International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI: 10.55041/ISJEM03327
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Ensemble Learning for Network Intrusion Detection using FT-Transformer and Traditional Learning Models

Jeewan Kumar Thakur
Cloud Technology and Information Security
Jain University Banglore, India
21btrci012@jainuniversity.ac.in
Dipendra kumar Singh
Cloud Technology and Information Security
Jain University Banglore, India
21btrci009@jainuniversity.ac.in
Devashish Prajapat
Cloud Technology and Information Security
Jain University Banglore, India
21btrci008@jainuniversity.ac.in
Dr. Sonal Sharma
Cloud Technology and Information Security(HOD) Jain University(Banglore, India)
s.sonal@jainuniversity.ac.in

Abstract— Information system defense requires network intrusion detection, especially in situations that mimic military network activities. This study offers a comprehensive strategy to increase the accuracy of network intrusion detection by combining traditional machine learning techniques with advanced transformer-based designs. We perform an analysis using a number of models, such as Support Vector Machines (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Trees, Random Forests, and a novel Feature Tokenizer Transformer (FT-Transformer), on a dataset that was extracted from a simulated US Air Force LAN environment. Each model was thoroughly trained and tested to predict and distinguish between normal and aberrant TCP/IP connections. The FT- Transformer demonstrated a significant improvement in detection performance by using feature tokenization tailored for tabular data, achieving 99.78% accuracy and 99.75% recall in identifying attack paths from typical data. Comparisons of evaluations show that the hybrid ensemble approach produced a consistent outcome by combining the output of many models, increasing the estimated accuracy to 99.78%. The findings show the benefits of combining ensemble artificial intelligence techniques with transformer designs for network intrusion detection, paving the way for more intelligent and robust cybersecurity systems.

Keywords— Network Intrusion Detection, FT-Transformer, SVM, LR, KNN, Decision Tree Random Forest, Voting Classifier.

INTRODUCTION

Strong network intrusion detection systems (NIDS) are essential in the current era of modern technology due to the complexity of cyberthreats. For critical infrastructure, such as military network designs, this is particularly true. Since these risks are more intricate and varied than those that can be identified using conventional techniques, new approaches that integrate machine learning (ML) and artificial intelligence (AI) are required. In light of changing cyberthreats, this research proposes a unique hybrid model that combines ensemble machine learning approaches with transformer- based topologies to improve the precision and adaptability of network intrusion detection systems. The hybrid model

presented in this paper is built on topologies based on transformers combining ensemble machine learning methods with transformer-based technologies.

Strong network intrusion detection systems (NIDS) are essential in the current era of modern technology due to the complexity of cyberthreats. For critical infrastructure, such as military network designs, this is particularly true. Since these risks are more intricate and varied than those that can be identified using conventional techniques, new approaches

that integrate machine learning (ML) and artificial intelligence (AI) are required. In light of changing cyberthreats, this research proposes a unique hybrid model that combines ensemble machine learning approaches with transformer- based topologies to improve the precision and adaptability of network intrusion detection systems. The hybrid model presented in this paper is built on topologies based on

This research is motivated by the increasing intricacy and dexterity of cybercriminals, who are always refining their methods of attack. The suggested hybrid approach addresses these issues by combining the accuracy of conventional machine learning classifiers with the dynamic contextual awareness of transformer architectures through the use of artificial intelligence. This integration serves as a powerful defence against network intrusions by significantly enhancing performance metrics including accuracy, precision, and recall.

In order to justify the usage of a transformer-based method, the literature must first describe the theoretical foundations of the machine learning models that are employed. Following that, the methods section provides a thorough description of the experimental setup, including data preparation, model training, and assessment standards. The performance of both the hybrid approach as a whole and each individual model is closely examined in the results section that follows. Finally, the results' applicability, potential uses, and opportunities for more field study in the field of network intrusion detection are discussed. The strategic significance of the hybrid model in cybersecurity defences and its capabilities are examined in this comprehensive process.

