International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, Delaware, USA
Volume 13– No. 1, August 2025 – www.ijais.org

Data-Driven Detection of Network Threats using Advanced Machine Learning Techniques for Cybersecurity

Bhavana Kamarthapu
Fairleigh Dickinson University

Mitra Penmetsa
University of Illinois at Springfield

Jayakeshav Reddy Bhumireddy
University of Houston

Rajiv Chalasani
Sacred Heart University

Srikanth Reddy Vangala
University of Bridgeport

Ram Mohan Polam
University of Illinois at Springfield

ABSTRACT

The more sophisticated and diverse the network threats become, the lower the conventional intrusion detection systems' precision and versatility. This work provides a Data Driven Intrusion Detection System (IDS) based on Artificial Neural Networks (ANN) in combination with Principal Component Analysis (PCA) to improve features and minimize dimensionality. A significant amount of preprocessing is performed on the proposed model including missing value handling, normalization and removal of outliers for quality data. The ANN model outperformed the benchmark models Random Forest and Isolation Forest, with 97.5% detection accuracy, 99.0% precision, 96.7% recall, and 95.7% F1-score on the NSL-KDD dataset. These findings also demonstrate that the ANN-based IDS can effectively identify complex and dynamic cyber threats and solve a number of real-world cybersecurity issues. In addition, the model shows strong generalization and efficient learning over validation criteria across dynamic network environments which validates the stability and practicability of the model.

Keywords

Threat Detection, NSL-KDD Dataset, Artificial Neural Network, PCA technique, Cybersecurity, Machine Learning.

1. INTRODUCTION

Organizations all over the world continue to encounter increasingly complex and ever-changing network threats that pose serious risks. A DoS attack, malware infection, brute force password attempts and insider exploits are some of these threats. By definition, insiders are authorized so they can especially bypass traditional security measures [1]. Given the increase in network size, diversity of attacks and interconnectivity of networks, it is a key problem to detect these diverse attacks quickly and accurately for keeping systems reliable and secure. In addition, the emergence of more complex attack techniques and the growing popularity of encrypted communication have undermined the efficiency of traditional detection techniques, therefor giving rise to the demand for new, more sophisticated methods.

Protecting sensitive data and system integrity is vital to defend networks against these growing threats and cybersecurity is the backbone of that effort [2][3]. Intrusion detection systems and firewalls that rely on signatures are traditionally great against known threats, but they often do not work when novel or unknown attacks are at play. The limitations of defence in depth leave gaps primarily against advanced forms of insider threats and zero-day exploits [4]. As a result, cybersecurity is going the way of smarter, more adaptive solutions ones that can scan and process big chunks of network data to recognize anomalous behaviour in real time. To minimize loss and keep trust in digital infrastructure, it is essential to be able to respond quickly to emerging threats.

Data-driven detection of network threats has become a powerful tool to perform cybersecurity with the help of ML techniques [5][6]. Even from historical data, machine learning models are able to identify patterns of harmful activity, both known and previously unknown. ML based systems that combine supervised learning with anomaly detection provide added capability to detect threats with higher concurrency and fewer false alarms [7]. Moreover, ML algorithms are continuously changing, and the number of cybersecurity datasets is growing, hence, The creation of detection frameworks that are more resilient and expandable is feasible [8]. This paper investigates how advanced ML methods can be utilized to construct effective network threat detection systems that bolster overall cybersecurity defences.

1.1 Motivation and Contributions of the Study

Cyber-attacks are becoming more frequent and complex, while traditional ML models fail to detect changing threats with sufficient precision, as they lack generalizability or the capability to capture complex patterns of data. The motivation for this study comes from the requirement for an intelligent, high performing and adaptive intrusion detection framework which can successfully detect known as well as novel attacks. In this domain, DL models, especially ANN, promises promising capabilities and for that it will be necessary to analyze the capabilities of these models and to fill in performance and reliability gaps of existing Intrusion detection systems. My key contributions of this research are as follows:

- A comprehensive IDS framework using ANN integrated with PCA for better feature extraction and down dimensionality of data is proposed.
- Takes care of missing values, normalizes and removes outliers to improve input quality, implements a rigorous

pre-processing pipeline.

- Demonstrates superior detection performance with an accuracy, significantly outperforming traditional models like Random Forest and Isolation Forest.
- Provides detailed evaluation metrics and visual analysis (confusion matrix, loss, and accuracy graphs) to validate the model's stability, learning efficiency, and real-world applicability.

1.2 Novelty and Justification of the Study

This study proposes a novel network threat using an ANN along with a PCA for feature extraction on the NSL-KDD dataset. In contrast to conventional models, this method adopts a pre-processing perspective, enhancing detection accuracy and computational efficiency by pre-processing techniques like dimensionality reduction, normalization, and outlier elimination. These results indicate that the ANN model can learn complex, non–linear patterns in network traffic data much better than conventional ML models, which justifies its implementation for real-time cybersecurity applications. The extensive evaluation using precision, recall, F1-score, and accuracy confusion matrix shows its robustness in detecting complex network threats.

