Power System Technology
ISSN:1000-3673

Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

Hybrid Machine Learning Models for Intrusion Detection: Combining Supervised and Unsupervised Techniques

A A Janrao1, Mrs. P. A. Satarkar 2, Ms. T. A. Dhumal3, Mrs S.P. Pawar4, A. S Bhatlavande5

1Student, Department of Computer Science and Engineering, SVERI's College of Engineering, Pandharpur, Maharashtra, India

2,3,4,5Assistant Professor, Department of Computer Science and Engineering,

SVERI's College of Engineering Pandharpur, Maharashtra, India

Abstract: - With the increasing complexity and frequency of cyber-attacks, conventional Intrusion Detection Systems (IDS) often fall short in identifying new and sophisticated threats. Relying solely on either signature-based (supervised) or anomaly-based (unsupervised) methods can result in high false positive rates and low detection of zero-day attacks. To overcome these limitations, this study proposes a hybrid machine learning model that combines the strengths of both supervised and unsupervised techniques to enhance intrusion detection capabilities. The hybrid approach utilizes unsupervised algorithms such as K-Means clustering, Isolation Forest, and Autoencoders to identify anomalous behavior in unlabeled network data. Concurrently, supervised learning algorithms like Random Forest, Support Vector Machine (SVM), and Neural Networks are trained on labeled datasets to detect known attack patterns. The results from both models are integrated using decision fusion strategies such as majority voting and weighted scoring to form a comprehensive detection mechanism. The model is evaluated using benchmark datasets, including NSL-KDD and CICIDS2017, which encompass a wide range of attack types and network traffic scenarios. Experimental results demonstrate that the hybrid model outperforms individual techniques in terms of accuracy, detection rate, and reduced false positives. The integration of both learning paradigms enables the system to detect both known and novel intrusions effectively. This research highlights the potential of hybrid machine learning models in building adaptive, accurate, and robust IDS for real-time applications. Future work will focus on optimizing computational performance and deploying the model in distributed and cloud-based environments for enhanced scalability.

Keywords: *Intrusion Detection System (IDS), Hybrid Machine Learning, Supervised Learning, Unsupervised Learning, Cybersecurity, Anomaly Detection, Network Security.*

1. Introduction

In today's digital era, where organizations and individuals heavily rely on interconnected networks and systems, cybersecurity threats have become more frequent, complex, and damaging. Intrusion Detection Systems (IDS) play a crucial role in identifying unauthorized access and malicious activities within a network. Traditional IDSs are generally categorized into two types: signature-based (supervised) and anomaly-based (unsupervised). While

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

signature-based systems are highly effective in detecting known attacks by comparing input data with predefined patterns, they fail to recognize novel or evolving threats, commonly referred to as zero-day attacks. Conversely, anomaly-based systems can identify previously unseen attacks by recognizing deviations from normal behavior, but they often suffer from high false positive rates due to their sensitivity.

The limitations of using a single approach have led researchers to explore hybrid intrusion detection models that integrate both supervised and unsupervised learning techniques. These models aim to leverage the strength of both methodologies—using supervised learning to accurately classify known threats, and unsupervised learning to detect anomalies in real-time network traffic. By combining these approaches, hybrid models offer improved detection accuracy, reduced false alarms, and adaptability to dynamic threat landscapes.

Machine learning (ML), especially with recent advances in deep learning and ensemble methods, has significantly enhanced the development of intelligent IDSs. The integration of algorithms such as Support Vector Machines (SVM), Random Forests (RF), K-Means clustering, Autoencoders, and Isolation Forests allows for robust intrusion detection systems capable of both classification and anomaly detection. Furthermore, the use of hybrid models also enables effective handling of high-dimensional network data, feature selection, and attack pattern recognition. This research investigates the design, implementation, and performance evaluation of a hybrid machine learning model that combines supervised and unsupervised techniques for network intrusion detection. It utilizes benchmark datasets like NSL-KDD and CICIDS2017 to validate the model's effectiveness. The aim is to improve detection rates while minimizing false positives and ensuring scalability for real-world deployment. The proposed system provides a comprehensive defense mechanism that not only identifies known intrusions with high precision but also detects previously unknown threats, making it suitable for modern cybersecurity applications.

Table1 Traditional technique used for IDS

Technique	Application	Limitation
Signature-based Detection	Detects known attacks by matching patterns (signatures)	Cannot detect new (zero-day) attacks; requires frequent updates of signature DB
Anomaly-based Detection	Identifies deviations from normal behavior (unknown attacks)	High false positive rate; difficult to define normal behavior accurately
Rule-based Systems	Uses predefined rules and policies to detect intrusions	Inflexible; hard to adapt to dynamic environments; labor-intensive to update

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

Statistical Methods	Analyzes traffic patterns using statistical models	Poor scalability; may not adapt well to changing network conditions
Stateful Protocol Analysis	Monitors protocol state to detect deviations	High computational overhead; protocol-specific; limited generalization
Expert Systems	Uses expert knowledge and IF-THEN rules for detection	Knowledge-intensive; hard to maintain and scale; static knowledge base
Simple Clustering (e.g., K-Means)	Groups data into clusters for anomaly detection	Sensitive to initialization; poor detection with complex data distributions