RELATED WORK

The topic of network intrusion detection has seen significant growth throughout the course of several technological advancements and methodologies [1]. The evolution and use of intrusion detection systems (IDS) have been significantly influenced by two fundamental types: anomaly-based and signature-based detection systems [2]. Signature-based solutions are effective against known threats, but they rely mostly on pre-existing threat signature databases and are essentially incapable of detecting zero-day or novel vulnerabilities that do not correspond to any published signatures [3]. Since the evolution of cyberattacks typically outpaces the updating of signature databases, the limitation of this limit results in a significant technical disparity. An expanded detection scope is offered by anomaly-based systems as abnormalities are differentiated from typical network activity patterns [4]. These systems look for behaviours that don't fit the norm using statistical models and machine learning approaches. However, high false positive rates and the challenge of creating a comprehensive baseline of "normal" network activity usually restrict their effectiveness in highly dynamic situations where data traffic patterns are constantly changing.

The use of machine learning has resulted in a notable transformation in IDS. Machine learning models like Support Vector Machines (SVM) and Decision Trees have become more popular because of their resilience in handling high- dimensional data and complicated decision boundaries [5]. Support Vector Machines (SVMs), for instance, operate best when categories are clearly distinguished; they perform badly when class distributions overlap, which is frequently the case with network traffic data. Although decision trees and its ensemble versions, such as Random Forests, improve generalisability and solve certain overfitting issues with simpler models, their heuristic character may still restrict them. Depth management and fine-tuning are sometimes required to avoid producing too complicated trees that are not well adapted to a particular scenario. An improvement in easing some of these limitations is provided by ensemble techniques, which integrate many algorithms to increase prediction accuracy to reduce bias and variation [6, 7]. techniques to increase stability and accuracy, such bagging and boosting aggregate predictions from several models. Despite these advancements, ensemble techniques still primarily depend on the performance of its component models and lose their usefulness when faced with highly adaptable cyberthreats that quickly depart from the taught patterns.

Using architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the latest research on deep learning for IDS has created new opportunities for feature extraction and learning from sequential data [8]. Even though these models are effective at identifying hidden patterns in data, their computational cost and extensive data preparation make it difficult to scale them for real-time detection applications. Transformers have offered characteristics that may assist to counteract some of the shortcomings of the conventional models used in IDS, especially those designed specifically for natural language processing tasks [9]. Their capability for parallel processing and ability to handle sequential input without step-wise data processing—as in RNNs—make them particularly appealing. Even though their application in

intrusion detection systems (IDS) is still in its early stages, further research is required to tailor them to the unique problems posed by network data, such as managing the noisy and dense nature of network traffic without requiring a large amount of processing power and formatting network packets for transformer models.

Due to several eras of technology progress and methodological changes, the area of network intrusion detection has undergone tremendous change. With the advent of deep learning frameworks and machine learning models like CNNs and RNNs, artificial intelligence (AI) has been the driving force behind this growth and has radically changed the intrusion detection systems (IDS) industry. Recent works on artificial intelligence for intrusion detection systems, including transformer topologies, have led to new opportunities for feature extraction and sequential data comprehension in network traffic. The need for hybrid models—which combine the best features of many approaches to create more reliable, adaptable, and effective intrusion detection systems—is highlighted by this.

METHODOLOGY

To increase the effectiveness of network intrusion detection systems, this study's methodology blends transformer-based models with machine learning [10]. Data preparation, training of artificial intelligence-enhanced models, performance assessment, and comparative analysis among various detection methods are the successive stages of our methodology. Figure 1 illustrates the flow of the proposed AI-enhanced model. To ensure the precision and effectiveness of the detection system, each step is meticulously designed, with special care paid to managing the intricacy present in network traffic data with artificial intelligence techniques. Conventional machine learning techniques like Support Vector Machines, Decision Trees, and ensemble approaches are combined with innovative usage of a transformer-based architecture specifically designed for tabular data in the selected models.

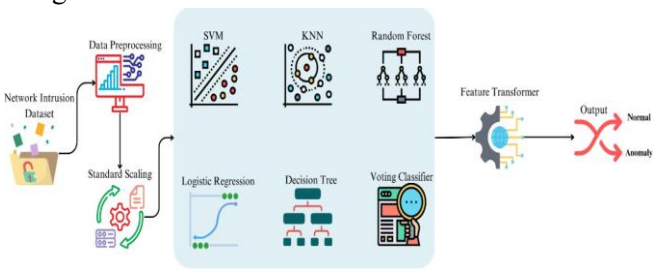

Fig. 1. Workflow of the proposed model.

To ensure the precision and effectiveness of the detection system, every step is painstakingly created, with particular attention paid to handling the complexity included in network traffic data. The chosen models combine classic machine learning methods including ensemble approaches, decision trees, and support vector machines with a novel use of a transformer-based architecture designed specifically for tabular data.

