphasizing its balanced performance with an equal emphasis on precision and recall. These results reflected the key contributions of the KNN classifier to our intrusion detection system and hold promise for further enhancement of overall system performance and robustness through ensemble approaches by integrating it with more models.

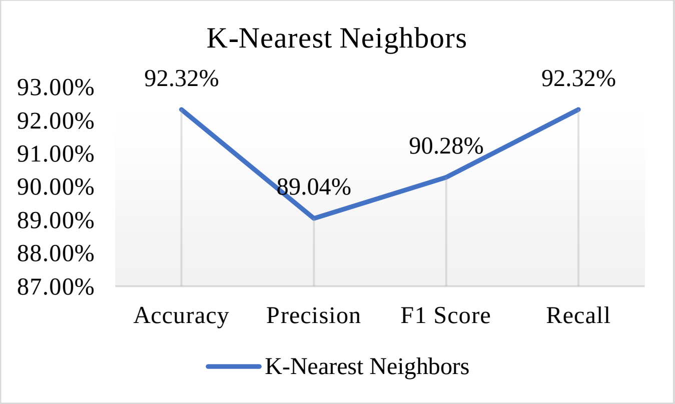


Figure 14. Performance Evaluation of K- Nearest Neighbors

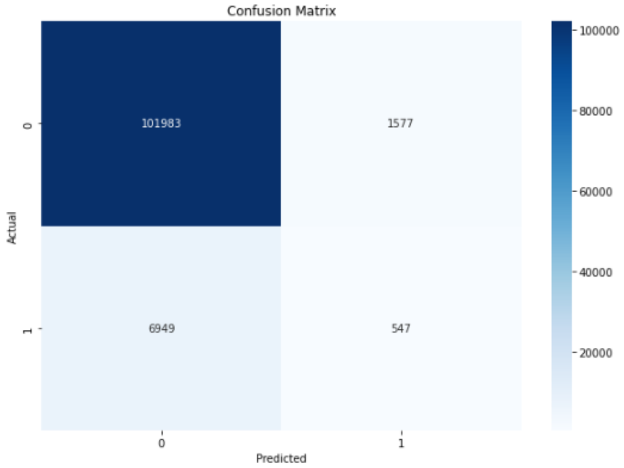


Figure 15. Confusion Matrix for K-Nearest Neighbors

4.7 Performance Evaluation of Multi-Layer Perceptron

The multi-layer perceptron classifier demonstrates robust performance on our intrusion detection dataset. This classifier correctly classified an astonishing 93.21% of instances, suggesting that a high proportion of instances was classified with accuracy shown in Figure 16. The precision of 89.57% gives an idea as to how well the model will make correct positive classifications while the recall of 93.21% is a measure of how well it identifies real positive instances. The F1 score of 90.09% thus makes a performance balance between precision and recall apparent, showing the model's reliability and robustness overall. This would explain the performance of this model in supporting our intrusion detection system, where it could be extended using different models through ensemble techniques for more effective and robust performance.