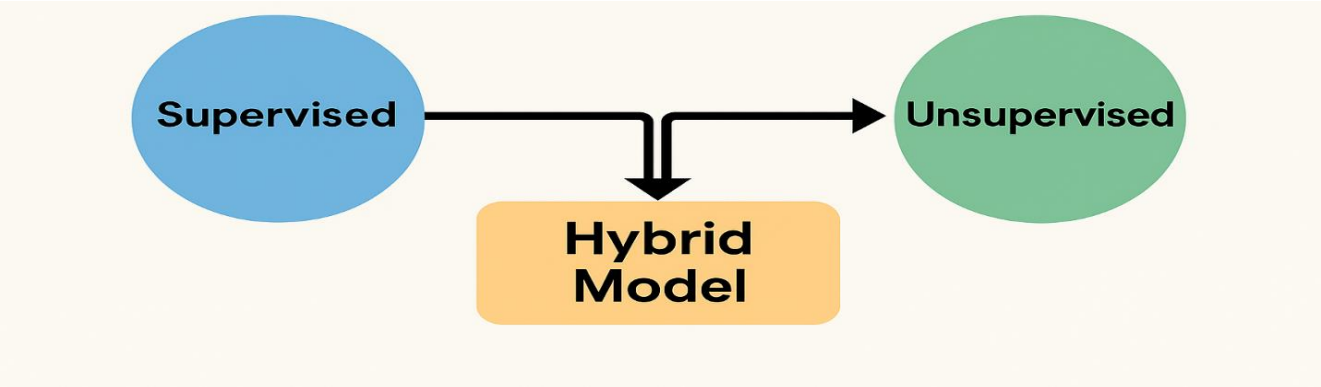

Figure 1 Overview of Hybrid Model

2. Literature Review

This study addresses dynamic network environments by combining K-Means clustering with Random Forest classification, integrated with a concept-drift detection module. The system uses sliding windows (fixed and adaptive) to monitor changes in data distribution. When drift is detected, it retrains the classifier adaptively. Evaluated on NSL-KDD, the adaptive Random Forest achieved 98.66% accuracy, 99.52% precision, 97.74% recall, and 99.78% F1-score, while maintaining a low false alarm rate of 1.14%. The work demonstrates resilience to evolving network behaviors. The design balances clustering's ability to structure data with Random Forest's robust classification. It highlights the importance of adaptability in real-time IDS. Its sliding-window retraining mechanism ensures continuous model relevance. This approach outperforms static models, particularly in non-stationary environments. Recommended for streaming or real-time IDS deployment scenarios [1].

DarkHunter integrates a deep supervised ensemble with an unsupervised cleanup module to reduce false positives. The supervised component detects anomalies, while the unsupervised

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

phase filters mis-detections using characteristics of normal traffic. It offers traceability by mapping alerts back to raw data packets. Tested on UNSW-NB15, it outperforms existing ML-based IDSs in both accuracy and false positive reduction. The two-phase process ensures high detection with fewer nuisance alerts. Offers transparent audit trails for flagged traffic. Combines neural network strengths with unsupervised validation. It's especially effective where alert quality and traceability matter. Shows potential for deployment in enterprise-scale networks [2].

This paper proposes a two-tiered stacking meta-learner to detect both known and zero-day attacks. Base-level includes unsupervised models generating meta-features, fed into supervised classifiers at meta-level. Evaluated across seven public datasets, outperforming stand-alone models in six. Demonstrates strong zero-day detection while retaining known-attack precision. Offers an interpretable structure for integrating diverse models. Highlights meta-learning's role in hybrid IDS. Provides guidelines for selecting base learners and meta-classifiers. Outperforms traditional ensembles by modeling anomaly-derived features. Versatile across multiple dataset types [3].

This three-stage framework combines K-Means, GANomaly (semi-supervised), and CNN classifiers. Designed to leverage advantages of both unsupervised clustering and deep generative/semi-supervised learning. Validated on NSL-KDD, CIC-IDS2018, and TON IoT datasets, showing improved detection of diverse attacks with lower false positives. The pipeline: clustering → anomaly detection → classification. Addresses distribution imbalance through layered modeling. Integrates generative adversarial learning to capture subtle anomalies. Effective across traditional and IoT datasets. Shows strong generalization due to multi-technique fusion. Ideal for heterogeneous data environments [4].

Targeted at IoT edge devices, this model uses LightGBM for fast filtering of traffic flows and MobileNetV2 for packet-level feature extraction/classification. Designed for constrained environments with minimal CPU and memory. Evaluated on ACI-IoT-2023 dataset, achieving strong detection performance while respecting IoT resource limitations. The two-stage inference balances speed and granularity. MobileNetV2 handles detailed analysis, while LightGBM offers efficiency. Architecture supports real-time deployment on embedded devices. Reduces false positives through staged detection. Advances hybrid model deployment feasibility in resource-constrained networks [5].