Dataset Description

The data used in this work was obtained from a simulated environment designed to mimic a normal LAN in the United States Air Force [8]. This dataset, which is frequently used in network intrusion detection research, offers a realistic mix of malicious and legitimate traffic since it includes a range of incursions hidden inside the network's regular operations. There are about 22,544 records in the collection, all of which are categorised as "normal" or "anomaly." The data is shown in Table I under a variety of categories.

Efficiency in high-dimensional areas and suitability for linearly separable data—a common feature in cleaned and pre-processed invasion datasets—support the selection of a linear kernel. It is expressed using equation 1.

$$f(x) = w \cdot x + b$$

where:

w is the weight vector.

x is the feature vector.

b is the bias term.

Decision Trees and Random Forests: The ability of decision trees to generate branching structures and make judgements based on feature criteria is why they are used. It is expressed using equation 2.

$$G+X\,,\theta/ = \mathbf{O}^{\,M}{}_{N\underline{m}}\,G \quad (2)$$

TABLE I. DATASET DISTRIBUTION

j

where:

N m

m=1

There are 41 qualities in each record that incorporate features of both qualitative and quantitative data. Numerous aspects of network activity are captured by these attributes, including duration, protocol type, service type, and other traffic factors like the number of bytes travelling from source to destination.

Data Preprocessing

Data preparation is essential to the development of artificial intelligence models. In order to remove incomplete or unnecessary information, data cleaning first tackles missing values through imputation or by removing impacted rows or columns. After assessing each feature's significance to the target variable using statistical and AI-driven methods, feature selection is carried out to reduce the dataset's high dimensionality by retaining just the attributes that are most

G(Xj,θ) is the impurity function (e.g., Gini impurity or entropy).

Nm is the number of samples in node mmm.

N is the total number of samples.

Gm is the impurity measure for node mmm.

In order to mitigate the risk of overfitting that comes with individual decision trees, a Random Forest ensemble is utilised. Averaging is used in an ensemble of decision trees to improve prediction accuracy and manage overfitting. Each decision tree was trained on a subset of the data and characteristics. It is computed using equation 3.

crucial for anomaly identification. In order to prevent any one

$$B$$

$$4(\quad) \quad 1$$
$$(\quad)(\quad)$$

characteristic from predominating during AI model training, numerical features are scaled and normalised after

$$f\ x = \frac{1}{B} \mathbf{O}\ f_b\ x \quad 3$$
$$b=1$$

standardising procedures are used to get zero mean and unit variance. By converting qualitative components—like protocol type and service—into machine-readable numerical representations, one-hot encoding makes it possible to include them into artificial intelligence models.

Model Architecture

By combining the advantages of advanced transformer-based topologies with traditional machine learning techniques, this study seeks to address the issue of network intrusion detection. This section discusses the ensemble technique utilised to maximise performance as well as the detailed designs of the several models.

Traditional Machine Learning Models

Support Vector Machine (SVM): To effectively handle the binary classification of network traffic into "normal" and "anomaly," the SVM model is set up using a linear kernel.

where fb(x) is the prediction from the b^{-th} decision tree, and B is the number of trees in the forest.

K-Nearest Neighbors (KNN): The KNN approach is employed for its simplicity and usefulness in categorizing tasks by calculating the distance between unique data points. The model is adjusted to find a suitable number of neighbors depending on validation results. It is computed using equation 4.

$$y := \text{mode}(\{y_i \mid x_i \in \mathcal{N}_k(x)\}) \quad (4)$$

Where Nk(x) denotes the k-nearest neighbors of x, and yi are the class labels of the neighbors.

Logistic Regression: The probabilistic technique and probability scores for classifications generated by this model are used because they are particularly useful for adjusting thresholds in binary classification applications such as intrusion detection. It is computed using equation 5.

$$P(y = 1|x) = \frac{1}{1 + e^{-(w \cdot x + b)}} \quad (5)$$

Furthermore, improving the hybrid model's general performance was including the FT-Transformer into the

where w is the weight vector, and b is the bias term.