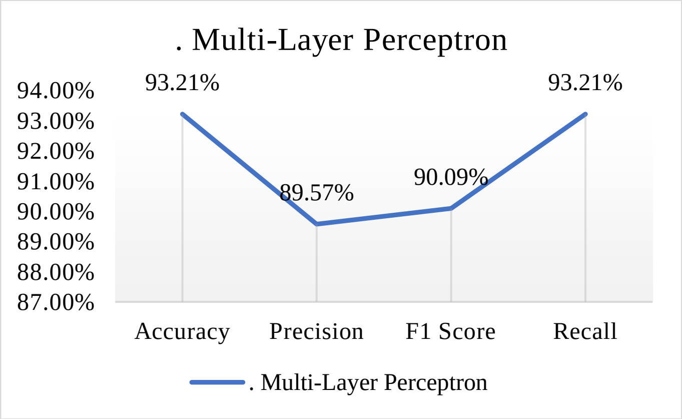


Figure 16. Performance Evaluation of Multi-Layer Perceptron

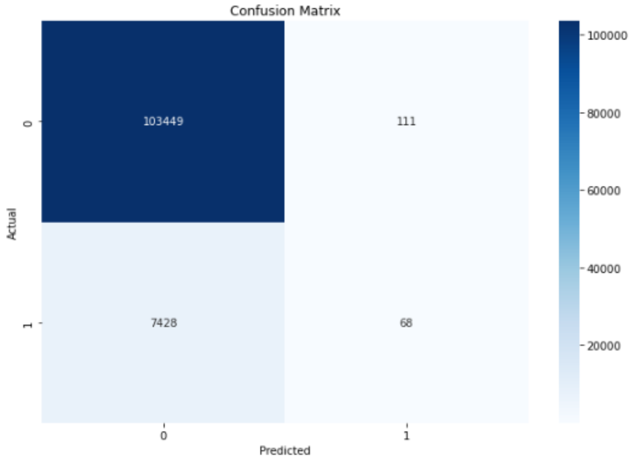


Figure 17. Confusion Matrix for Multi-Layer Perceptron

4.8 Performance Evaluation of Convolutional Neural Network

The Convolutional Neural Network (CNN), however, performed remarkably well on our intrusion detection dataset. With 93.14% accuracy, the models have classified a large number of instances correctly shown in Figure 18. Accuracy of 86.95% indicates that the model performs well in predicting the numbers of positive classes among all the instances on the training data, while a 93.14% recall specificity indicates how well this model shows positive events. The F1 score of 89.94% represents a compromise between recall and accuracy, further demonstrating how well-rounded and stable the model could be. These results underscore the contribution of the CNN toward our intrusion detection system and its potential for incorporation with other models in ensemble techniques for boosting performance and robustness.

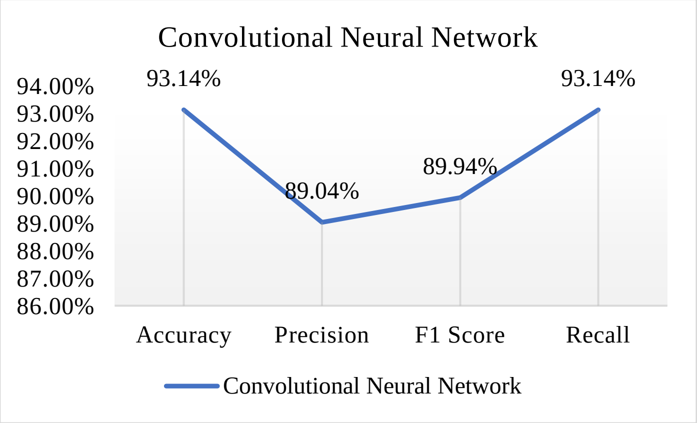


Figure 18. Performance evaluation of Convolutional Neural Network

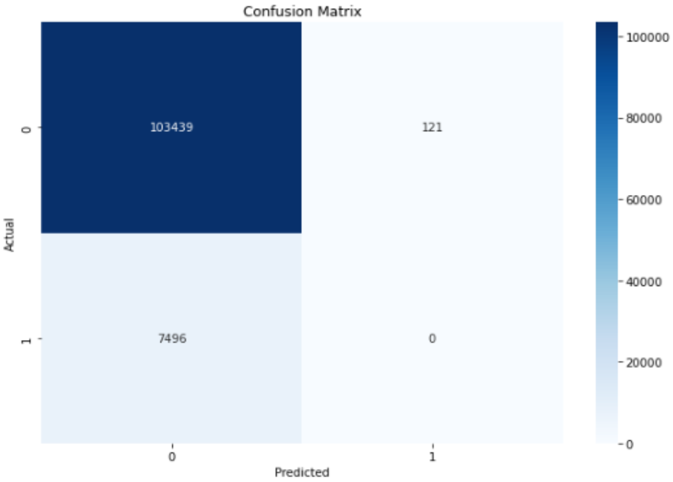


Figure 19. Confusion Matrix for Convolutional Neural Network

4.9 Performance of Proposed Ensemble Models

Ensemble or ensemble learning the benefits of many base models of Random Forest, Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbor, Logistic Regression, Multi-Layer Perceptron, and Convolutional Neural Networks performed well with our dataset. The accuracy, precision, and recall of the stacked model were 96.33%, 95.53%, and 95.64% respectively shown in Figure 20. This shows how much the algorithm can detect real positive cases and give highly accurate positive predictions. Ensemble's remarkable accomplishment highlights its promise to strengthen the intrusion detection system in terms of power and resistance via a combination of the many advantages of multitudes of algorithms, thereby improving an even more significant and accurate detection mechanism.

