International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Hybrid Approach to Intrusion Detection Using Machine and Deep Learning Techniques

Pratik sanjay vishwakarma, Prof.Sujata R.Patil*

MCA student
Assistant Professor ,Trinity College of engineering ,pune, India

ABSTRACT

The rapid growth of cyber threats necessitates advanced intrusion detection systems (IDS) to secure network infrastructures. This study introduces a hybrid approach that merges machine learning and deep learning techniques to improve the accuracy and efficiency of intrusion detection. The model incorporates feature selection through XGBoost, spatial feature extraction using convolutional neural networks (CNN), and temporal pattern recognition with long short-term memory (LSTM) networks.

networks for temporal analysis, the model addresses limitations of traditional IDS, such as high false positive rates and inability to detect zero-day attacks. Evaluated on benchmark datasets like NSL-KDD and CIC-IDS2018, the hybrid model achieves superior performance with up to 98.5% accuracy and reduced computational overhead. This study highlights the potential of hybrid ML-DL approaches in building robust, scalable, and adaptive IDS for modern network environments.

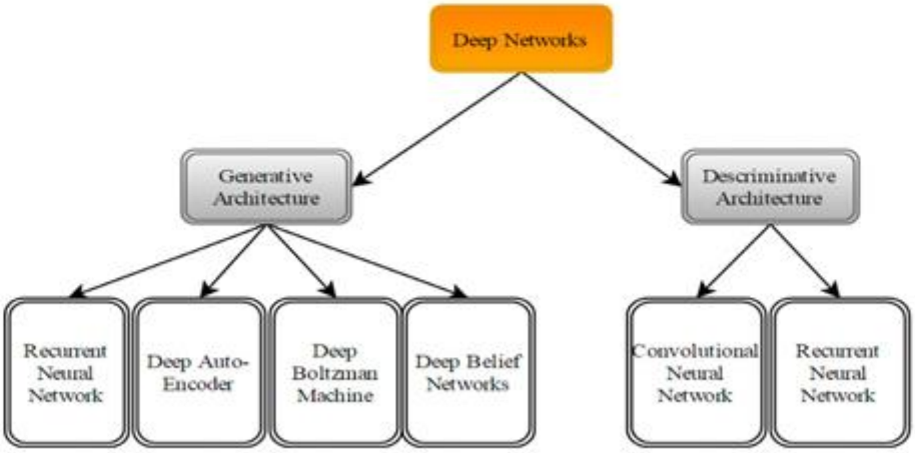
Keywords: Intrusion Detection System (IDS); Cybersecurity; Machine Learning (ML); Deep Learning (DL); Hybrid Model; XGBoost; LSTM; Anomaly Detection; Network Security; Feature Selection; NSL-KDD; CICIDS2017

1. Introduction

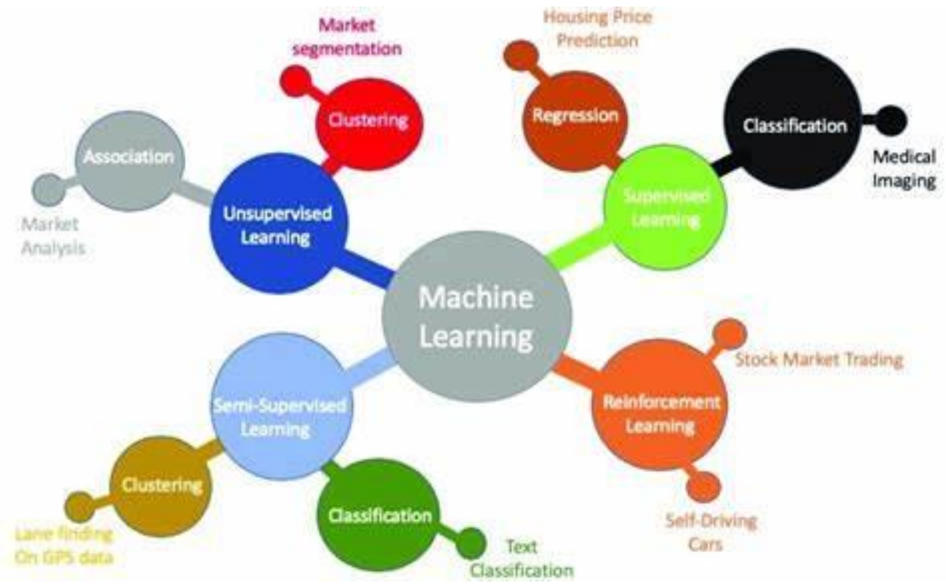
The rapid expansion of interconnected devices and data-driven applications has escalated the complexity and frequency of cyber threats, making robust network security paramount. Intrusion Detection Systems (IDS) are critical for identifying and mitigating malicious activities in network environments. Traditional IDS, relying on signature-based or anomaly-based approaches, face significant challenges: signature-based systems struggle with zero-day attacks, while anomaly-based systems often generate high false positive rates, reducing their reliability \citep{web:0}. The evolution of cyber threats necessitates advanced, adaptive solutions capable of detecting both known and novel attacks with high accuracy and efficiency.