1.3 Structure of the Paper

The study is organized as follows: Section 2 discusses pertinent work on Machine Learning Network Threat Detection. Section 3 describes the techniques, methodologies, and materials employed. Section 4 contains the experimental findings, result analysis, and a discussion of the suggested system. Section 5 outlines the conclusion and future work.

2. LITERATURE REVIEW

This section briefly reviews recent intrusion detection methods and behavioral models. Each method demonstrates effective detection of malicious activities across various network environments.

A and Sundarakantham (2019) approach utilizes SVM for classifying network traffic data. Their system analyzes large-scale traffic data to identify malicious activities within networks. The proposed SVM-based model achieves 97.29% classification accuracy, outperforming other traditional methods [9].

Chu, Lai, and Liu's (2019) Present-day intrusion detection methods can't handle the complexity and exponential growth. The suggested method is assessed using a Modbus protocol gas pipeline dataset and contrasted with current approaches. The approach's suitability for intrusion detection is demonstrated by the industrial control systems' 97.56% accuracy, 2.42% FPR, and 2.51% miss rate, respectively [10].

Begli, Derakhshan and Karimipour (2019) suggest an architecture that carries out a safe remote medical system. They want to provide a safe framework for remote healthcare systems that protects system data from typical network threats, such as U2R and DoS assaults. They accomplished this by designing an IDS using the SVM, ML technique. Evaluation metrics of the layered architecture of IDS, once their technique is put into practice, demonstrate the effectiveness of their suggested framework [11].

Srivastava, Agarwal and Kaur (2019) utilized state-of-the-art machine learning techniques based on feature reduction to identify odd patterns in the recently provided dataset. It has achieved a remarkable accuracy rate of 86.15 percent. Intrusion detection technology is growing quickly. Network traffic data invasions were previously discovered by supervised learning methods. But in addition to traffic's sharp increase in recent years, network risks are also changing [12].

Kim, Hong, and Han (2018) present an insider threat detection method using a Markov Chain Model to identify anomalous user behaviour based on transition probabilities between activity states. By modelling user behaviour sequences and applying ML algorithms, the system effectively classifies insider threats. Experiments on 15% of the CERT insider threat dataset confirm the model's behaviour-based categorization with an accuracy of up to 97% (not include Naive Bayes) [13].

Table 1 summarizes Network Traffic using ML Techniques literature study, including methodologies, datasets, main results, limits, and planned initiatives.

Table 1. Summary of the Study on Data-Driven Detection of Network Threats for Cybersecurity

Author	Methodology	Data	Key Findings	Limitations & Future Work
A and Sundarakantham (2019)	Support Vector Machine (SVM)	Large-scale network traffic data	Achieved 97.29% accuracy; outperformed traditional methods	Future work could explore hybrid models or real-time detection capabilities
Chu, Lai, and Liu (2019)	Custom ML-based IDS for ICS	Gas pipeline dataset (Modbus protocol)	97.56% accuracy, 2.42% false-positive rate; effective for industrial systems	Needs adaptation to more varied ICS environments and real-time deployment
Begli, Derakhshan, and Karimipour (2019)	SVM-based IDS in layered architecture	Remote healthcare system dataset	Efficient detection against DoS and U2R attacks; secure framework demonstrated	Future enhancements can include hybrid models and deeper security integration
Srivastava, Agarwal, and Kaur (2019)	Feature reduction + ML classification	Recent network traffic dataset	86.15% accuracy; effective anomaly detection using novel feature reduction	Requires improvement in handling high-dimensional data and evolving attack patterns
Kim, Hong, and Han (2018)	Markov Chain + ML for behavior analysis	CERT insider threat dataset (15% sample)	Up to 97% accuracy (excluding Naive Bayes); effective behavior-based detection	Needs full dataset evaluation and comparison with deep learning techniques

3. METHODOLOGY

The proposed methodology begins with the NSL-KDD dataset, which undergoes a detailed data pre-processing phase to ensure high-quality input for the machine learning model. This entails dealing with missing values, eliminating outliers, and normalizing the data to guarantee feature consistency. In order to reduce dimensionality by keeping just the most pertinent qualities, features are subsequently extracted using Principal Component Analysis (PCA). An artificial neural network (ANN) is used to segregate and classify the training and testing sets of processed data. Using measures like accuracy, precision, recall, and F1-score, the model's performance is evaluated to give insight into how well it can recognize and classify different kinds of network intrusion. The entire flowchart from data pre-processing to performance evaluation is demonstrated in Figure 1.