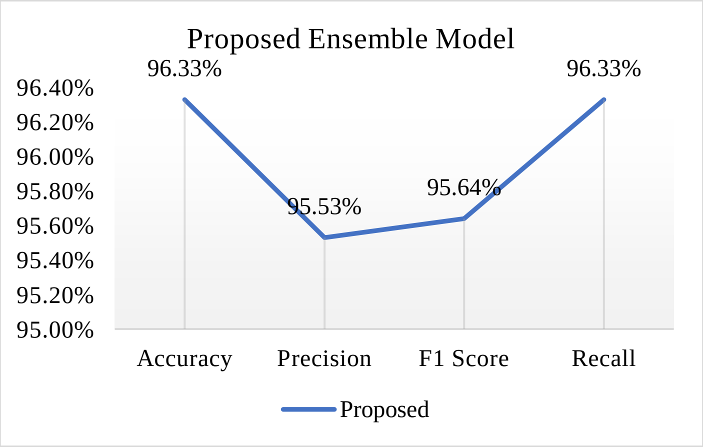


Figure 20. Performance of Proposed Ensemble Model

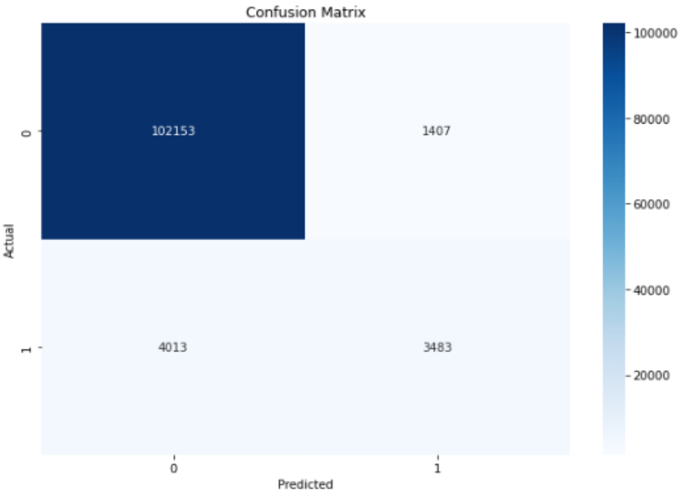


Figure 21. Confusion Matrix for Performance of Proposed Ensemble Model

4.10 Performance Comparison of Proposed LDX Model

The comparison of the performance of various models used in this study, including the proposed ensemble model, Gradient Boosting, Logistic Regression, K-Nearest Neighbors, Multi-Layer Perceptron, Convolutional Neural Network, Random Forest, Decision Tree, and Gaussian Naive Bayes. The evaluation metrics used for comparison are accuracy, recall, F1-score, and precision.

Table I sets a comparative study of various machine learning algorithms using four performance metrics Accuracy, Precision, F1-Score, and Recall. The proposed LDX Model demonstrates its superiority over all machines by being the only algorithm to rank first in accuracy 96.33% with balanced measures of precision 95.53% recall 96.33%, and F1-score 95.64%, therefore, having the most predictive power. Among classical models, Gradient Boosting comes in quite efficient at 93.35%, shortly followed by Logistic Regression, MLP, and CNN in the range of 93%. KNN shows a competitive performance of 92.32% but with slightly lesser precision. The accuracy of Random Forest and Decision Tree remains mediocre with rates of 88.15% and 87.82%, suggesting possible overfitting. Interestingly Gaussian Naïve Bayes has the lowest accuracy with 72.21% but has the highest precision with 94.38%, which implies that it predicts a small number of false positives but has poor recall and overall performance. Results evidenced the comparative advantage of ensemble and deep learning models over simpler algorithms, thereby conferring more robustness to the proposed LDX model in attaining the optimal classification performance.

