ISSN - 2456-1142

International journal of humanities and information technology Volume 7, Issue 1, 2025

An Enhanced Intrusion Detection Architecture Using ML-Based Anomaly Detection on Large-Scale Data

Krishna Bhardwaj Mylavarapu[1], Jenitha Pilli[2], Prathik Kumar Jannu[3], Javed Ali Mohammad[4], Sri Harsha Panchali[5], Usha Mohani Kavirayani[6]

[1]MS in Computer Science, University of Illinois Springfield

[2]MS in Computer Science, University of Louisiana at Lafayette

[3]Computer Science Engineering, JNTU Hyderabad

[4]Masters in Data Science, New England College

[5]Information Systems Engineer, CrowdStrike Inc

[6]MS in Computer Science, Kent State University

Abstract

Intrusion detection (ID) has become a major issue in the area of safe and dependable communication infrastructure due to the faster pace of large-scale networked systems and the mounting sophistication of cyber-attacks. This article discusses an ID architecture for very big network data that is based on anomaly detection influenced by machine learning (ML). The proposed method makes use of the Synthetic Minority Oversampling Technique (SMOTE), data processing, data cleaning, data normalization, label encoding, and class balancing in order to enhance data quality and solve the issue of class imbalance. A Random Forest (RF) classifier serves as the primary detection model due to its strength, capacity for ensemble learning, and ability to handle high-dimensional data. Another tool used to identify intrusions when users and servers exchange sensitive data is an intrusion detection system (IDS). RF model would be the most effective among other ML and DL models due to its accuracy, which implies that this model could capture unusual and dangerous network activity. The experimental findings prove that the proposed architecture is an effective, scalable, and reliable method of detecting anomalies in large-scale network systems.

Keywords: Anomaly detection, Intrusion Detection System (IDS), Cybersecurity, NSL-KDD, Network Traffic Analysis, Ensemble Learning.

International journal of humanities and information technology (2025)

Introduction

In the midst of the quick advancement of internet technology and smart gadgets, the quantity of infiltrations. The act of entering a location or virtual area without authorization is known as intrusion. Confidentiality and system security greatly depend on ID [1]. Another tool used to identify intrusions when users and servers exchange sensitive data is an IDS [2]. Thus, detecting new threats is one of the most important factors to improve the systems' security. In addition to preventing access to unwanted data, the firewalls also hide valuable system elements to the curious eyes. Hackers can use ports such as SMTP, HTTP among others, which are often left open by systems to send malicious traffic. Thus, it becomes necessary to have advanced IDS. It is simpler to categorize known threats when IDSs are divided into signature-based and misuse-based detection techniques. It searches the network traffic [3][4] subsequent to malicious transmissions. Anomaly-based IDSs, which include user behavior from the past into their learning process, are the other type of IDS. IDSs that are anomaly-based help find new threats. Depending on the data it analyzes, Additionally, the IDS may be separated into groups based on hosts and networks.

In the context of ML, being able to identify abnormalities in IoT networks depends on choosing the right feature sets [5][6][7]. These feature sets are necessary for ML model training in order for them to correctly classify and identify anomalous network activity [8][9]. Because IoT devices generate enormous amounts of data, feature sets must be carefully selected in order for the ML-based detection method to be successful. Numerous methods have been investigated in earlier studies; difficulties associated with computing complexity and insufficient feature sets lead to creative solutions.

Significance and Contribution of Paper

The relevance of this work is that it has focused on the increasing requirement of proper and scalable IDS in large and complicated network settings. Since cyber-attacks

are constantly changing their levels and sophistication, conventional security strategies are not able to accurately detect anomalous behaviors. This study shows how ensemble learning is safe against class imbalance and, therefore, noisy data, and can easily enhance detection accuracy by using an ML-based anomaly detection solution that employs a RF classifier. The suggested solution to actual network security systems is proven to be useful by its excellent performance in accurately detecting malicious activity, reducing false alarms, and improving network protection overall, on the NSL-KDD dataset. The following summarizes this paper's primary contributions:

- Obtained NSL-KDD data set on Kaggle which offers a practical base on anomaly detection.
- Used a large amount of data pre-processing such as cleaning, normalization, label encoding, and SMOTE in order to balance the distribution of classes and enhance the quality of the data.
- Used a RF classifier on anomaly detection that creates a dependable and accurate IDS by utilizing its ensemble learning capacity.
- To evaluate the model's efficacy, a comprehensive performance assessment was developed using the F1-score metrics, precision, recall, and accuracy.