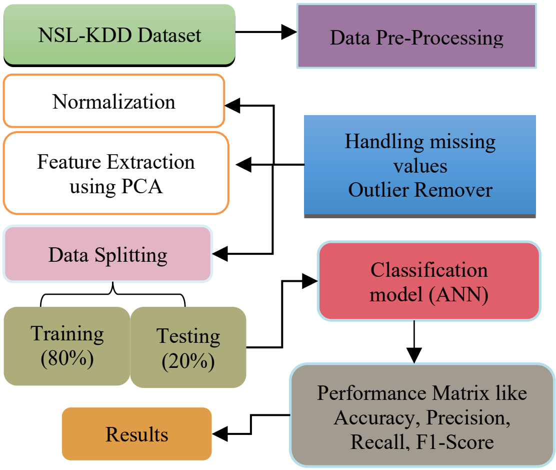

Fig 1: Proposed Flowchart for Network Threat Using Cybersecurity

Below are briefly explained the following steps of the flowchart:

3.1 Data collection

The 2009 NSL-KDD dataset, which includes two test sets (KDDTest+ and KDDTest21) and a training set (KDDTrain+), is presented in this study as an enhanced benchmark for network intrusion detection. It provides more representation to rare attacks and balances number of instances by difficulty level, unlike the KDD Cup '99 original dataset, but without duplicate entries. This enables researchers to fine-tune and analyzed machine learning models on completed datasets, improving the validity and fairness of intrusion detection assessments.

3.2 Data analysis and visualization

Data visualization involves displaying data visually. Data visualization tools use maps, graphs, and charts to show data patterns and trends. Below, provide the graphics from the Network Threats visualization:

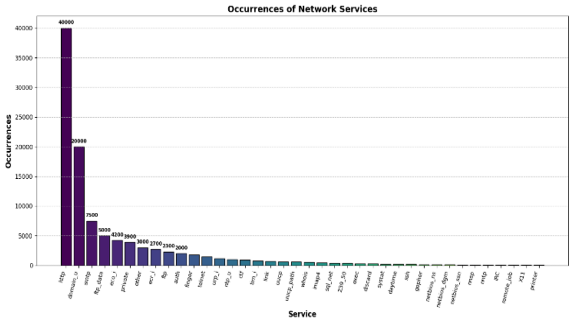

Fig 2: Categorical Feature 'Service' Vs. Occurrences

In the Figure 2 present a bar chart which shows the frequency of different network services in a dataset. The x-axis shows different kinds of services (such as HTTP, DNS, FTP, and SSH). Each service's number of instances is shown on the y-axis. The chart reveals a highly imbalanced distribution, with a few services like HTTP and DNS dominating in frequency, each with over 10,000 occurrences, while most other services have significantly fewer appearances, many below 1,000. Thus, the imbalance indicates that proper treatment should be applied to data during preprocessing as well as model training to prevent biased learning outcomes.

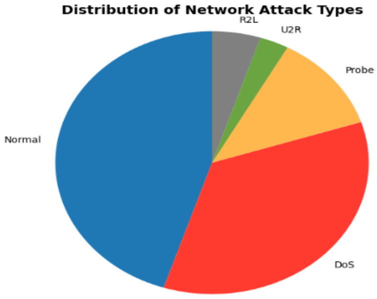

Fig 3: Attack Type vs. Normal

Figure 3 shows the distribution of network traffic of various kinds in a dataset. Each category is represented by a colour and Displayed are Probe, DoS (Denial of Service), Normal R2L (Remote to Local) and U2R (User to Root). As can be seen from the chart, the vast majority of traffic fits into the "Normal" and "DoS" categories, with Normal being the biggest of the two. However, only a small share belongs to the "Probe" class and "U2R" and "R2L" classes represent a tiny fraction of the data, suggesting that class imbalance should be considered during training and evaluation of model.

3.3 Data Preprocessing

A fundamental and first step in converting unprocessed data into information that may be used is data preparation. This research incorporates procedures like managing missing data to provide consistent and high-quality input to ML algorithms, outlier detection, normalization and feature extraction, which have been necessary to work with complex network threat data in this study. Pre-processing procedures were methodically performed to the NSL-KDD dataset in order to achieve that:

- **Handling Missing Values:** To preserve data integrity while handling missing values, gaps in the dataset must be found and filled. The common techniques are removal of incomplete records or imputing the missing values by methods like mean, median or interpolation.

- **Outlier Detection and Removal:** Finding outliers is a crucial first step in many data mining projects. Here, the outlier identification process aims to separate out the hundreds of characteristics that are impacted by outlier tools.

3.4 Normalization:

The method of normalizing data involves scaling the attribute values such that it is equally significant and lie within the same numerical interval or scale. A common normalization function is Z-score normalization. Equation (1) is defined as:

$$z_{score} = \frac{X-\mu}{\sigma} \quad (1)$$

Where, X is the input of a model μ and σ represent the mean and the standard deviation calculated over X respectively.

3.5 Feature Extraction

The two main purposes of feature extraction are invariance and data compression. Usually taken from an oversampled collection of measurements, A properly selected set of features makes up the feature vector. Redundancy in the representation is intended to be removed by the feature selection stage. By collecting the greatest variation in the data, PCA, a linear dimensionality reduction approach, helps minimize overfitting and enhance IDS performance by identifying the most significant features. It streamlines data by eliminating duplication and noise, but it ignores class labels and non-linear correlations. Figure 4 illustrates the use of PCA in the analysis.