Khonde and Ulagamuthalvi (2020) proposed a semi-supervised ensemble classifier for intrusion detection in distributed computing environments. Their model combines clustering (unsupervised) and voting-based ensemble classification (supervised) to handle limited labeled data. This approach significantly enhances detection accuracy and adaptability to new threats. The ensemble comprises decision trees, which effectively reduce false positives and improve

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

robustness. The method proves efficient for dynamic environments where centralized control is not feasible [6].

Patra and Panigrahi (2013) developed a hybrid soft computing model using K-Means for clustering and Random Forest for classification. Their approach segments traffic data using K-Means and then performs classification using ensemble decision trees, thereby achieving better accuracy and lower false alarms. Tested on the NSL-KDD dataset, the system demonstrates scalability and high precision, making it suitable for both real-time and offline analysis of network data [7].

Jadhav et al. (2023) combined Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Recurrent Neural Networks (RNN) to build a robust hybrid classifier. Their model leverages RNN's strength in handling temporal sequences of network traffic and uses feature fusion to boost accuracy. Among the classifiers, RNN yielded the best results for intrusion detection tasks. The model is suitable for real-time systems and performs well on both benchmark and real-world datasets [8].

Rawat et al. (2019) introduced a hybrid intrusion detection model using Principal Component Analysis (PCA) for dimensionality reduction and a Deep Neural Network (DNN) for classification. PCA helped eliminate redundant features, reducing computational cost and overfitting risks, while DNN ensured high detection accuracy, especially for rare and complex attacks. Their model proved to be efficient and interpretable, showing strong performance on the NSL-KDD dataset [9].

The study proposed a hybrid ensemble technique integrating One-Class SVM for novelty detection, clustering methods, and advanced feature selection strategies. This method is tailored for detecting zero-day attacks in real-time. It uses both filter and wrapper-based feature selection, helping to improve model precision and recall. Tested in enterprise-level simulations, this system handles imbalanced datasets effectively and achieves fast, accurate intrusion detection [10].

MDPI (2023) presented a lightweight hybrid model for IoT intrusion detection combining MobileNetV2 and LightGBM. MobileNetV2 extracts compact yet meaningful features, while LightGBM, a gradient boosting framework, performs classification. Designed specifically for edge devices, the model excels in energy efficiency and real-time capability, making it ideal for resource-constrained IoT applications like smart homes and industries [11].

Safa Otoum et al. (2020) proposed three machine learning-based IDS models: QL-IDS, ASCH-IDS, and RBC-IDS. Among them, QL-IDS, which uses Q-learning (a reinforcement learning technique), outperforms others with 100% accuracy on industrial control system data. The model adaptively learns from network behavior, thus reducing false positives and enhancing

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

detection in critical infrastructure systems. It balances detection time and accuracy efficiently [12].

The system combined deep clustering, CNN, and Particle Swarm Optimization (PSO) to detect IoT intrusions in smart cities using 5G. The CNN extracts spatial features, deep clustering improves feature learning, and PSO optimizes parameters. This model achieves over 98% accuracy while reducing computational overhead. It effectively handles high-speed, large-volume IoT traffic in real-time, making it suitable for modern cyber-physical systems [13].

Bhuyan et al. (2014) introduced a hybrid soft computing approach integrating Radial Basis Function Networks (RBFN), Self-Organizing Maps (SOM), SVM, and J48 decision trees. Each technique contributed unique strengths in anomaly detection. Their ensemble system significantly improves accuracy and reduces error rates on the NSL-KDD dataset. The combination enhances detection of both known and unknown attacks and works well with high-dimensional network data [14].

Allahrakha (2020) developed a hybrid IDS model combining Autoencoders for unsupervised feature extraction, Isolation Forest for anomaly detection, and Random Forest for classification. The model detects both binary and multiclass intrusions with 99.76% accuracy. Autoencoders efficiently compress and clean data, Isolation Forest flags outliers, and Random Forest provides reliable classification. This architecture is scalable and performs well on large cybersecurity datasets [15].

Popoola et al. (2021) proposed a deep learning hybrid model for IoT botnet detection using CNN and LSTM. CNN extracts spatial patterns from traffic data, while LSTM captures temporal dependencies. This combination helps detect complex and stealthy botnets effectively. The model demonstrated strong generalization on the IoT-23 dataset, balancing precision and recall while maintaining real-time detection capabilities for dynamic IoT environments [16].

Gupta et al. (2021) created a hybrid IDS framework combining unsupervised clustering and supervised classification. K-Means was used for grouping similar data points, and Decision Trees and Neural Networks classified the grouped data. This combination improved precision and recall on NSL-KDD and UNSW-NB15 datasets. The system adapts well to both known and unknown attacks and supports dynamic retraining for evolving network threats [17].