Novelty & Justification of the Study

This research has been justified by the fact that there has been a growing need in the need to have reliable and efficient IDS that have the capacity to handle massively distorted network traffic data that is not always easily identified by conventional security measures. Although many machine and deep learning algorithms have been proposed, many of them are computationally complex, overfitting, or do not generalize to real-world conditions. This work is novel in the sense that a rigorous data preprocessing process incorporating SMOTE-based class balancing is systematically coupled with a well-optimized framework based in terms of low false alarm rates and detection accuracy, the RF with an early detection detector is more effective. Contrary to sophisticated deep learning models, the suggested solution shows that ensemble-based ML model has the ability to efficiently generalize intricate intrusion patterns, with increased interpretability and stability, as well as scalability. The high level of empirical validity and comparative outperformance of the proposed ID architecture against existing ones are indicative of the novelty and pragmatic usefulness of the architecture.

Structure of Paper

Here is the outline of the paper: The second section summarizes the studies conducted on ID. In Section III, we detailed the methodology, dataset, and procedures that will be used to implement the model. The experimental results are presented in section IV. Section V concludes with a discussion of the results, highlighting their limits and suggesting areas for future study.

Literature Review

Recent studies on ML-Based Anomaly Detection are thoroughly discussed in the literature review. Table I provides a summary of the results, which include the methodology utilized, performance outcomes, important discoveries, limits, and suggestions for further research.

Chauhan and Vamsi (2019) To detect the anomalous ozone measurements in air quality data, different ML methods were used. In the comparative study of testing unsupervised ML methods, Isolation Forests was better than One Class Support Vector machine. Also, these anticipated anomalies were examined with the help of Z-Score to identify a failure sensor. Isolation Forest was the best in classifying abnormalities in Ozone values based on Air Quality sensor data with an accuracy of 92.7% compared to One Class Support Vector Machine [10].

Shriram and Sivasankar (2019) assessed the various unsupervised anomaly detection methods using performance metrics like recall, accuracy, area under the curve, and F-score. This work incorporates EE, IF, LOF, and OneClassSVM as unsupervised learning techniques. The testing was conducted using the satellite and shuttle datasets. The effectiveness of several unsupervised learning strategies was contrasted with that of supervised learning strategies such as SVM and k-NN. According to the records of the shuttle and the satellite, the findings are that unsupervised learning techniques to identify anomalies are similar or better than supervised learning techniques [11].

Zaman and Lung (2018) The current trend of anomaly identification is based on ML methods of categorization. Test the seven different ML methods with the calculation of information entropy using the Kyoto 2006+ dataset, and evaluate the effectiveness of each of the methods. The findings have shown that the vast majority of the ML techniques are over 90 percent accurate, recalls, and precise to this particular data set. Nevertheless, the Radial Basis Function (RBF) is the most effective algorithm of the seven algorithms analyzed in this paper based on the Receiver Operating Curve (ROC) measure's applicability [12].

Aljamal et al. (2019) To employ anomalies to identify unknown threats, a hybrid detection mechanism that incorporates it is necessary to explore the benefits of signature-based detection techniques. To improve the accuracy of the system, the paper suggests implementing an anomaly detection system at the level of the Cloud Hypervisor that combines SVM classification with K-means clustering. they assess the suggested approach by comparing the findings to those of prior research and analyzing data from the UNSW-NB15 project. The suggested K-means clustering algorithm outperforms the alternatives when evaluating accuracy [13].

Dilraj, Nimmy and Sankaran (2019) There is need to come up with an innovative approach whereby such gadgets power consumption is utilized. To do it, a smart home situation is replicated with the help of brute-force and smart cameras,

and DDoS attacks have been performed to gather power profile changes. Develop ML models to identify irregularities in the history of power consumption. The suggested method has a 94.04 detection rate. According to study, power consumption can serve as a possible measure to determine an imbalance in the IoT-based smart home [14].

Vartouni, Kashi and Teshnehlab (2018) proposed an isolated forest as a classifier and a DNN as a feature learning method. also used the CSIC 2010 data set to compare the method with those that do not use feature extraction models. The results indicate that deep models are more precise than the techniques that do not extract features [15].

Luo and Zhong (2017) a method of achieving enhanced accuracy in engine gas route anomaly detection by enriching with deep learning anomaly detection. To be able to train robust features on labelless datasets, the first model constructed was a stacked denoising autoencoder model. In order to find abnormalities, learning characteristics were then fed into a Gaussian distribution-based anomaly detection system. To confirm that the proposed method is technically feasible, an experiment was carried out to look at real rapid data from a certain kind of turbofan gas turbine engine's recorder. When compared to conventional approaches, the results showed that this approach might increase the accuracy of anomaly identification [16].

Mehmood and Md Rais (2016) The anomaly-based detection method has the large FPR, it is able to detect new attacks. Several ML algorithms are developed in order to tackle this problem. In this work, the different supervised anomaly-based detection algorithms are compared. The algorithms have been tested using the KDD99 data, which is the reference data of anomaly-based detector models. The outcome demonstrates that none of the approaches obtain high detection rates across all KDD99 dataset types [17].