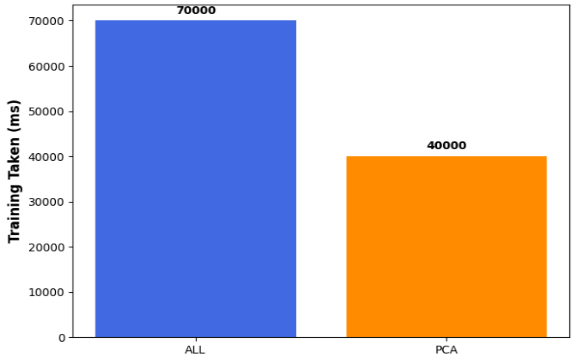

Fig 4: Before and After PCA

Figure 4 compares the training time (in milliseconds) taken with all features versus after applying PCA. It shows that PCA significantly reduces training time from 70,000 ms to 40,000 MS, demonstrating improved computational efficiency.

3.6 Data Splitting

The process of separating process of dividing Data splitting is the process of dividing a dataset into discrete subsets to make it easier to test, assess, and train machine learning models. 20% of the NSL-KDD dataset was used for testing, while the remaining 80% was used for training. The testing set was used to evaluate the model's performance after it had been built and trained using the training set.

3.7 Classification of Artificial Neural Networks

ANN is a machine learning system that learns from neuron connections like the human brain. It is based on the human nervous system. The learning network is made up of linked nodes, or neurons, grouped in three layers: input, hidden, and output. Because it assesses and aggregates node weights, the hidden layer, which connects the input and output layers' neurons, is critical to the learning process. Finally, the output layer displays the training phase's results. The weight of each node-to-node link reflects how important the input values are to the neuron. This weight is used throughout the learning process by the activation function, which calculates the input values and generates the resulting knowledge. Figure 5 displays the architecture of a typical neural network.

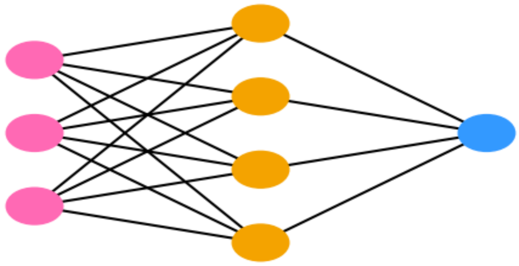

Fig 5: Structure of ANN with 3 Layers

An ANN gathers expertise by learning how network traffic behaves generally using historical data from previous projections as inputs. After then, it uses data from the current instant to anticipate traffic levels for the following time frame. ANNs are helpful for traffic prediction and pattern recognition in traffic data because of their excellent pattern recognition capability.

3.8 Performance Metrics

The learning network is a network of Three layers are made up of linked nodes, or neurons: input, hidden, and output. The hidden layer, which links the input and output layers' neurons, is crucial to learning since it assesses and aggregates node weights. Lastly, the output layer displays the outcome of the training process. The significance of the input values to the neuron is indicated by the weight of each node-to-node link. This weight is used throughout the learning process by the activation function, which computes the input values and produces the resulting knowledge. To compare model performance in network threat tasks, these metrics must be interpreted accurately.

Accuracy: Accuracy is the most evident metric. One may calculate this by dividing the number of accurate predictions by the number of occurrences and then multiplying the result by 100. The following Equation (2) is mentioned below:

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \times 100 \quad (2)$$

Precision: The TP to total positive predictions is used to check the positive predictions of the system and is defined as Equation (3):

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall: It assesses how well the model can distinguish insider threats in network traffic when they are indeed part of a fraudulent activity [14]. It reports on how many actual positives are correctly found by the model with the shown in Equation (4):

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1-score: The F1-measure, often known as the score, is a combination of two single metrics: the accuracy and recall harmonic mean, which is defined as Equation (5):

$$F1 - Score = 2\frac{(Precision*Recall)}{(Precision+ Reall)} \qquad (5)$$

These matrices are used to assess the model's efficacy for network threats using ML techniques for cybersecurity.

4. RESULT ANALYSIS AND DISCUSSION

DP models are evaluated for their ability to detect security flaws throughout a network using measures including precision, accuracy, F1-score, recall, and confusion matrix. If the collection of data with accurate labels, it may use a confusion matrix to see how many actual outcomes the model correctly predicted. The tabulation provides the TP, TN, FP, and FN numbers, which are useful for calculating these performance metrics. These figures together show how accurate the model is at recognizing attack traffic from ordinary traffic. A clear and accurate view of these metrics is essential for looking at the outcomes when using models for network threat detection.