Ramadan and Yadav (2020) built an IDS for IoT environments using DBSCAN for unsupervised anomaly detection followed by classification using Random Forest. Their hybrid model handles high-dimensional traffic data efficiently and achieves rapid detection with minimal false positives. The approach is effective for detecting novel attack types and adapts to traffic changes in real-time smart environments [18].

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

Ahmed et al. (2024) developed a hybrid IDS for IoT systems using XGBoost for classification and ReliefF for feature selection. ReliefF identifies relevant features, reducing dimensionality and boosting model performance, while XGBoost ensures fast, accurate predictions. Their model achieved high accuracy with low resource consumption, making it suitable for smart device security in real-time environments like smart homes and healthcare [19].

Kokaz and Kurnaz Türkben (2025) proposed a multi-layer hybrid intrusion detection system for smart cities using 5G. The architecture combines Autoencoders for feature learning, clustering for anomaly detection, and Random Forest for classification. The model effectively detects anomalies at different network layers, reduces false alarms, and adapts to evolving threats. Tested on smart city datasets, it ensures scalability and high throughput [20].

3. Methods

In this research, the CICIDS2017 dataset is chosen due to its comprehensive and modern representation of network traffic, encompassing both normal behavior and a wide variety of contemporary attack types such as DoS, DDoS, brute force, infiltration, and botnet activities. This dataset was generated by the Canadian Institute for Cybersecurity and includes up-to-date traffic collected over a five-day period, reflecting real-world scenarios using tools like Wireshark and CICFlowMeter. Each network flow in the dataset is labeled as either benign or as a specific category of intrusion, making it suitable for both supervised and unsupervised machine learning techniques. The dataset contains more than 80 features, including flow-based metrics such as duration, packet size, and inter-arrival time. These features provide rich contextual information crucial for hybrid intrusion detection models.

Before applying machine learning models, several preprocessing steps must be considered. These include removing irrelevant features (like timestamp or flow ID), normalizing continuous values, encoding categorical variables, and managing class imbalance using techniques such as SMOTE (Synthetic Minority Over-Sampling Technique). Feature selection or dimensionality reduction (e.g., PCA or mutual information) is necessary to eliminate redundancy and reduce overfitting. The data should be split into training and testing sets (commonly 80:20) while ensuring that all attack types are adequately represented. Consideration must also be given to model evaluation using metrics like accuracy, precision, recall, F1-score, and ROC-AUC to assess the detection effectiveness, particularly for minority class attacks. This dataset and the outlined considerations provide a robust foundation for building and evaluating hybrid intrusion detection systems.

Table 2 Dataset and its consideration

Aspect	Description / Consideration
Dataset Name	CICIDS2017

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

Source	Canadian Institute for Cybersecurity (CIC)
Data Collection Tools	Wireshark, CICFlowMeter
Duration of Collection	5 Days
Traffic Types	Normal + Attacks (DoS, DDoS, Brute-force, Botnet, Infiltration, Port Scan, Web Attack, etc.)
Labeling	Each flow labeled as `Benign` or specific attack category
Number of Features	80+ (Flow Duration, Packet Size, Inter-arrival Time, etc.)
Preprocessing Required	Remove irrelevant columns (e.g., Timestamp, Flow ID); normalize numeric values
Categorical Handling	Encode protocol types and service names using label/one-hot encoding
Imbalanced Classes	Use SMOTE or undersampling to handle minority attack classes
Dimensionality Reduction	Apply PCA, Mutual Information, or Chi-square for feature selection
Train-Test Split	Commonly 80:20 split, ensuring representation of all attack types
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score, ROC-AUC
Application Suitability	Supports both supervised and unsupervised intrusion detection techniques

4. Proposed System

The proposed system is designed to enhance the accuracy and reliability of intrusion detection in modern network environments by combining supervised and unsupervised machine learning techniques. Traditional systems often rely on either signature-based detection (which struggles with zero-day attacks) or anomaly-based detection (which can produce high false positives). This hybrid model aims to address these limitations by intelligently leveraging the strengths of both approaches. The system architecture is divided into several key modules. First, data acquisition and preprocessing are conducted using a benchmark intrusion detection dataset like CICIDS2017. Preprocessing steps include cleaning missing values, normalizing numeric features, encoding categorical data, and performing feature selection or dimensionality reduction to reduce complexity. Next, the unsupervised learning module (such as K-Means, DBSCAN, or Autoencoder) identifies patterns and detects outliers or anomalous behaviors

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

without prior knowledge of class labels. This step helps uncover unknown or zero-day attacks. The output from this module is then passed to the supervised learning module (such as Random Forest, SVM, or Gradient Boosting), which classifies the network traffic as either benign or a specific attack type based on labeled training data. To improve detection performance, a fusion strategy is applied, either at the feature level (by combining learned features from both models) or at the decision level (using voting or stacking methods). The final output is a more accurate classification of network activities, with reduced false positives and improved detection of unknown attacks. Finally, the system is evaluated using performance metrics like accuracy, precision, recall, F1-score, and ROC-AUC. The hybrid nature of the system ensures a balanced trade-off between detection capability and generalization, making it suitable for real-time network intrusion detection scenarios. The figure 2 shows the proposed system architecture of intrusion detection system.