Research Methodology

The proposed approach to NSL-KDD dataset is where ID starts, which is first treated using an intrusion detector to improve the data's quality. To remove the problem of an unbalanced dataset, this step consists of data cleaning, addressing missing or null values, feature normalization and label encoding, as well as balancing the SMOTE feature in the courses. After processing, the data is split into training and testing sets in an 80:20 ratio. Building an RF classifier using a training set enables the establishment of infiltration patterns. Finally, standard measures such as F1-score, recall, accuracy, and precision are employed to assess the model's efficacy on the test dataset. The approach taken during the investigation is illustrated in Figure 1.

The phases are explained in depth and step-by-step below, shown in the flowchart:

Data Collection

To be more precise, NSL-KDD use an enhanced KDDCup99 dataset. Researchers from the Canadian Institute for Cybersecurity at the University of New Brunswick have reduced the size of the original KDDCup99 dataset to make it more concise and usable.

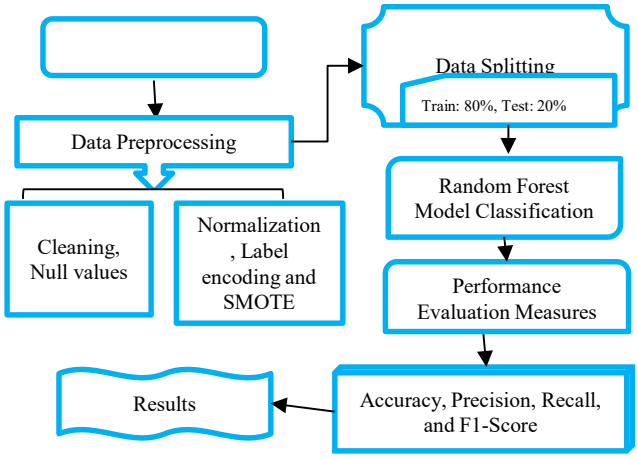

Figure 1: Flowchart for Intrusion Detection

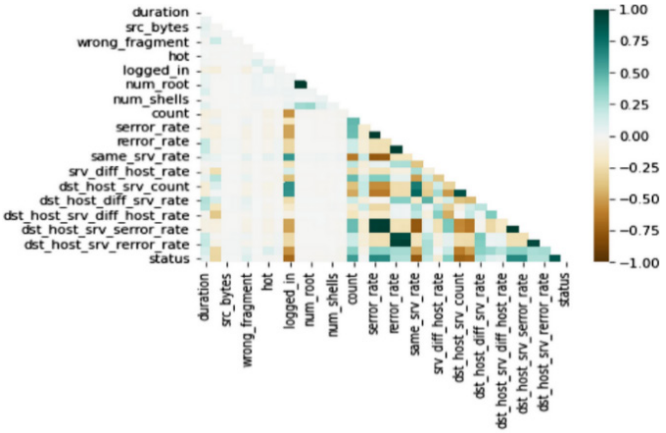

Figure 2: Heatmap for NSL-KDD Dataset

Data Analysis and Visualization

The practice of visually representing information and data is known as data visualization. Tools for data visualization that use visual aids like graphs, charts, and maps, provide a simple way to observe and comprehend data trends and patterns. The NSL-KDD data visualization is shown below:

Figure 2 displays a correlation heatmap that displays the relationships between different features in the NSL-KDD dataset. The intensity of the association may be shown by color-coding the correlation values, which range from 1 (very negative correlation) to 1 (highly positive correlation).

You can see the breakdown of the NSL-KDD dataset into attack and normal classes in Figure 3, which is a pie chart. Normal traffic accounts for 51% of the data, while DoS assaults account for 35%. Echoing the dataset's emphasis on many forms of intrusion, other categories are Probe (9%), Root-to-Local (RTL) (3%), and User-to-Root (UTR) (2%).

Figure 4 depicts the NSL-KDD dataset's class distribution, which represents the normal and attack instances utilized

Table 1: Summary of literature Overview and Comparative Analysis of Anomaly Detection Techniques