Table 2. Outcome of ANN model on Network Threats

Measures	Artificial Neural Network
Accuracy	97.5
Precision	99.0
Recall	96.7
F1-score	95.7

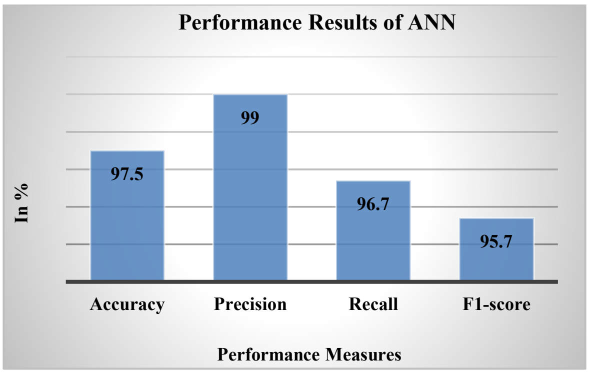

Fig 6: Performance results for Artificial Neural Network

The outcomes attained by the model suggested in this investigation are displayed in Figure 6 and Table 2. The 97.5% accuracy rate of the model suggests that it can correctly detect insider security threats. The extreme effectiveness of a precision score of 99.0% is based on guaranteeing that genuine traffic does not end up in the false positive bucket. With a 96.7% recall and a 95.7% F1-score, real harmful activity is correctly identified while demonstrating a strong balance between detection and accuracy. This indicates that the approach is consistently effective in identifying insider threats in dynamic networks.

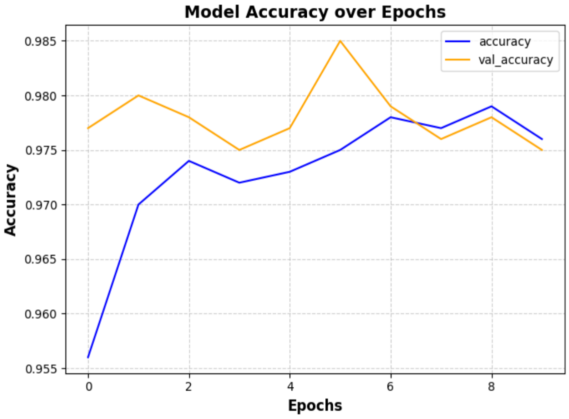

Fig 7: Accuracy validation Graph

The proposed ANN model’s training and validation accuracies are presented in Figure 7 after 10 epochs. It can see on the blue line that training accuracy steadily moves upward with only small variations, ending at about 97.5%. The orange line, which charts validation accuracy, corresponds closely to the trend of the training loss. The highest point at epoch 5 was 98.5%, showing that the model generalizes well. Although there are some minor deviations among epochs, both graphs stay fairly similar, which indicates that the model doesn’t overfit when performing both learning and generalization.

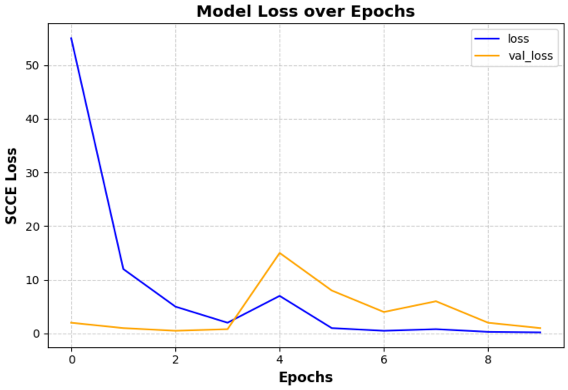

Fig 8: Loss vs. Validation Loss Graph

Figure 8 displays the findings of the suggested ANN model's training and validation loss curves using Sparse Categorical Cross-Entropy (SCCE) loss. The blue curve shows a sharp decline from above 50 to nearly zero within the first few epochs, and that must mean the training loss is still converging rapidly. The orange line, the validation loss, fluctuates a little but stays low throughout the run, with one short jump around epoch 4 and then levels off. The model is robust and generalizable in subsequent epochs when training and validation loss values closely correspond, indicating good learning with little overfitting.

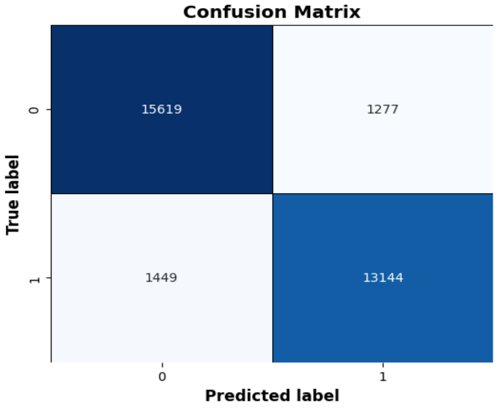

Fig 9: Confusion Matrix for ANN Model

Figure 9 displays the confusion matrix of the suggested ANN model's network traffic binary classification performance. The quantity of normal and attack cases that were accurately detected is considerable (15,619 TN and 13,144 TP, respectively). Finally, misclassifications where normal traffic was labeled as an attack or vice versa result in 1,277 FP, meaning normal traffic was labeled as an attack and 1,449 FN, meaning an attack was labeled as normal traffic. To demonstrate how well the model can differentiate between malicious and legitimate traffic, the overall distribution of values and demonstrated that it does so with relatively low error rates.

4.1 Comparative Analysis and Discussion

A comparison of network threats in terms of cybersecurity is presented here. Performance of different ML and DL models, such as forest [15], RF [16] and the ANN model, which is proposed in this thesis, is evaluated. Table 3 provides a detailed description of how the comparison is made from key performance metrics such as accuracy.