Table 1. Performance Comparison Proposed Model's

Models	Accuracy	Precision	F1-Score	Recall
Proposed LDX Model	96.33%	95.53%	95.64%	96.33%
Gradient Boosting	93.35%	91.14%	90.97%	93.35%
Logistic Regression	93.22%	86.98%	93.25%	89.99%
KNN	92.32%	89.04%	90.28%	92.32%
MLP	93.21%	89.57%	90.09%	93.21%
CNN	93.14%	89.04%	89.94%	93.14%
Random forest	88.15%	87.64%	87.89%	88.15%
Decision Tree	87.82%	87.55%	87.82%	87.69%
Gaussian Naive Bayes	72.21%	94.38%	79.13%	72.21%

5. Conclusions and Future Work

The results highlight that the efficiency of any intrusion detection system can be improved optimally using the ensemble models. With the strengths of various machine learning and deep learning algorithms combined, ensemble models perform well in terms of better accuracy and robustness in detecting cyber threats. All the algorithms were trained and tested on our intrusion detection dataset, and evaluation metrics such as accuracy, precision, recall, and F1 score were evaluated. Our results suggest that ensemble technique whereby the advantages of different individual algorithms are leveraged together, delivers the highest performance. The ensemble model attained a great accuracy of 96.33%, precision of 95.53%, recall of 96.33%, and F1 score of 95.64%. This signifies the leverage of the ensemble approach in significantly improving detection capabilities. Future work will focus on: integrating BERT and GPT models for encrypted traffic analysis, measuring framework adaptability through response time and false-positive rate benchmarks, and testing against zero-day attacks in AWS/Azure cloud environments.