Author(s) & Year	*Technique(s) Used*	*Dataset*	*Key Findings*	*Future Work / Research Gaps*
Chauhan and Vamsi (2019)	One-Class SVM, Isolation Forest, Z-Score validation	Air Quality Ozone Sensor Data	Distancing oneself When it came to identifying unusual ozone readings, Forest fared better than One-Class SVM (92.7% accuracy), while Z-score was useful for confirming sensor errors.	Future work may include real-time deployment, adaptive thresholding, and testing on multi-pollutant air quality datasets with concept drift.
Shriram and Sivasankar (2019)	One-Class SVM, LOF, Isolation Forest, Elliptic Envelope; compared with SVM, k-NN	Shuttle and Satellite datasets	In all four metrics—precision, recall, F-score, and AUC—unsupervised anomaly detection approaches were on par with or even outperformed supervised methods.	Further studies can explore scalability on high-dimensional IoT data and hybrid semi-supervised approaches.
Zaman and Lung (2018)	Seven ML classifiers with entropy-based feature selection (including RBF)	Kyoto 2006+ Network Traffic Dataset	Most ML techniques achieved >90% precision, recall, and accuracy; RBF performed best based on ROC-AUC.	Future research can focus on online learning models and robustness against evolving cyberattack patterns.
Aljamal et al. (2019)	Hybrid K-means Clustering + SVM	UNSW-NB15 Network Intrusion Dataset	Hybrid anomaly detection model achieved slightly higher accuracy compared to existing approaches.	Future work may integrate deep learning models and evaluate performance in real cloud hypervisor environments.
Dilraj, Nimmy and Sankaran (2019)	Machine Learning models using power consumption features	IoT-based Smart Home (simulated attacks: DDoS, brute force)	Power consumption proved to be an effective anomaly indicator with 94.04% detection accuracy.	Further work could consider real-world smart home deployments and multi-modal features (network + power).
Vartouni, Kashi and Teshnehlab (2018)	Isolation Forest + Deep Neural Network for Feature Learning	CSIC 2010 Dataset	Deep feature extraction significantly improved anomaly detection accuracy compared to non-feature-based methods.	Future studies may explore lightweight deep models and transfer learning for resource-constrained environments.
Luo and Zhong (2017)	Stacked Denoising Autoencoders + Gaussian-based anomaly detection	Aircraft Engine Gas Path (QAR data)	Learned deep features enhanced anomaly detection accuracy over traditional methods.	Future work can include real-time fault prognosis and integration with predictive maintenance frameworks.

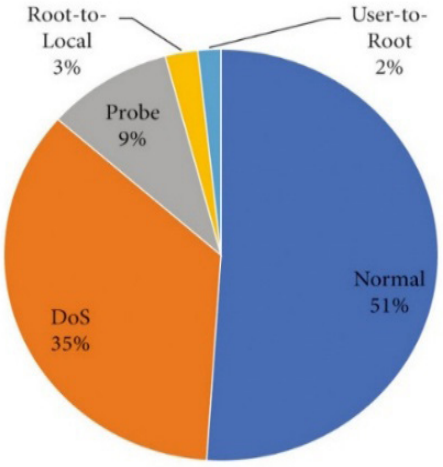

Figure 3: Pie Chart of NSL-KDD Dataset Distribution

for ID. The dataset has a slightly higher proportion of normal samples than attack samples, suggesting a slight class imbalance. Nonetheless, both classes are adequately represented, this indicates that there is sufficient data for efficient model training and assessment.

Data Preparation

Data pre-processing is an important stage for ML techniques because it the cleaning data and dealing with null values, categorical values, redundant, and irrelevant data. So, to build a model with high performance and good accuracy, the pre-processing must be accurate.

- **Cleaning the data:** It is necessary to eliminate all of the missing (nan) and values of infinity found in the original data.
- **Dealing with null values:** Strategies for keeping the dataset consistent include deleting rows with empty values or filling in missing data with the median, mode, or mean.

Normalization

The attribute values are normalized using Z-score normalization [18]. The attribute value is normalized so that after normalization, the standard deviation is one, while the mean is zero. This normalizing attribute is sometimes referred to as zero mean normalization for the Z-score. Its mathematical Equation is as follows: (1).

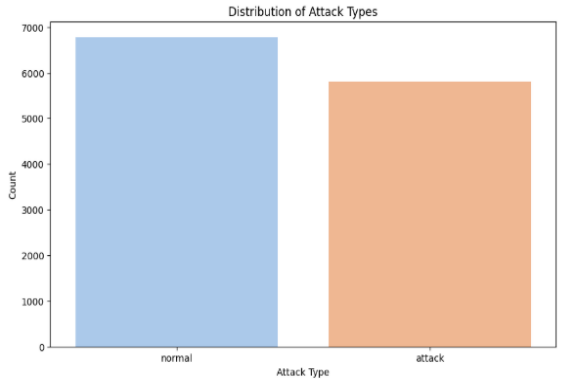

Figure 4: Distribution of NSL-KDD Dataset

$$a(i) = \frac{a(i) - mean(A)}{std(A)} \quad (1)$$

Here, A is the attribute, and $a(i)$ is i th value that the equation above update.

Label Encoding

Label encoding is the process of numerically altering category data for use by machine learning algorithms. For machine learning model training and to make model creation easier, numerical representations of categorical data are required.

Handling class imbalance using Synthetic Minority Oversampling Technique (SMOTE)

The SMOTE creates fictitious samples for the synthetic by filling in the current data in order to correct class imbalance, has been used to training data to produce a balanced dataset.

Data Splitting

The dataset is split 80:20, such that 80% used to construct the model, with the remaining 20% being utilized for testing to assess the model's performance using previously unseen data.

Model Classification

In order for RFs or Random Decision Trees to work, it is necessary to build and regularly train a large number of "mini" decision trees or estimators on a particular dataset[19]. In essence, the RF uses the majority vote of these estimators to assess output. The mathematical function represents the estimators' average (or majority vote) for a specific input sample x 0. This is how classification is done more formally. Equation (2) represents the precise formula:

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \quad (2)$$

The symbol B represents the number of estimators, f_b indicates the class that specific estimator b forecasts sample x^′ to belong to, and f^ represents the class that sample x′ belongs to (as predicted by the RF).