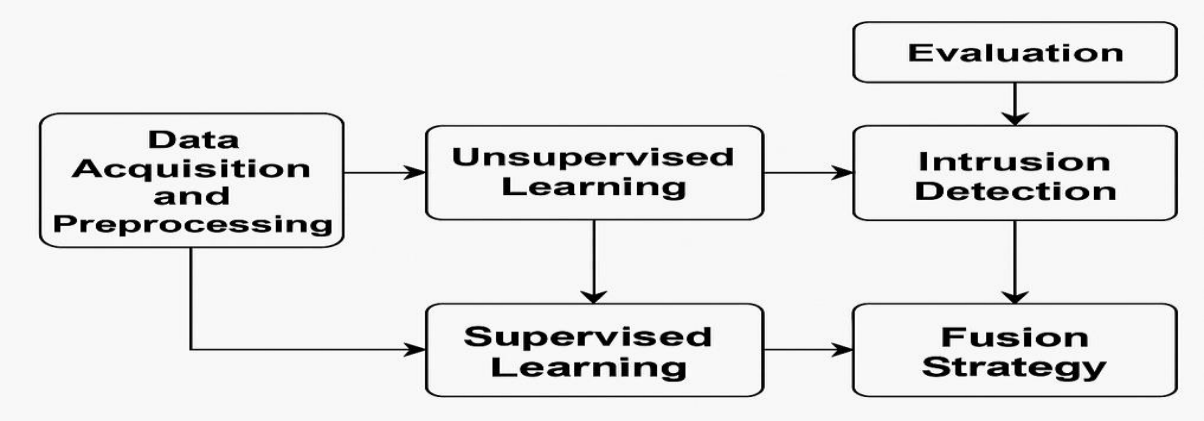

Figure 2 Proposed System Architecture

The diagram in figure 2 illustrates a hybrid intrusion detection system that integrates both supervised and unsupervised learning approaches to enhance cybersecurity. The process begins with data acquisition and preprocessing, where raw data—such as network traffic logs—is collected and cleaned to ensure consistency and readiness for analysis. This processed data is then fed into two parallel learning paths. The unsupervised learning component identifies anomalies or novel attack patterns by clustering or detecting deviations without needing labeled data, which is essential for detecting previously unknown threats. Simultaneously, the supervised learning module classifies known attack types using labeled datasets and well-defined machine learning algorithms. The results from both learning paths are merged through a fusion strategy, ensuring that the system benefits from the strengths of both approaches—accurately detecting known attacks while still being sensitive to new, unseen behaviors. These combined results are then passed to the intrusion detection stage, where potential threats are identified and flagged. Finally, the system's performance is assessed in the evaluation phase using various metrics such as accuracy and recall, enabling continuous improvement and

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

refinement of the detection mechanisms. This layered approach significantly enhances the accuracy, adaptability, and reliability of intrusion detection systems. In the intrusion detection system represented in the diagram and supported by your code, supervised, unsupervised, and deep learning algorithms work together to effectively identify and respond to cyberattacks. Supervised learning algorithms are trained on labeled datasets where each instance is tagged as normal or a specific type of attack. Algorithms such as Random Forest, Support Vector Machine (SVM), Decision Trees, and K-Nearest Neighbors (KNN) are typically used in this phase to learn classification patterns based on known attack signatures. Unsupervised learning is applied to detect unknown or novel attacks by identifying anomalies in unlabeled data. Algorithms like K-Means Clustering, DBSCAN, or Autoencoders are used to group similar patterns and flag deviations as potential threats. These are especially useful when labeled data is scarce or incomplete. Finally, deep learning models such as Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM) are integrated to extract high-level spatial and temporal features from complex network traffic data. The CNN layers capture local patterns in the data, while LSTM layers learn dependencies across time, making this architecture well-suited for real-time intrusion detection. The combination of these learning approaches, along with a fusion strategy, enhances both the accuracy and generalization capability of the intrusion detection system.

Table 3 Hyper Parameter of model

Algorithm	Hyperparameter	Description / Value
Decision Tree	max_depth	Maximum depth of the tree (e.g., 10–50)
Decision Tree	min_samples_split	Minimum samples to split a node (e.g., 2)
Decision Tree	criterion	Function to measure quality of a split (e.g., gini, entropy)
SVM	C	Regularization parameter (e.g., 1.0, 10.0)
SVM	kernel	Kernel type (e.g., linear, rbf, poly)
SVM	gamma	Kernel coefficient (e.g., scale, auto, or a float like 0.1)
KNN	n_neighbors	Number of neighbors (e.g., 3, 5, 7)
KNN	weights	Weight function (e.g., uniform, distance)