Table 3. Comparative analysis of Proposed and Existing models based on Network Threats

Models	Accuracy
Isolation forest [15]	89.54
Random Forest[16]	93
ANN	97.5

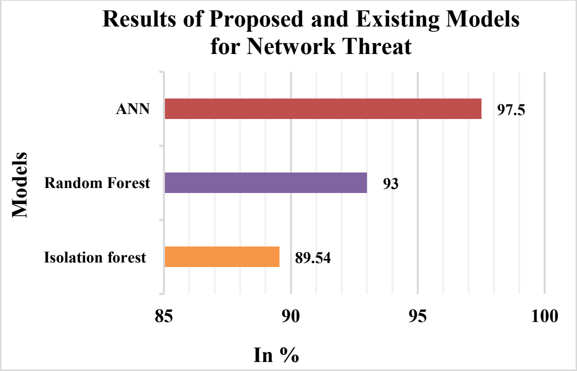

Fig 10: Results of Proposed and Existing Models for Network Threat

The results of each model performance comparison in Table 3 and Figure 10 show that ANN model accuracy is 97.5% which is higher than baseline models RF and forest with accuracy of 93% and 89.54% respectively. It demonstrates how an unsupervised anomaly detection method, Isolation Forest, fails to achieve high accuracy compared to the other methods, especially when detecting complex attack patterns in labeled datasets. Although more effective, RF uses an ensemble learning method which does not capture intricate, non-linear relationships in network traffic data as well as the ANN. This highlights that ANN's high accuracy emphasize its capability in detecting subtle threat signatures and is the winner of intrusion detection performance with respect to others with improvements 4.5% and 7.96% than RF and Isolation Forest, respectively.

The results of the proposed ANN model show a clear benefit in network threats with its capacity to recognize and depict intricate, nonlinear patterns in data about network traffic. because of its DL architecture, it is able to learn and understand subtle and sophisticated threat signatures beyond the reach of simple ML algorithms. While forest and RF are ridden by its own unsupervised nature and relies on the aggregation of decision tree ensembles, respectively, ANN is capable of modeling intricacies in its data with its multilayer structure better. Due to this, the ANN becomes a more powerful and reliable approach to detect myriad, evolving cybersecurity threats.

5. CONCLUSION AND FUTURE DIRECTION

In the modern cybersecurity landscape, where attacks are becoming increasingly targeted and complex, accurate detection of network intrusions is critical. Traditional methods often fail to cope with evolving threats, making data-driven and intelligent approaches essential. This study suggested an intrusion detection system that combines Principal Component Analysis (PCA) with Artificial Neural Networks (ANN), where PCA served as an effective preprocessing and dimensionality reduction technique, removing redundant features and enhancing learning efficiency. By focusing on the most relevant data attributes, the ANN achieved an accuracy of 97.5%, significantly outperforming conventional models such as Random Forest and Decision Tree ensembles. The superior performance of this integrated framework demonstrates the importance of combining robust dimensionality reduction with deep learning architectures for improved detection outcomes. Looking forward, the model will be extended to handle multi-class classification, enabling simultaneous detection and categorization of diverse attack types. Additionally, future research will explore advanced deep learning methods, Long Short-Term Memory (LSTM) networks and attention-based processes, among others, to improve the system's capacity to identify changing and sequential cyberattack patterns. Such advancements will allow seamless integration into real-time detection frameworks, ensuring more adaptive and resilient defenses against increasingly sophisticated cyber threats.

6. REFERENCES

[1] Kim, J. et al. (2019) 'Insider threat detection based on user behavior modeling and anomaly detection algorithms', Applied Sciences (Switzerland) [Preprint]. Available at: https://doi.org/10.3390/app9194018.

[2] Theis, M.C. et al. (2019) 'Common Sense Guide to Mitigating Insider Threats', CERT Division [Preprint].

Available at: https://doi.org/10.1184/R1/12363665.v1.

[3] Kolluri, V. (2016) ‘A Pioneering Approach To Forensic Insights: Utilization AI for Cybersecurity Incident Investigations’, International Journal of Research and Analytical Reviews, 3(3).

[4] Homoliak, I. et al. (2019) ‘Insight Into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures’, ACM Computing Surveys, 52(2), pp. 1–40. Available at: https://doi.org/10.1145/3303771.

[5] Toupas, P. et al. (2019) ‘An intrusion detection system for multi-class classification based on deep neural networks’, in Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019. Available at: https://doi.org/10.1109/ICMLA.2019.00206.

[6] Kolluri, V. (2018) ‘A Thorough Examination of Fortifying Cyber Defenses : AI in Real Time Driving Cyber Defence Strategies Today’, International Journal of Emerging Technologies and Innovative Research, 5(3).