Evaluation Metrics

The model's efficacy was evaluated using several measures, including ROC analysis, f1-score, recall, accuracy, and precision[20]. The following are the metrics:

- **True positive (TP):** recognizing an assault before it is launched.
- **True negative (TN):** Normalcy detection when it's genuinely normal.
- **False positive (FP):** False alarm occurs when an assault is detected when it is truly normal.
- **False negative (FN):** recognizing normal while it is under threat.

Accuracy

The algorithm may be used to determine the percentage of correctly categorized classes TP and TN over all classifications: Equation (3)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Precision

It is defined as the probability that a positive prognosis would materialize Equation (4)

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

Recall

The True Positive Rate is all that it is. It shows the proportion of invasions that were accurately identified. Equation (5) displays the equation:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

F1-Score

It is defined as the accuracy and sensitivity parameters' harmonic mean. Equation (6) illustrates the formula:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Results Analysis and Discussions

An Intel(R) Core (TM)2 Duo CPU T6670@ 2.20GHz with 4GB of RAM was used to conduct the aforementioned experiment in Matlab. Using the various assessment criteria in the NSL-KDD dataset, Table II shows the results of the RF model's ID. An impressive 96.2% accuracy rate shows that the algorithm is doing a good job of classifying generally. The precision of the model is 97% and indicates the high effectiveness of the model in determining whether a case of intrusion actually occurs, with few false alarms, with a recall of 95% indicating the effectiveness of the model to identify most of the real attacks. The high amplitude RF model of 96% balance F1 further testifies to the power and dependability of this RF model, which makes it suitable for its purpose of detecting intrusion with accuracy and efficiency.

Table 2: Results of RF Model for Intrusion Detection Using NSL-KDD Dataset

Metrics	*RF*
Accuracy	96.2
Precision	97
Recall	95
F1-Score	96

Figure 5 illustrates the RF model's accuracy with respect to training epochs and shows a high and consistent training accuracy throughout the learning process. The accuracy of the training reaches its maximum slowly and levels off at 98.4-98.5, which means that the model has been learned and converged. Although such variations have been observed, the validation accuracy is similar to the training accuracy (around 96.5-97.7%), indicating the high capability to generalize and little overfitting. All in all, the graph indicates that the RF model attains stable and high performance with high predictive stability.

The loss curves of the RF model training and validation with epochs are given in Figure. 6. The training loss gradually reduces until it maintains a low constant. Conversely, the validation loss exhibits clear changes in data between epochs with highs and lows, demonstrating variations in the model's performance on anonymous data. Notwithstanding these oscillations, the loss of validation is in a reasonable range and is not characterized by a steady increase, which speaks in favor of the fact that the model does not overfit dramatically. All in all, the loss behavior confirms that the RF model can learn in a stable manner with reasonable generalization behavior.

Figure 7 shows the confusion matrix of the RF ID model, which shows how accurate the model is at differentiating between malicious and normal transmissions. As evidence