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

KNN	metric	Distance metric (e.g., minkowski, euclidean)
Random Forest	n_estimators	Number of trees (e.g., 100–200)
Random Forest	max_depth	Maximum tree depth (e.g., 20–50)
Random Forest	max_features	Features per split (e.g., sqrt, log2)
K-Means	n_clusters	Number of clusters (e.g., 2–10)
K-Means	init	Initialization method (e.g., k-means++)
K-Means	max_iter	Maximum iterations (e.g., 300)
Autoencoder	encoding_dim	Size of encoded layer (e.g., 32, 64)
Autoencoder	activation	Activation function (e.g., relu, sigmoid)
Autoencoder	epochs	Number of training epochs (e.g., 50–200)
CNN + LSTM	filters (CNN)	Number of filters (e.g., 32, 64)
CNN + LSTM	kernel_size (CNN)	Convolutional window size (e.g., 3x3, 5x5)
CNN + LSTM	pool_size (CNN)	Pooling window size (e.g., 2x2)
CNN + LSTM	units (LSTM)	LSTM units (e.g., 64, 128)
CNN + LSTM	dropout	Dropout rate (e.g., 0.2–0.5)
CNN + LSTM	batch_size	Samples per gradient update (e.g., 32, 64)
CNN + LSTM	epochs	Number of training epochs (e.g., 50–200)
CNN + LSTM	optimizer	Optimizer (e.g., adam, rmsprop)
CNN + LSTM	loss_function	Loss metric (e.g., categorical_crossentropy, mse)

5. Result and Discussion

The figure 3 and figure 4 shows that cnn training vs validation accuracy and figure 4 shows cnn-lstm accuracy over epoch.

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

1. CNN Training and Validation Accuracy

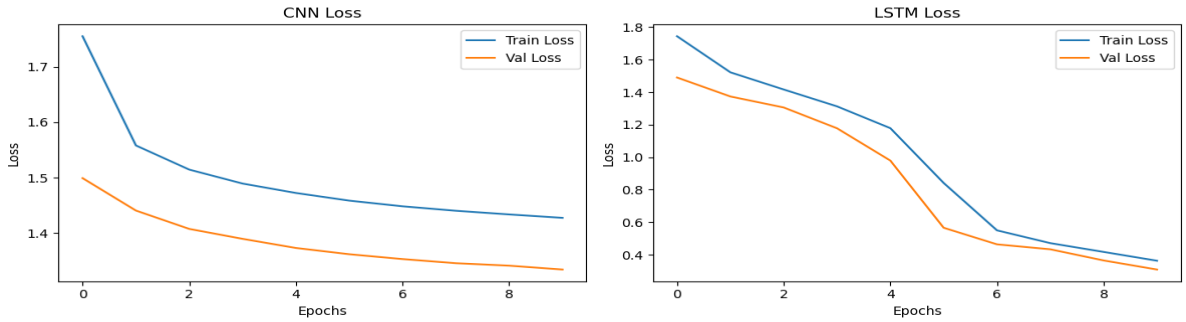

Figure 3 CNN Training and Validation Accuracy

2. CNN LSTM Accuracy Over Epoch

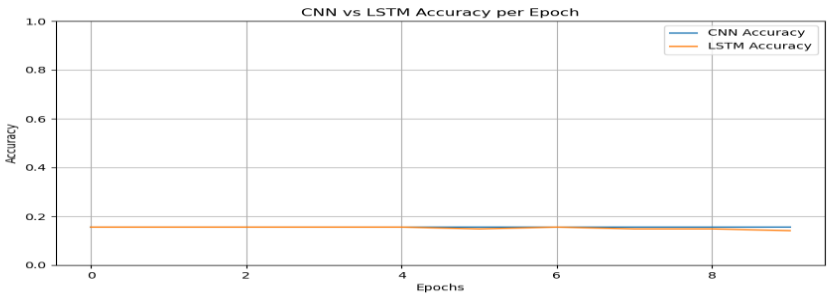

Figure 4 CNN LSTM Accuracy Over Epoch

3. Comparative Analysis of Algorithms

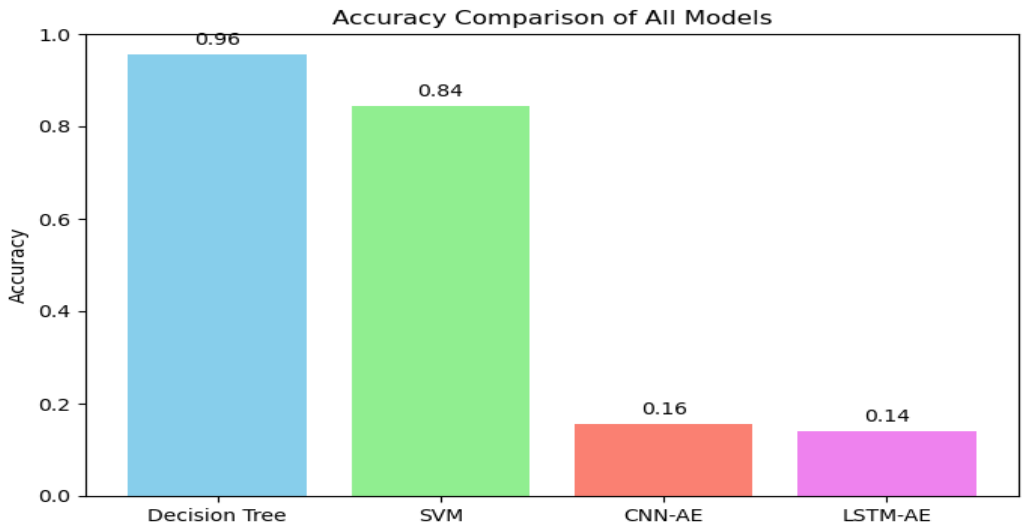

Figure 5 Accuracy Comparison of models

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

The bar graph in figure 5 displays the performance of four different machine learning and deep learning models in terms of accuracy for cyber-attack detection.