[7] Kanimozhi, V. and Jacob, P. (2019) ‘Calibration Of Various Optimized Machine Learning Classifiers InNetwork Intrusion Detection System On The Realistic Cyber Dataset Cse-Cic-Ids2018 Using Cloud Computing’, International Journal of Engineering Applied Sciences and Technology, 04(06), pp. 209–213. Available at: https://doi.org/10.33564/IJEAST.2019.v04i06.036.

[8] Abideen, M.Z. ul, Saleem, S. and Ejaz, M. (2019) ‘VPN Traffic Detection in SSL-Protected Channel’, Security and Communication Networks, 2019(5), pp. 1–17. Available at: https://doi.org/10.1155/2019/7924690.

[9] Halimaa, A.A. and Sundarakantham, K. (2019) ‘Machine learning based intrusion detection system’, in Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019. Available at: https://doi.org/10.1109/ICOEI.2019.8862784.

[10] Chu, A., Lai, Y. and Liu, J. (2019) ‘Industrial Control Intrusion Detection Approach Based on Multiclassification GoogLeNet-LSTM Model’, Security and Communication Networks [Preprint]. Available at: https://doi.org/10.1155/2019/6757685.

[11] Begli, M., Derakhshan, F. and Karimipour, H. (2019) ‘A Layered Intrusion Detection System for Critical Infrastructure Using Machine Learning’, in Proceedings of 2019 the 7th International Conference on Smart Energy Grid Engineering, SEGE 2019. Available at: https://doi.org/10.1109/SEGE.2019.8859950.

[12] Srivastava, A., Agarwal, A. and Kaur, G. (2019) ‘Novel Machine Learning Technique for Intrusion Detection in Recent Network-based Attacks’, in 2019 4th International Conference on Information Systems and Computer Networks, ISCON 2019. Available at: https://doi.org/10.1109/ISCON47742.2019.9036172.

[13] Kim, D.W., Hong, S.S. and Han, M.M. (2018) ‘A study on classification of insider threat using markov chain model’, KSII Transactions on Internet and Information Systems [Preprint]. Available at: https://doi.org/10.3837/tiis.2018.04.027.

[14] Farnaaz, N. and Jabbar, M.A. (2016) ‘Random Forest Modeling for Network Intrusion Detection System’, in Procedia Computer Science. Available at: https://doi.org/10.1016/j.procs.2016.06.047.

[15] Aldairi, M., Karimi, L. and Joshi, J. (2019) ‘A trust aware unsupervised learning approach for insider threat detection’, in Proceedings - 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science, IRI 2019. Available at: https://doi.org/10.1109/IRI.2019.00027.

[16] Lee, J. et al. (2019) ‘Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles’, IEEE Access [Preprint]. Available at: https://doi.org/10.1109/ACCESS.2019.2953095.

[17] Kalla, D., Smith, N., & Samaah, F. (2023). Satellite Image Processing Using Azure Databricks and Residual Neural Network. International Journal of Advanced Trends in Computer Applications, 9(2), 48-55.

[18] Kuraku, D. S., & Kalla, D. (2023). Phishing Website URL’s Detection Using NLP and Machine Learning Techniques. Journal on Artificial Intelligence-Tech Science.

[19] Varadharajan, V., Smith, N., Kalla, D., Samaah, F., Polimetla, K., & Kumar, G. R. (2024). Stock Closing Price and Trend Prediction with LSTM-RNN. Journal of Artificial Intelligence and Big Data, 4, 877.

[20] Kalla, D., Kuraku, D. S., & Samaah, F. (2021). Enhancing cyber security by predicting malwares using supervised machine learning models. International Journal of Computing and Artificial Intelligence, 2(2), 55-62.

[21] Kuraku, D. S., Kalla, D., Smith, N., & Samaah, F. (2023). Safeguarding FinTech: elevating employee cybersecurity awareness in financial sector. International Journal of Applied Information Systems (IJAIS), 12(42).

[22] Kalla, D., Smith, N., Samaah, F., & Polimetla, K. (2022). Enhancing Early Diagnosis: Machine Learning Applications in Diabetes Prediction. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-205. DOI: doi. org/10.47363/JAICC/2022 (1), 191, 2-7.

[23] Kuraku, S., Kalla, D., Samaah, F., & Smith, N. (2023). Cultivating proactive cybersecurity culture among IT professional to combat evolving threats. International Journal of Electrical, Electronics and Computers, 8(6).

[24] Kalla, D., Smith, N., Samaah, F., & Polimetla, K. (2024). Hybrid Scalable Researcher Recommendation System Using Azure Data Lake Analytics. Journal of Data Analysis and Information Processing, 12, 76-88.

[25] Kuraku, D. S., & Kalla, D. (2023). Impact of phishing on users with different online browsing hours and spending habits. International Journal of Advanced Research in Computer and Communication Engineering, 12(10).

[26] Kalla, D., & Kuraku, S. (2023). Phishing website url’s detection using nlp and machine learning techniques. Journal of Artificial Intelligence, 5, 145.

[27] Kuraku, D. S., Kalla, D., & Samaah, F. (2022). Navigating the link between internet user attitudes and cybersecurity awareness in the era of phishing challenges. International

Advanced Research Journal in Science, Engineering and Technology, 9(12).