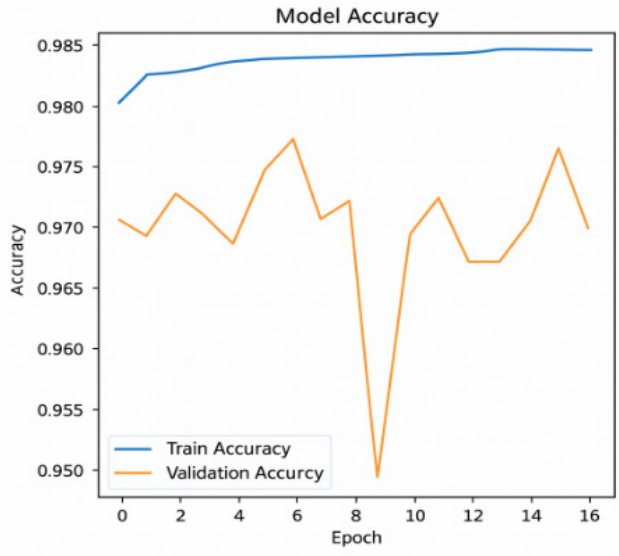

Figure 5: Accuracy Graph of RF Model

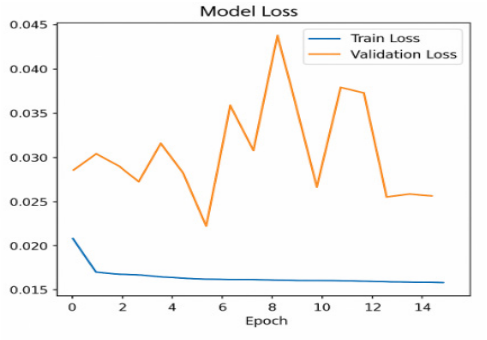

Figure 6: Loss Graph of RF Model

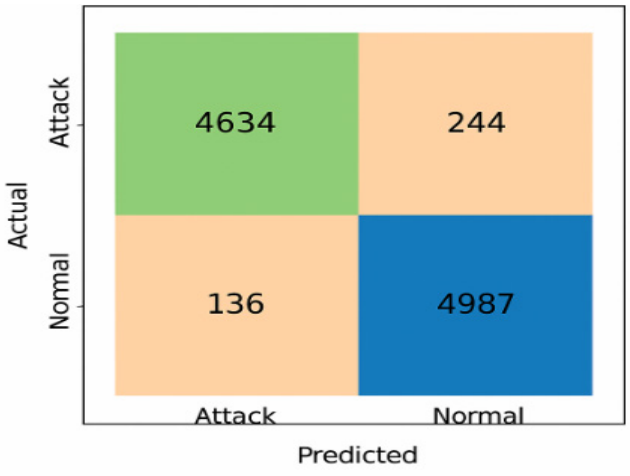

Figure 7: Confusion Matrix of Random Forest

Table 3: comparison for anomaly detection using machine and deep learning

Models	*Accuracy*
MLP [21]	80.5
LSTM[22]	73.18
RF	98.79

of its strong detection capabilities, the model correctly identifies 4,634 attacks and 4,987 routine occurrences. False classification is not very high with 244 attack cases classified as normal and 136 normal cases as attacks. The prevalence of the correctly categorized samples on the main diagonal shows that it is a balanced and highly accurate performance.

A comparison of the accuracy of several models is shown in Table III used in ID. Multilayer Perceptron (MLP) has an accuracy of 80.5% with a moderate level of detection and the LSTM model has a relatively low accuracy of 73.18%, which indicates poor results in this respect. Conversely, RF model outperforms all deep learning models with an accuracy of 98.79%. This significant enhancement signals the efficiency of ensemble-based learning in the representation of intricate intrusion patterns and it shows the excellence of the RF model for accurate and real-time ID.

Conclusion and Future Work

As networked systems have proliferated, and cyber-attacks have become so complex, creating an accurate and reliable ID mechanism has become a very important concern regarding the safety of networks. This work used the NSL-KDD dataset using RF classifier to assess a successful IDS. With a 96.2% accuracy rate, 97% precision rate, 95% recall rate, and 96% F1-score, the test results indicate that the RF model performed exceptionally well, exhibiting a low false alarm rate and a high detection rate. The analysis of accuracy and loss curves verified that the learning behavior is steady and that good generalization is attained. Additionally, the confusion matrix was utilized to support the balanced categorization of normal and attack traffic. Further, the RF model was compared with MLP and LSTM models, which verified that the former model had superior performance and the highest accuracy.

Future research can also entail the extension of the framework to the multi-class intrusion classification of the finer-grained attack identification instead of binary detection. It might also be explored to use the advanced deep learning architectures like hybrid CNNLSTM or transformer-based ones to represent the complex temporal and spatial patterns of attack. Besides, the model should be assessed on more recent and real-time data, such as IoT and cloud-based network traffic, in order to enhance its empirical implementation. They can be used to enhance interpretability and minimize computational complexity by adding feature selection methods and explainable AI (XAI).

References

[1] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," *Int. Conf. Signal Process. Commun. Eng. Syst. - Proc. SPACES 2015, Assoc. with IEEE*, no. July, pp. 92–96, 2015, doi: 10.1109/SPACES.2015.7058223.

[2] A. Jain, B. Verma, and J. L. Rana, "Anomaly Intrusion Detection Techniques: A Brief Review," *Int. J. Sci. Eng. Res.*, vol. 5, no. 7, pp. 1372–1383, 2014.

[3] A. Verma and V. Ranga, "On evaluation of network intrusion detection systems: Statistical analysis of CIDDS-001 dataset using machine learning techniques," *Pertanika J. Sci. Technol.*, 2018.

[4] S. Fenanir, F. Semchedine, and A. Baadache, "A Machine Learning-Based Lightweight Intrusion Detection System for the Internet of Things," *Rev. d'Intelligence Artif.*, vol. 33, no. 3, pp. 203–211, Oct. 2019, doi: 10.18280/ria.330306.

[5] S. M. Othman, F. M. Ba-Alwi, N. T. Alsohybe, and A. Y. Al-Hashida, "Intrusion detection model using machine learning algorithm on Big Data environment," *J. Big Data*, vol. 5, no. 1, 2018, doi: 10.1186/s40537-018-0145-4.

[6] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. M. A. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," *Internet of Things*, vol. 7, Sep. 2019, doi: 10.1016/j.iot.2019.100059.

[7] S. Garg, "Predictive Analytics and Auto Remediation using Artificial Inteligence and Machine learning in Cloud Computing Operations," *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 2, 2019, doi: 10.5281/zenodo.15362327.