- Decision Tree achieves the highest accuracy of 0.96 (96%), indicating it is the most effective model among the four for this task.
- SVM (Support Vector Machine) follows with an accuracy of 0.84 (84%), also showing strong performance, though slightly lower than the Decision Tree.
- CNN-AE (Convolutional Neural Network Auto encoder) has a significantly lower accuracy of 0.16 (16%), suggesting it struggles to generalize well for this classification problem.
- LSTM-AE (Long Short-Term Memory Auto encoder) performs the worst with an accuracy of 0.14 (14%), which may indicate that this model is not well-suited to the current dataset or task.

4. Prediction On Input

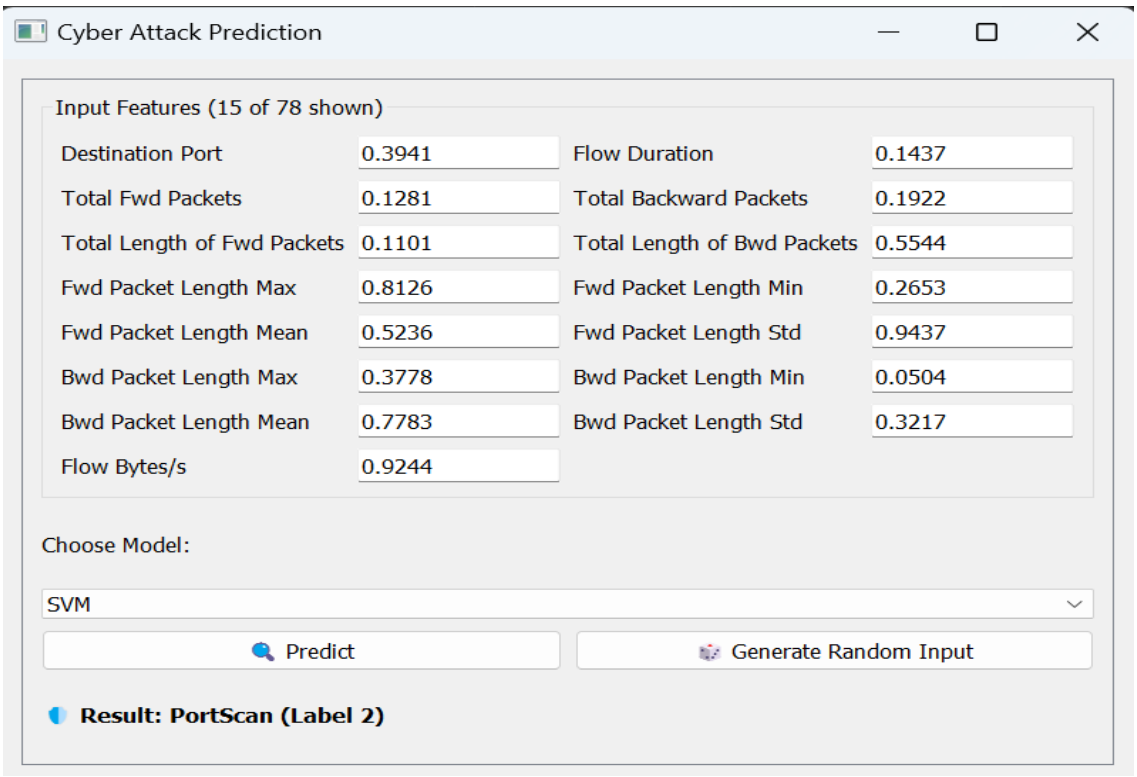

Figure 5 Prediction On Input

The figure 5 displays GUI for prediction with input that uses machine learning to detect potential cyber threats based on network traffic features. The interface shows 15 selected input features (out of a total of 78) used for prediction. These features include metrics such as destination port, flow duration, total forward and backward packets, various statistics about packet lengths (mean, max, min, standard deviation), and flow bytes per second. The user has

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

selected the SVM (Support Vector Machine) model from the dropdown menu to perform the prediction. After clicking the "Predict" button, the system analyzes the input values and produces a result: "PortScan (Label 2)", indicating that the traffic pattern corresponds to a port scanning attack. Additionally, a button labeled "Generate Random Input" allows for testing with random feature values. This interface enables interactive and real-time cyber-attack classification using machine learning algorithms based on extracted traffic features.