[28] Kuraku, D. S., Kalla, D., Smith, N., & Samaah, F. (2023). Exploring How User Behavior Shapes Cybersecurity Awareness in the Face of Phishing Attacks. International Journal of Computer Trends and Technology.

[29] Sreeramulu, M. D., Mohammed, A. S., Kalla, D., Boddapati, N., & Natarajan, Y. (2024, September). AI-driven Dynamic Workload Balancing for Real-time Applications on Cloud Infrastructure. In 2024 7th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 7, pp. 1660-1665). IEEE.

[30] Kalla, D., Mohammed, A. S., Boddapati, V. N., Jiwani, N., & Kiruthiga, T. (2024, November). Investigating the Impact of Heuristic Algorithms on Cyberthreat Detection. In 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT) (Vol. 1, pp. 450-455). IEEE.

[31] Chandrasekaran, A., & Kalla, D. (2023). Heart disease prediction using chi-square test and linear regression. Computer Science & Information Technology, 13, 135-146.

[32] Chinta, P. C. R., Katnapally, N., Ja, K., Bodepudi, V., Babu, S., & Boppana, M. S. (2022). Exploring the role of neural networks in big data-driven ERP systems for proactive cybersecurity management. Kurdish Studies.

[33] Routhu, K., Bodepudi, V., Jha, K. M., & Chinta, P. C. R. (2020). A Deep Learning Architectures for Enhancing Cyber Security Protocols in Big Data Integrated ERP Systems. Available at SSRN 5102662.

[34] Moore, C. (2023). AI-powered big data and ERP systems for autonomous detection of cybersecurity vulnerabilities. Nanotechnology Perceptions, 19, 46-64.

[35] Bodepudi, V., & Chinta, P. C. R. (2024). Enhancing Financial Predictions Based on Bitcoin Prices using Big Data and Deep Learning Approach. Available at SSRN 5112132.

[36] Chinta, P. C. R. (2023). The Art of Business Analysis in Information Management Projects: Best Practices and Insights. DOI, 10.

[37] Boppana, S. B., Moore, C. S., Bodepudi, V., Jha, K. M., Maka, S. R., & Sadaram, G. AI And ML Applications In Big Data Analytics: Transforming ERP Security Models For Modern Enterprises.

[38] Katnapally, N., Chinta, P. C. R., Routhu, K. K., Velaga, V., Bodepudi, V., & Karaka, L. M. (2021). Leveraging Big Data Analytics and Machine Learning Techniques for Sentiment Analysis of Amazon Product Reviews in Business Insights. American Journal of Computing and Engineering, 4(2), 35-51.

[39] Chinta, P. C. R., Moore, C. S., Karaka, L. M., Sakuru, M., Bodepudi, V., & Maka, S. R. (2025). Building an Intelligent Phishing Email Detection System Using Machine Learning and Feature Engineering. European Journal of Applied Science, Engineering and Technology, 3(2), 41-54.

[40] Moore, C. (2024). Enhancing Network Security With Artificial Intelligence Based Traffic Anomaly Detection In Big Data Systems. Available at SSRN 5103209.

[41] Chinta, P. C. R., Moore, C. S., Karaka, L. M., Sakuru, M., & Bodepudi, V. (2025). Predictive Analytics for Disease Diagnosis: A Study on Healthcare Data with Machine Learning Algorithms and Big Data. J Cancer Sci, 10(1), 1.

[42] KishanKumar Routhu, A. D. P. Risk Management in Enterprise Merger and Acquisition (M&A): A Review of Approaches and Best Practices.

[43] Bodepudi, V. (2023). Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights. Journal of Artificial Intelligence and Big Data, 3(1), 10-31586.

[44] Chinta, P. C. R. (2022). Enhancing Supply Chain Efficiency and Performance Through ERP Optimisation Strategies. Journal of Artificial Intelligence & Cloud Computing, 1(4), 10-47363.

[45] Krishna Madhav, J., Varun, B., Niharika, K., Srinivasa Rao, M., & Laxmana Murthy, K. (2023). Optimising Sales Forecasts in ERP Systems Using Machine Learning and Predictive Analytics. J Contemp Edu Theo Artific Intel: JCETAI-104.

[46] Jha, K. M., Velaga, V., Routhu, K., Sadaram, G., Boppana, S. B., & Katnapally, N. (2025). Transforming Supply Chain Performance Based on Electronic Data Interchange (EDI) Integration: A Detailed Analysis. European Journal of Applied Science, Engineering and Technology, 3(2), 25-40.

[47] Sadaram, G., Sakuru, M., Karaka, L. M., Reddy, M. S., Bodepudi, V., Boppana, S. B., & Maka, S. R. (2022). Internet of Things (IoT) Cybersecurity Enhancement through Artificial Intelligence: A Study on Intrusion Detection Systems. Universal Library of Engineering Technology, (2022).

[48] Maka, S. R. (2023). Understanding the Fundamentals of Digital Transformation in Financial Services: Drivers and Strategic Insights. Available at SSRN 5116707.