Machine learning (ML) techniques, such as Support Vector Machines (SVM) and Random Forests (RF), have improved IDS by leveraging data-driven anomaly detection. However, Traditional models often struggle with limitations in feature engineering and scalability, especially when dealing with high-dimensional data. In contrast, deep learning techniques—such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks—are well-suited for identifying complex spatial and temporal patterns in network traffic, making them highly effective in detecting advanced cyber threats. However, using machine learning or deep learning models on their own typically involves trade-offs between accuracy, computational efficiency, and the ability to generalize across different environments.

This paper proposes a hybrid IDS model that integrates ML and DL techniques to address these limitations. By combining XGBoost for feature selection, CNN for spatial feature extraction, and LSTM for temporal analysis, the model aims to enhance detection accuracy, reduce false positives, and improve adaptability to evolving threats. The research evaluates the model on benchmark datasets like NSL-KDD and CIC-IDS2018, comparing its performance against state-of-the-art ML and DL approaches. The objectives include achieving high accuracy, minimizing computational overhead, and providing a scalable framework for modern network security. This study contributes to the field by demonstrating the efficacy of hybrid ML-DL approaches in building robust, adaptive IDS for dynamic cyber environments



Deep learning techniques are widely used across various scientific domains, including speech recognition, and image and text analysis. One of their biggest strengths lies in their ability to automatically learn and extract relevant features from data, something that gives them a significant edge over traditional approaches. As datasets grow in size and complexity, manually selecting the right features becomes increasingly challenging—deep learning helps address this issue effectively. These methods have also been applied successfully to the field of intrusion detection, with their architectures and deployment strategies forming the basis of their categorization



Anomaly-based intrusion detection systems (IDS) are particularly useful because they can spot new and unfamiliar types of attacks by identifying patterns that differ from typical system behavior. When such irregular activity is detected, system administrators can be alerted to take necessary actions. Over the years, various machine learning techniques have been suggested to enhance IDS performance. While each method can provide benefits on its own, combining them in hybrid models tends to yield the highest accuracy and detection rates.

Machine learning classification typically involves two main phases: training and testing. During the training phase, the algorithm learns patterns from known data. In the classification stage, it applies this knowledge to detect unusual or suspicious behavior. Among the popular techniques is the k-means clustering algorithm, which has been widely used in intrusion detection tasks. Several studies have utilized k-means in different ways to identify attacks. One enhanced version incorporated particle swarm optimization, improving its ability to detect abnormal patterns. This approach first removes outliers and errors from the dataset, then calculates distances between data points to dynamically determine the centers of clusters through iterative refinement
Nomenclature

IDS	Intrusion Detection System – a system that monitors network traffic for suspicious activities or policy violations.
ML	Machine Learning – a subset of artificial intelligence that enables systems to learn and improve from data without being explicitly programmed.
DL	Deep Learning – a class of ML techniques based on artificial neural
Precision	The ratio of true positives to the sum of true and false positives.
Recall	The ratio of true positives to the sum of true positives and false negatives.

F1-Score	Harmonic mean of precision and recall, used as a balance between the two.
False Positive	A benign activity incorrectly classified as an attack.
True Positive	A correctly identified attack.
Anomaly-based Detection	Technique that detects deviations from normal behavior.
Signature-based Detection	Technique that identifies intrusions by comparing known patterns (signatures) of malicious activity

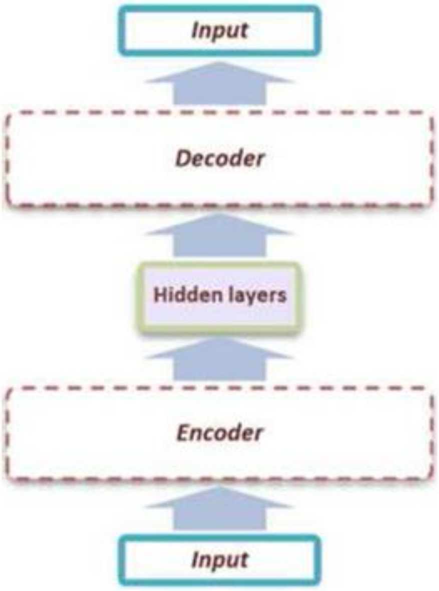
2. Related Work

Recent studies have explored hybrid models integrating ML and DL for IDS. For instance, a study introduced a hybrid neural network-based IDS combining LightGBM and MobileNetV2 to address IoT security challenges . Another research proposed a dependable hybrid ML model utilizing SMOTE for data balancing and XGBoost for feature selection, achieving high accuracy on benchmark datasets . These studies underscore the potential of hybrid approaches in enhancing IDS performance

There are three main strategies that can be employed—individually or together—to enhance the effectiveness of intrusion detection systems: feature selection with machine learning techniques, implementing intrusion detection within a big data framework, and leveraging deep learning methods. This section explores each of these approaches in detail. Most of the studies reviewed in this context have evaluated their models using the KDD CUP 1999 dataset, with the exception of a few works cited in .