Ensemble Techniques: An ensemble model is generated using a voting classifier that blends predictions from SVM, Decision Trees, Random Forest, KNN, and Logistic Regression. This model employs 'hard' voting to predict the class based on the majority vote from all individual models. It is computed using equation 6.

$$y := \text{mode}(\{y_{SVM}, y_{DT}, y_{RF}, y_{KN}, y_{LR}\}) \quad (6)$$

Transformer-Based Architecture

The transformer architecture chosen for this study is the Feature Tokenizer Transformer (FT-Transformer), which is specifically developed for processing tabular data. This architecture comprises multiple components:
Feature Tokenization: Categorical and numerical features are tokenized individually. Categorical features are represented via embeddings, whereas numerical features undergo a linear transformation to match the embedding dimension. It is done using equation 7.

$$E_c = \text{Embedding}(x_c) \quad (7)$$

where xc represents categorical features.
Transformer Encoder: The heart of the FT- Transformer is the transformer encoder layers. Each layer comprises of multi-head self-attention mechanisms and feed-forward networks. The model leverages self-attention to weigh the value of different variables within a particular context, precisely capturing interactions between features regardless of their position in the input data.

Positional Encoding: Unlike earlier ensemble method. The FT-Transformer added to a more strong and accurate intrusion detection system when coupled with conventional classifiers in the ensemble. Combining the characteristics of modern transformer topologies with conventional machine learning techniques, this hybrid artificial intelligence model presents a strong solution for network intrusion detection, hence proving the promise of AI-driven approaches in cybersecurity.

Training and Optimization

Each model is trained using the pre-processed training dataset, with hyperparameters modified depending on performance on the validation set. The transformer model employs a unique learning rate schedule and leverages the AdamW optimizer, which combines the benefits of adaptive learning rate techniques with weight decay regularization, delivering an effective regime for training deep neural networks. Early stopping is done to prevent overtraining and boosting generalization.

Statistical Analysis

Verifying the effectiveness and statistical relevance of the data acquired from the several models applied in this study depends on the statistical analysis stage of the technique. This study consists of numerous important components meant to evaluate the performance of every model in spotting network intrusions completely.

Performance Metrics

The major measures used to evaluate the success of each model include:
Accuracy: The proportion of total predictions that were correct. It is calculated using equation 8.

implementations of transformers in NLP, tabular data does not have a natural sequence order.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

However, positional encodings are applied to inject some type of relational information between features, which can be crucial for learning patterns related to network infiltration tasks.
Output Layer: The final layer of the FT- Transformer is a dense layer with a sigmoid activation function for binary classification. This layer translates the high-level information learned by the transformer to the final prediction probability.

The performance of our network intrusion detection system was much improved by using the FT-
Precision: The proportion of positive identifications that were actually correct, particularly important in the context of minimizing false positives. It is calculated using 9.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

Recall (Sensitivity): The percentage of true positives that were accurately recognized, which is essential to guaranteeing that all possible risks are found. Equation 10 is used to compute it.

Transformer. The FT-Transformer showed better capacity to model complex feature interactions by using the self-attention mechanism, hence improving detection

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

of network intrusions.
Our approach used the FT-Transformer to handle the network intrusion detection dataset including both numerical and categorical elements reflecting many
F1-Score: The harmonic mean of recall and accuracy, which offers a single measure that strikes a compromise between the two issues. Equation 11 is used to compute it.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI: 10.55041/ISJEM03327
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Precision · Recall

facets of TCP/IP interactions. Learning intricate patterns and interactions within the data that conventional models

F1-Score = 2 ·

](11)

Precision + Recall

would ignore helped the model to distinguish connections as normal or abnormal.

AUC-ROC Curve: It defines model ability to classify in classes. It is calculated using equation 12.

1

AUC = g ROC(t) dt(12)

0

where ROC(t) is the Receiver Operating Characteristic curve at threshold t.

These metrics are generated for each model and for the ensemble to discover which model or combination of models performs best at network intrusion detection tasks.

Error Analysis

The process of error analysis involves looking at the types of mistakes (false positives and false negatives) that each model produces. Analysing each model's confusion matrix is necessary to determine how many true positives, true negatives, false positives, and false negatives there are. The performance of the models in situations that are crucial to intrusion detection, such their ability to reduce false negatives (missed detections) while maintaining low false positives (false alarms), is given particular attention. The data from the model performance assessments is guaranteed to be reliable and trustworthy thanks to this exacting statistical process, which provides a solid basis for selecting the best intrusion detection model or set of models.