[8] J. Veeramreddy, V. V. R. Prasad, and K. M. Prasad, "A Review of Anomaly based Intrusion Detection Systems," *Int. J. Comput. Appl.*, vol. 28, no. 7, pp. 26–35, Aug. 2011, doi: 10.5120/3399-4730.

[9] P. Pathak, A. Shrivastava, and S. Gupta, "A survey on various security issues in delay tolerant networks," *J Adv Shell Progr.*, vol. 2, no. 2, pp. 12–18, 2015.

[10] A. Chauhan and P. R. Vamsi, "Anomalous Ozone Measurements Detection Using Unsupervised Machine Learning Methods," in *2019 International Conference on Signal Processing and Communication (ICSC)*, IEEE, Mar. 2019, pp. 69–74. doi: 10.1109/ICSC45622.2019.8938256.

[11] S. Shriram and E. Sivasankar, "Anomaly Detection on Shuttle data using Unsupervised Learning Techniques," in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, IEEE, Dec. 2019, pp. 221–225. doi: 10.1109/ICCIKE47802.2019.9004325.

[12] M. Zaman and C.-H. Lung, "Evaluation of machine learning techniques for network intrusion detection," in *NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium*, IEEE, Apr. 2018, pp. 1–5. doi: 10.1109/NOMS.2018.8406212.

[13] I. Aljamal, A. Tekeoglu, K. Bekiroglu, and S. Sengupta, "Hybrid Intrusion Detection System Using Machine Learning Techniques in Cloud Computing Environments," in *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)*, IEEE, May 2019, pp. 84–89. doi: 10.1109/SERA.2019.8886794.

[14] M. Dilraj, K. Nimmy, and S. Sankaran, "Towards Behavioral Profiling Based Anomaly Detection for Smart Homes," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2019. doi: 10.1109/TENCON.2019.8929235.

[15] A. M. Vartouni, S. S. Kashi, and M. Teshnehlab, "An anomaly detection method to detect web attacks using Stacked Auto-Encoder," in *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, IEEE, Feb. 2018, pp. 131–134. doi: 10.1109/CFIS.2018.8336654.

[16] H. Luo and S. Zhong, "Gas turbine engine gas path anomaly detection using deep learning with Gaussian distribution," in *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, IEEE, Jul. 2017, pp. 1–6. doi: 10.1109/PHM.2017.8079166.

[17] T. Mehmood and H. B. Md Rais, "Machine learning algorithms in context of intrusion detection," in *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*, IEEE, Aug. 2016, pp. 369–373. doi: 10.1109/ICCOINS.2016.7783243.

[18] H. Liu and B. Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," *Appl. Sci.*, vol. 9, no. 20, p. 4396, Oct. 2019, doi: 10.3390/app9204396.

[19] S. Sapre, P. Ahmadi, and K. Islam, "A Robust Comparison of the KDDCup99 and NSL-KDD IoT Network Intrusion Detection Datasets Through Various Machine Learning Algorithms," *arXiv*, 2019.

[20] G. Meena and R. R. Choudhary, "A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA," in *2017 International Conference on Computer, Communications and Electronics, COMPTELIX 2017*, 2017. doi: 10.1109/COMPTELIX.2017.8004032.

[21] R. R. Devi and M. Abualkibash, "Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD-99 and NSL-KDD Datasets - A Review Paper," *Int. J. Comput. Sci. Inf. Technol.*, vol. 11, no. 03, pp. 65–80, Jun. 2019, doi: 10.5121/ijcsit.2019.11306.

[22] Y. Ding and Y. Zhai, "Intrusion Detection System for NSL-KDD Dataset Using Convolutional Neural Networks," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 2018, pp. 81–85. doi: 10.1145/3297156.3297230.

[23] Gupta, A. K., Polu, A. R., Narra, B., Buddula, D. V. K. R., Patchipulusu, H. H. S., & Vattikonda, N. (2024). Leveraging deep learning models for intrusion detection systems for secure networks. *Journal of Computer Science and Technology Studies*, *6*(2), 199-208.

[24] Narra, B., Buddula, D. V. K. R., Patchipulusu, H., Vattikonda, N., Gupta, A., & Polu, A. R. (2024). The integration of artificial intelligence in software development: Trends, tools, and future prospects. *Available at SSRN 5596472*.

[25] Achuthananda, R. P., Bhumeka, N., Dheeraj Varun Kumar, R. B., Hari Hara, S. P., & Navya, V. (2024). Evaluating machine learning approaches for personalized movie recommendations: A comprehensive analysis. *J Contemp Edu Theo Artific Intel: JCETAI-115*.

[26] Waditwar, P. (2024) The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations. Open Journal of Business and Management, 12, 4073-4085. doi: 10.4236/ojbm.2024.126204.

[27] Bitkuri, V., Kendyala, R., Kurma, J., Mamidala, J. V., Attipalli, A., & Enokkaren, S. J. (2024). A Survey on Blockchain-Enabled ERP Systems for Secure Supply Chain Processes and Cloud Integration. *International Journal of Technology, Management and Humanities*, *10*(04), 126-135.