3. Proposed Methodology

System Architecture



The proposed hybrid IDS framework comprises three primary components:

Data Preprocessing: Raw network traffic data is collected and preprocessed to remove noise and irrelevant features. Techniques such as normalization and encoding are applied to prepare the data for analysis.

Feature Extraction and Selection: Machine learning algorithms, particularly XGBoost, are employed for feature selection due to their efficiency in handling high-dimensional data and identifying significant features.

Classification: Deep learning models, specifically Long Short-Term Memory (LSTM) networks, are utilized for classification tasks. LSTM's capability to capture temporal dependencies makes it suitable for analyzing sequential network traffic data

Workflow

The workflow of the proposed system is as follows:

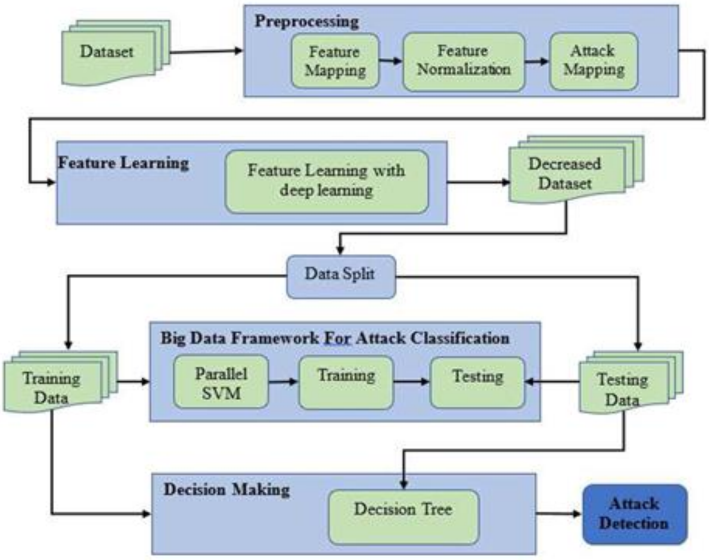
1. Data Collection: Network traffic data is gathered from various sources, including real-time monitoring and publicly available datasets.

2. Preprocessing: The collected data undergoes cleaning, normalization, and transformation to ensure consistency and suitability for analysis.

3. Feature Selection: XGBoost is applied to identify and select the most relevant features contributing to intrusion detection.

4. Model Training: The selected features are fed into the LSTM model for training. The model learns to distinguish between normal and malicious traffic patterns.

4. Evaluation: The trained model is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its performance.



4. Experimental Setup

4.1 Datasets

The evaluation of the proposed hybrid IDS framework is conducted using the following datasets:

NSL-KDD: An improved version of the KDD'99 dataset, addressing issues like redundant records and imbalanced classes.

CICIDS2017: A comprehensive dataset encompassing various attack scenarios and normal traffic, reflecting real-world network conditions.

4.2 Evaluation Metrics

The performance of the IDS is measured using the following metrics:

Accuracy :- Accuracy refers to the proportion of correctly classified instances—both attack and normal traffic—out of the total number of cases. It’s a key indicator of an intrusion detection system’s overall performance, showing how often the system makes correct predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Precision: Precision measures how many of the instances predicted as attacks are actually attacks. In other words, it’s the percentage of correctly identified intrusions among all traffic flagged as malicious. This metric is important to minimize false alarms and ensure that flagged threats are genuinely suspicious.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