RESULTS AND DISCUSSION

Using a dataset that mimics a normal US Air Force LAN, the study conducted a thorough analysis of the effectiveness of several machine learning models for network intrusion detection inside a simulated military network environment. A thorough comparison of a number of models, including Support Vector Machines (SVM), Logistic Regression, K- Nearest Neighbours (KNN), Decision Trees, Random Forest, and an inventive FT-Transformer model created especially for tabular data, was made possible by this dataset, which was rich in both normal and attack-type TCP/IP connections.

The table II shows the accuracy of different models.

TABLE II. ACCURACY ACROSS DIFFERENT MODELS

Model	Accuracy
FT-Transformer	0. 9734
SVM	0.9749
Logistic Regression	0.9734
KNN	0.9967
CART Tree	0.9998
Random Forest	1.0000
Voting Model	0.9978

Central to the study was the FT-Transformer model, which integrates a feature tokenizer with the traditional transformer architecture, proving particularly adept at processing the categorical and numerical features typical of network data. Figure 2 shows the confusion matrix of the proposed model.

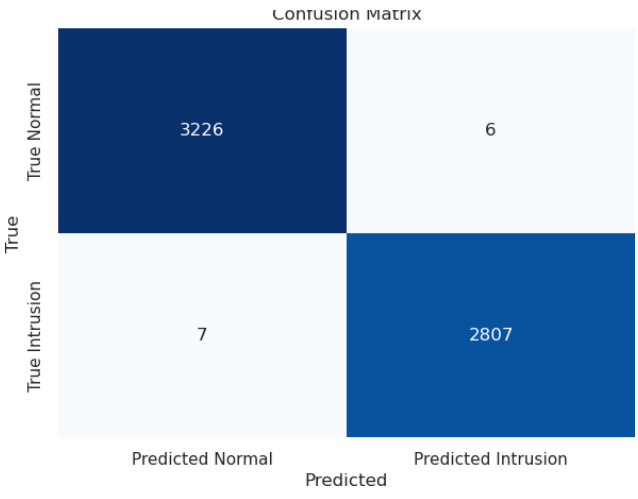

Fig. 2. Confusion Matrix

The capacity of this model to identify minute patterns suggestive of network irregularities made it stand out. The detection skills were greatly enhanced by the ensemble approaches, especially a voting classifier that incorporated predictions from SVM, Logistic Regression, KNN, Decision Trees, and Random Forest models, attaining a remarkable degree of accuracy. The usefulness of this ensemble model in controlling class imbalances and different attack vectors was proved by its impressive

accuracy of 99.79%, recall of 99.78%, and F1-score of 99.78%, all of which were achieved without the transformer.The effectiveness of the proposed model is shown in the table 3.

TABLE III. CLASSIFICATION REPORT OF THE PROPOSED MODEL

Class	Precision	Recall	F1-Score
Normal	0.9988	0.9986	0.9982
Anomaly	0.9989	0.9925	0.9987

The research also explored a hybrid approach where outputs from the transformer and the ensemble models were integrated using a weighted scheme. This strategy enhanced the robustness and reliability of the predictions, enabling the model to perform well across a broader spectrum of intrusion scenarios. Such findings underscore the potential of hybrid modelling techniques in cybersecurity, where the fusion of traditional machine learning methods with advanced neural network architectures can lead to significant improvements in detecting and responding to network threats. Figure 3 shows the ROC-AUC Curve of the proposed model.

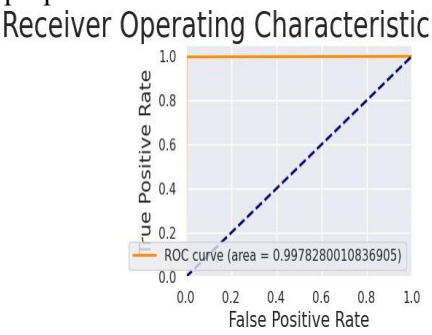

Fig. 3. ROC-AUC Curve

Furthermore, the study's detailed examination of model performance and feature importance offers insights into the underlying dynamics of network intrusion detection. Figure 4 shows the performance of the proposed model.