References

1. Alotaibi, Y. and M. Ilyas, Ensemble-learning framework for intrusion detection to enhance internet of things' devices security. Sensors, 2023. 23(12): p. 5568.
2. Srinivasan, S. and P. Deepalakshmi, Enhancing the security in cyber-world by detecting the botnets using ensemble classification based machine learning. Measurement: Sensors, 2023. 25: p. 100624.
3. Mokbal, F.M.M., et al., An efficient intrusion detection framework based on embedding feature selection and ensemble learning technique. Int. Arab J. Inf. Technol., 2022. 19(2): p. 237-248.
4. Mishra, A.K. and S. Paliwal, Mitigating cyber threats through integration of feature selection and stacking ensemble learning: the LGBM and random forest intrusion detection perspective. Cluster Computing, 2023. 26(4): p. 2339-2350.
5. Lower, N. and F. Zhan. A study of ensemble methods for cyber security. in 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). 2020. pp. (1001-1009).IEEE.
6. Hossain, M.A. and M.S. Islam, Ensuring network security with a robust intrusion detection system using ensemble-based machine learning. Array, 2023. 19: p. 100306.
7. Selvalakshmi, B., et al., Enhancing E-Commerce Data Privacy in India's Rapidly Evolving Cybersecurity Landscape Through AI-Driven Intrusion Detection Systems, in Strategic Innovations of AI and ML for E-Commerce Data Security. 2025, IGI Global. p. 261-280.
8. Jiang, Y. and Y. Atif, A selective ensemble model for cognitive cybersecurity analysis. Journal of Network and Computer Applications, 2021. 193: p. 103210.
9. Akram, U., et al., IoTTPS: Ensemble RKSVM Model-Based Internet of Things Threat Protection System. Sensors, 2023. 23(14): p. 6379.
10. Nadeem, M.W., et al., Fusion-Based Machine Learning Architecture for Heart Disease Prediction. Computers, Materials & Continua, 2021. 67(2).
11. F. Rustam, M. F. Mushtaq, A. Hamza, M. S. Farooq, A. D. Jurcut, and I. Ashraf. Denial of service attack classification using machine learning with multi-features. Electronics.. 11 (22), p. 3817.
12. R. Majeed, N. A. Abdullah, M. Umer, and M. Nappi. Intelligent cyber-security system for IoT-aided drones using voting classifier. Electronics. 10 (23), p. 2926.
13. Akram, Urooj, et al. "A comprehensive survey on Pi-Sigma neural network for time series prediction." Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 9.3-3 (2017): 57-62.
14. R. Fernandes, J. Silva, O´. Ribeiro, I. Portela, and N. Lopes. The Impact of I. identifiable Features in ML Classification Algorithms with the HIKARI-2021 Dataset. IEEE.
15. Sanober, I. and R.N. Mir, Feature Reduction and Dataset Balancing in Intrusion Detection Systems: A Comprehensive Evaluation on Hikari-2021 and Legacy Datasets. Available at SSRN 5009537.
16. T. Ludovico. "Improving Intrusion Detection Systems: Challenges with Public Datasets and the Role of Explainable AI: A Practical Guide Using NFS-2023-TE and HIKARI-2021." 2024.
17. Zhou, Y. and P. Wang, An ensemble learning approach for XSS attack detection with domain knowledge and threat intelligence. Computers & Security, 2019. 82: p. 261-269.
18. M. Rizwan et al., "Depression classification from tweets using small deep transfer learning language models," IEEE Access, vol. 10, pp. 129176–129189, 2022.
19. A. Ferriyan, A.H. Thamrin, K. Takeda, and J. Murai. "Generating Network Intrusion Detection Dataset Based on Real and Encrypted Synthetic Attack Traffic." Applied Sciences, 2021.
20. Karim, A., et al., Anticipating impression using textual sentiment based on ensemble LRD model. Expert Systems with Applications, 2025. 263: p. 125717.
21. Geng, G., et al., Random forest model that incorporates solar-induced chlorophyll fluorescence data can accurately track crop yield variations under drought conditions. Ecological Informatics, 2025. 85: p. 102972.
22. Hu, L., et al., Significance-based decision tree for interpretable categorical data clustering. Information Sciences, 2025. 690: p. 121588.
23. M. F. Mushtaq et al., An innovative cognitive architecture for humanoid robot. International Journal of Advanced Computer Science and Applications, 2017. 8(8).
24. Kwon, D., et al., Evaluating unbalanced network data for attack detection, in Proceedings of the 2023 on Systems and Network Telemetry and Analytics. 2023. p. 23-26.
25. Swati Chaudhari, Pratyush Shukla, and Archana Thakur. "A Comprehensive Investigation on the Identification of Real and Encrypted Synthetic Network Attacks using Machine Learning Algorithms." International Journal of P2P Network Trends and Technology, vol. 14, no. 1, pp. 1-6, 2024.
26. J. Vitorino, M. Silva, E. Maia, and I. Praça. "An Adversarial Robustness Benchmark for Enterprise Network Intrusion Detection." International Symposium on Foundations and Practice of Security, 2023.
27. U. Akram et al., "An improved pi-sigma neural network with error feedback for physical time series prediction" Int. J. Adv. Trends Comput. Sci. Eng, vol. 8, pp. 1-7, 2019.
28. Ferriyan, A., et al., Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. applied sciences, 2021. 11(17): p. 7868.

29. Mushtaq, M. F., et al., A Survey on the Cryptographic Encryption Algorithms. International Journal of Advanced Computer Science and Applications, 2017. 8(11): pp. 333-344.
30. M. F. Mushtaq et al., “Key schedule algorithm using 3-dimensional hybrid cubes for block cipher,” International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, 2019.
31. Sabzekar, S., et al., The Impact of Network Indices Integration on Traffic Flow Imputation Accuracy: A Machine Learning Approach. IEEE Transactions on Intelligent Transportation Systems, 2025.
32. A. Sarwar et al., “IoT networks attacks detection using multi-novel features and extra tree random-voting ensemble classifier (ER-VEC),” Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 12, 2023.
33. Lazzarini, R., H. Tianfield, and V. Charissis, A stacking ensemble of deep learning models for IoT intrusion detection. Knowledge-Based Systems, 2023. 279: p. 110941.