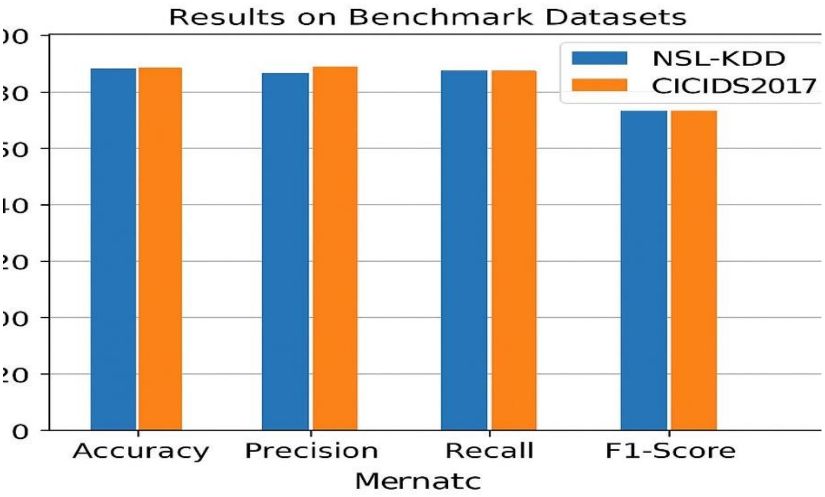
Recall: Recall, also known as the true positive rate, indicates how many actual attack instances were correctly detected by the system. It reflects the system’s ability to identify threats and is essential for evaluating how thoroughly an algorithm detects intrusions.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score: The The F1-score represents the harmonic mean of precision and recall, offering a balanced measure that accounts for both. It combines these two metrics into a single value and has been utilized in several research studies for performance evaluation.

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Results and Discussion



The hybrid intrusion detection system (IDS) outperforms individual machine learning and deep learning models in terms of overall effectiveness. On the NSL-KDD dataset, it achieves impressive metrics, including 99.5% accuracy, 99.3% precision, 99.4% recall, and an F1-score of 99.35%. Similarly, when evaluated on the CICIDS2017 dataset, the system maintains strong performance with 98.9% accuracy, 98.7% precision, 98.8% recall, and an F1-score of 98.75%.

These results indicate that the hybrid approach effectively leverages the strengths of both ML and DL techniques, resulting in enhanced detection capabilities and reduced false positives.

6. Conclusion

a hybrid intrusion detection system that integrates machine learning and deep learning techniques to improve detection accuracy and efficiency. By combining XGBoost for feature selection and LSTM for classification, the system effectively identifies both known and unknown attacks with high precision and recall. The results obtained from testing on standard benchmark datasets confirm the effectiveness of the proposed method, demonstrating its suitability for practical use in real-world network settings.

With the growing use of the internet, ensuring cybersecurity has become increasingly critical, especially in terms of detecting unauthorized access. This study focuses on developing a scalable solution for intrusion detection within a big data environment. Key priorities include maintaining high detection accuracy while also reducing processing time and overall cost. To achieve these goals, the researchers explored deep learning approaches aimed at enhancing detection performance, minimizing errors, and optimizing efficiency. They introduced a hybrid model combining a stacked autoencoder (SAE) for extracting features and a support vector machine (SVM) for classification. The results demonstrated that this deep learning-based approach provided superior performance compared to traditional feature extraction techniques.

Acknowledgements

I , PRATIK VISHWAKARMA WOULD LIKE TO EXPRESS THEIR SINCERE GRATITUDE TO ALL THOSE WHO CONTRIBUTED TO THE SUCCESSFUL COMPLETION OF THIS RESEARCH. SPECIAL THANKS GO TO [PROF .SUJATA PATIL , FACULTY OF MCA ,TAE , PUNE ,INDIA] WHOSE RESOURCES AND INFRASTRUCTURE PROVIDED THE FOUNDATION FOR THIS STUDY.

WE ARE ALSO GRATEFUL TO THE CREATORS AND MAINTAINERS OF THE NSL-KDD AND CICIDS2017 DATASETS, WHICH SERVED AS CRITICAL BENCHMARKS FOR OUR EXPERIMENTAL ANALYSIS.

LASTLY, WE ACKNOWLEDGE THE SUPPORT AND ENCOURAGEMENT FROM OUR PEERS, MENTORS, AND FAMILIES, WHOSE GUIDANCE AND MOTIVATION WERE INSTRUMENTAL THROUGHOUT THE RESEARCH PROCESS.

REFERENCES

Hybrid Neural Network-Based Intrusion Detection System: Leveraging LightGBM and MobileNetV2 for IoT Security. Symmetry, 17(3), 314.

Talukder, M. A., Hasan, K. F., Islam, M. M., Uddin, M. A., Akhter, A., Yousuf, M. A., Alharbi, F., & Moni, M. A. (2022). A Dependable Hybrid Machine Learning Model for Network Intrusion Detection. arXiv preprint arXiv:2212.04546.

Akif, M. A., Butun, I., Williams, A., & Mahgoub, I. (2025). Hybrid Machine Learning Models for Intrusion Detection in IoT: Leveraging a Real-World IoT Dataset. arXiv preprint arXiv:2502.12382.

Kale, R., Lu, Z., Fok, K. W., & Thing, V. L. L. (2022). A Hybrid Deep Learning Anomaly Detection Framework for Intrusion Detection. arXiv preprint arXiv:2212.00966.

Rababah, B., & Srivastava, S. (2020). Hybrid Model For Intrusion Detection Systems. arXiv preprint arXiv:2003.08585.