[28] Mamidala, J. V., Bitkuri, V., Attipalli, A., Kendyala, R., Kurma, J., & Enokkaren, S. J. (2024). Machine Learning Approaches to Salary Prediction in Human Resource Payroll Systems. *Journal of Computer Science and Technology Studies*, *6*(5), 341-349.

[29] Waditwar, P. (2024) AI for Bathsheba Syndrome: Ethical Implications and Preventative Strategies. Open Journal of Leadership, 13, 321-341. doi: 10.4236/ojl.2024.133020

[30] Attipalli, A., Kendyala, R., Kurma, J., Mamidala, J. V., Bitkuri, V., & Enokkaren, S. J. (2024). Privacy Preservation in the Cloud: A Comprehensive Review of Encryption and Anonymization Methods. *International Journal of Multidisciplinary on Science and Management IJMSM*, *1*(1).

[31] Tamilmani, V., Maniar, V., Singh, A. A., Kothamaram, R. R., Rajendran, D., & Namburi, V. D. (2024). A Review of Cyber Threat Detection in Software-Defined and Virtualized Networking Infrastructures. *International Journal of Technology, Management and Humanities*, *10*(04), 136-146.

[32] Singh, A. A. S., Kothamaram, R. R., Rajendran, D., Deepak, V., Namburi, V. T., & Maniar, V. (2024). A Review on Model-Driven Development with a Focus on Microsoft PowerApps. *International Journal of Humanities, Science Innovations and Management Studies*, *1*(1), 43-56.

[33] Padur, S. K. R. (2024). AI-augmented platform engineering: Redefining developer experience through autonomous, self-optimizing enterprise systems. *International Journal of Science, Engineering and Technology*.

[34] Gangineni, V. N., Tyagadurgam, M. S. V., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2024). AI-Powered Cybersecurity Risk Scoring for Financial Institutions Using Machine Learning Techniques (Approved by ICITET 2024). *Journal of Artificial Intelligence & Cloud Computing*.

[35] S. R. Sagili, C. Goswami, V. C. Bharathi, S. Ananthi, K. Rani and R. Sathya, "Identification of Diabetic Retinopathy by Transfer Learning Based Retinal Images," 2024 9th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2024, pp. 1149-1154, doi: 10.1109/ICCES63552.2024.10859381.

[36] S. R. Sagili and T. B. Kinsman, "Drive Dash: Vehicle Crash Insights Reporting System," 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA), Pune, India, 2024, pp. 1-6, doi: 10.1109/ICISAA62385.2024.10828724.

[37] Padur, S. K. R. (2024). Securing Oracle Integration Cloud ERP ecosystems, zero trust architecture, data governance, and compliance automation. *International Journal of Science, Engineering and Technology*, *12*(4), 10-5281.

[38] S. R. Sagili, S. Chidambaranathan, N. Nallametti, H. M. Bodele, L. Raja and P. G. Gayathri, "NeuroPCA: Enhancing Alzheimer's disorder Disease Detection through Optimized Feature Reduction and Machine Learning," 2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichirappalli, India, 2024, pp. 1-9, doi: 10.1109/ICEEICT61591.2024.10718628.

[39] S. R. Sagili, V. K, B. Puli, P. Sundaramoorthy, M. R and K. N V, "Advancing Cervical Cancer Identification using Generative-based Adversarial Networks: An Integrative Learning Methodology," 2025 6th International Conference for Emerging Technology (INCET), BELGAUM, India, 2025, pp. 1-5, doi: 10.1109/INCET64471.2025.11140170.

[40] Routhu, K. K. (2024). Beyond Automation: AI-Powered Employee Engagement Journeys in Oracle HCM Cloud. *KOS Journal of AIML, Data Science, and Robotics*, *1*(1), 1-6.

[41] Routhu, K. K. (2024). The future of HCM: Evaluating Oracle's and SAP's AI-powered solutions for workforce strategy. *Journal of Artificial Intelligence, Machine Learning & Data Science*, *2*(2), 2942-2947.

[42] Sannapureddy, R., Nadella, V. M., & Nelavelli, S. (2024). Edge–Cloud Continuums for Latency-Sensitive Tasks. *International Journal of AI, BigData, Computational and Management Studies*, *5*(4), 189-201.

[43] Arigela, A. K., Brahmareddy, A., Sreenivas, T. S., Selvan, M. P., Venu, N., & Lal, D. K. (2024, December). Optimizing Energy Efficiency and Latency in IoT Devices Through AI-Based Adaptive Protocols in Fog-Edge Computing Environments. In *Congress on Smart Computing Technologies* **(pp. 595-607).** Singapore: Springer Nature Singapore.

[44] Nadella, V. M. (2024). AI-Native 6G Network Management. *American International Journal of Computer Science and Technology*, *6*(1), 23-37.