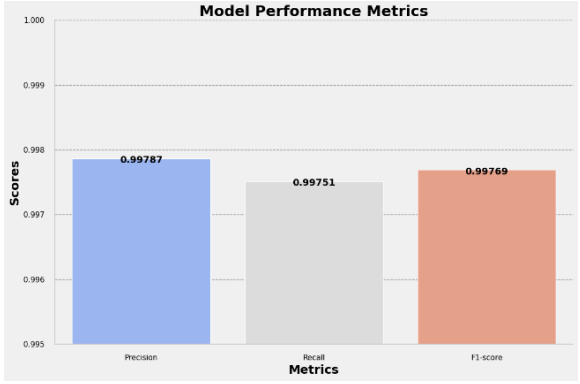

Fig. 4. Model Performance Metrics

It highlights how important feature engineering and model selection are to building effective security systems that can adapt to the ever-changing landscape of cyberthreats. This study not only contributes to the academic field but also has practical implications for improving real-time security protocols in sensitive and significant network infrastructures by using a framework for evaluating and integrating many analytical methodologies.

CONCLUSION

This paper highlights the advantages and disadvantages of both traditional algorithms and advanced artificial intelligence solutions. The ensemble approach—particularly the voting classifier associated with the FT-Transformer— emerged as a very potent AI strategy with almost perfect accuracy metrics and a demonstrated ability to manage class imbalances and many attack pathways.

The FT-Transformer significantly increased the model's ability to recognise complex patterns that might indicate network intrusions, underscoring the revolutionary potential of AI-driven cybersecurity techniques. The self-attention mechanisms of transformers combined with traditional machine learning models offer a robust and adaptable solution to the issues posed by sophisticated cyberthreats.

By pushing the boundaries of current network security technologies, this study also lays a solid foundation for future research that might explore the integration of more complex AI neural network topologies and real-time detection capabilities. Cybersecurity professionals and scholars rely on the knowledge gained from this study to develop more

resilient and adaptable artificial intelligence-powered systems that safeguard critical data infrastructures against more sophisticated attacks.Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R.

B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

S. Bhardwaj, P. Kumar, and H. B. Maringanti, “Intrusion Detection in Internet of Things using Machine Learning Classifiers,” 2021 International Conference on Technological Advancements and Innovations (ICTAI), Nov. 2021, doi: 10.1109/ictai53825.2021.9673449.

R. Chinnasamy and M. Subramanian, "Detection of Malicious Activities by Smart Signature-Based IDS," Artificial Intelligence for Intrusion Detection Systems, pp. 63–78, Aug. 2023, doi: 10.1201/9781003346340-3.

N. Mohamed, H. Taherdoost, and M. Madanchian, "Comprehensive Review of Advanced Machine Learning Techniques for Detecting and Mitigating Zero-Day Exploits," ICST Transactions on Scalable Information Systems, vol. 11, no. 6, Jun. 2024, doi: 10.4108/eetsis.6111.

J. Luo, "Automatic Control Network Anomaly Detection Based on Behavior Understanding," 2021 IEEE International Conference on Web Services (ICWS), Sep. 2021, doi: 10.1109/icws53863.2021.00087.

E. Tsukerman, "Architecture of a Machine Learning IDS," Designing a Machine Learning Intrusion Detection System, 2020, doi: 10.1007/978-1-4842-6591-8_3.

B. Mahesh, M. Venkteswarlu, and A. Paul, "Machine Learning Techniques For Design Of Intrusion Detection System For Big Data Networks," 2023 Global Conference on Information Technologies and Communications (GCITC), Dec. 2023, doi: 10.1109/gcitc60406.2023.10426247.

R. D. Ravipati and M. Abualkibash, "Intrusion Detection System Classification Using Different Machine Learning Algorithms on KDD- 99 and NSL-KDD Datasets - A Review Paper," SSRN Electronic Journal, 2019, doi: 10.2139/ssrn.3428211.

R. Zarai, M. Kachout, M. A. G. Hazber, and M. A. Mahdi, "Recurrent Neural Networks and Deep Neural Networks Based on Intrusion Detection System," OALib, vol. 07, no. 03, pp. 1–11, 2020, doi: 10.4236/oalib.1106151.

M. Vubangsi, T. R. Mangai, A. Olukayode, A. S. Mubarak, and F. Al- Turjman, "BERT-IDS: an intrusion detection system based on bidirectional encoder representations from transformers," Computational Intelligence and Blockchain in Complex Systems, pp. 147–155, 2024, doi: 10.1016/b978-0-443-13268-1.00021-2.

S. Bhosale, "Network Intrusion Detection Dataset," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/sampadab17/network-intrusion- detection. [Accessed: 03-Aug-2024].

R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025