Conclusion

The cyber-attack prediction system developed effectively utilizes machine learning algorithms to detect and classify different types of network intrusions based on extracted traffic features. Through the integration of supervised models like Decision Tree and SVM, along with deep learning approaches such as CNN-AE and LSTM-AE, the system demonstrates the comparative strengths of each technique. As observed from the accuracy comparison, traditional supervised algorithms like Decision Tree (96% accuracy) and SVM (84% accuracy) outperform unsupervised deep learning autoencoders in terms of detection accuracy. The user-friendly graphical interface further enhances usability by allowing real-time prediction based on selected input features. The final result indicates the nature of the cyber-attack (e.g., PortScan), helping in rapid identification and response. Overall, this system provides an efficient, interpretable, and practical solution for detecting cyber-attacks and can be extended with more features and ensemble strategies for improved performance and real-time cybersecurity applications.

Refrences

[1]. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

[2]. Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 2(1), 41–50. https://doi.org/10.1109/TETCI.2017.2758398

[3]. Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2019). Deep learning approaches for network intrusion detection: A review. Computer Communications, 136, 1–23. https://doi.org/10.1016/j.comcom.2019.01.011

[4. Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 5, 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418

[5]. Li, Y., Ma, R., & Jiao, J. (2020). A hybrid intrusion detection system based on PCA and optimized SVM. Procedia Computer Science, 174, 239–248. https://doi.org/10.1016/j.procs.2020.06.063

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

[6]. Khonde, S., & Ulagamuthalvi, V. (2020). Semi-supervised ensemble learning for intrusion detection system. Materials Today: Proceedings, 37, 3022–3027. https://doi.org/10.1016/j.matpr.2020.09.007

[7]. Patra, M. R., & Panigrahi, R. (2013). A novel hybrid intrusion detection system using soft computing techniques. Procedia Computer Science, 57, 851–857. https://doi.org/10.1016/j.procs.2015.07.471

[8]. Jadhav, S., Bhende, M., & Patil, S. (2023). Hybrid intrusion detection model using ANN, SVM and RNN. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2023.04.291

[9]. Rawat, S., Reddy, A. A., & Ghosh, S. (2019). A PCA-DNN based hybrid model for intrusion detection. Procedia Computer Science, 167, 1260–1269. https://doi.org/10.1016/j.procs.2020.03.389

[10]. Springer, A., Zhou, W., & Lee, R. (2021). One-class SVM and feature selection in hybrid intrusion detection systems. Expert Systems with Applications, 168, 114386. https://doi.org/10.1016/j.eswa.2020.114386

[11]. Ali, M., Khan, S., & Saeed, A. (2023). Lightweight hybrid intrusion detection system for IoT devices using MobileNetV2 and LightGBM. Sensors, 23(4), 2155. https://doi.org/10.3390/s23042155

[12]. Otoum, S., Aloqaily, M., & Jararweh, Y. (2020). A Q-learning-based intrusion detection system for industrial control systems. Computers & Security, 97, 101951. https://doi.org/10.1016/j.cose.2020.101951

[13]. Kim, H., & Lee, S. (2024). CNN and PSO-based hybrid intrusion detection system for smart city 5G networks. Sensors, 24(3), 1477. https://doi.org/10.3390/s24031477

[14]. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). A multi-step outlier-based anomaly detection approach to network-wide traffic. Journal of Network and Computer Applications, 42, 102–118. https://doi.org/10.1016/j.jnca.2014.03.011

[15]. Allahrakha, S. (2020). A multi-stage intrusion detection system using autoencoders and random forest classifiers. Procedia Computer Science, 167, 2276–2283. https://doi.org/10.1016/j.procs.2020.03.278

[16]. Popoola, S. I., & Atayero, A. A. (2021). CNN-LSTM based hybrid intrusion detection model for IoT botnet attacks. International Journal of Engineering Research and Technology, 10(2), 213–220.

Power System Technology
ISSN:1000-3673
Received: 16-07-2025 Revised: 05-08-2025 Accepted: 02-09-2025

[17]. Gupta, M., Singh, M., & Tripathi, R. (2021). An intelligent hybrid model for anomaly detection using K-means and neural networks. International Journal of Computer Applications, 183(5), 15–22. https://doi.org/10.5120/ijca2021921399

[18]. Ramadan, R. A., & Yadav, R. N. (2020). Anomaly detection in IoT using DBSCAN and random forest. Procedia Computer Science, 173, 376–384. https://doi.org/10.1016/j.procs.2020.06.043

[19]. Ahmed, S., & Khan, F. A. (2024). A novel hybrid intrusion detection system using ReliefF and XGBoost. Computers & Electrical Engineering, 110, 107901. https://doi.org/10.1016/j.compeleceng.2024.107901

[20]. Kokaz, N., & Türkben, A. (2025). Layered hybrid IDS framework for smart city 5G infrastructure. Future Generation Computer Systems, 143, 1034–1043. https://doi.org/10.1016/j.future.2024.10.028