iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.77217

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

A Comprehensive Review using Machine Learning and Deep Learning Methods for Intelligent Intrusion Detection in the Domain of Cybersecurity

Miss. Alwiya Shaikh[1], Dr. Anamika Jain[2]
[1, 2]*Computer Engineering, Pune University*

Abstract: ***Organizations today depend heavily on cloud computing, IoT devices, and large-scale well-connected systems. This expansion brings serious cybersecurity risks though. Threats like ransomware, zero-day exploits, DDoS attacks, and advanced persistent threats are happening more often and getting more sophisticated. Current signature-based intrusion detection systems (IDS) struggle with these evolving attacks because they depend on predefined rules and known patterns. This limits how well they work against new threats. That's why researchers are exploring machine learning (ML) and deep learning (DL) techniques to build smarter intrusion detection solutions. This paper reviews ML-based, DL-based, and hybrid ML-DL methods for intrusion detection. We examine benchmark datasets including UNSW-NB15 and CIC-IDS2017, looking at their features, benefits, and drawbacks. Class imbalance and high-dimensional feature spaces create problems. The review covers several important models: Random Forest, XGBoost, Convolutional Neural Networks, Long Short-Term Memory networks, and Autoencoders. Hybrid architectures look most promising since they combine deep feature extraction with traditional machine learning classifiers, getting better detection accuracy and fewer false positives. We also identify research gaps, discuss computational challenges, and suggest future directions for building scalable, real-time, and interpretable IDS solutions that meet today's cybersecurity needs.***

Keywords: ***Intrusion Detection System, Cyber security, Machine Learning, Deep Learning, Hybrid IDS, XGBoost, LSTM, Autoencoders.***

I. INTRODUCTION

Cloud computing, IoT, and distributed architectures have changed how finance, healthcare, transportation, and smart cities work, improving efficiency and connectivity. But this growing dependence on networked systems has expanded the attack surface too. Systems face more complex cyber threats now. Modern attacks including ransomware, distributed denial-of-service (DDoS) attacks, zero-day exploits, botnets, and advanced persistent threats (APTs) happen more often, at larger scales, and are more complex. These threats put data confidentiality, integrity, and availability at risk [3]. Intrusion Detection Systems (IDS) are crucial for cybersecurity. They monitor network or host activities to spot malicious behaviour. Traditional signature-based IDS use predefined rules and known attack patterns to catch threats they've seen before. This keeps false alarm rates low. But these systems can't recognize unknown intrusion patterns without prior signatures, making them ineffective against new and evolving attacks [4]. Keeping signature databases updated takes a lot of resources and becomes impractical when threats change quickly.

Anomaly-based IDS solve this by learning normal system behaviour and flagging deviations that might indicate security attacks. But these systems create more false alarms. Network attack patterns change constantly. User access behaviours vary too [4]. How you deploy IDS depends on what your network needs. You can use Network-Based IDS or Host-Based IDS. Catching DDoS and port scanning attacks requires understanding normal network behaviour at different network points. Detecting insider attacks requires system log analysis [5]. Traditional IDS architectures have trouble keeping up with network complexity [6].

A. Background of Intrusion Detection

Intrusion Detection Systems (IDS) play a critical role in securing legacy and modern networks. Their importance keeps growing as cloud computing and IoT adoption evolving. Traditional signature-based IDS catch known attacks but need frequent updates. They can't handle new threats. Anomaly-based IDS spot deviations from normal behaviour but create lots of false positives. Networks vary too much [6]. There are two main IDS types. Network-Based IDS examine traffic. Host-Based IDS check system logs. Traditional IDS architectures can't keep up with how complex networks have become.

That's where machine learning and deep learning come in to contribute. These techniques can automatically spot complex attack patterns that traditional methods miss to identify. Researchers have used these advance methods to build hybrid IDS models. These adapt better and provide stronger cybersecurity protection [7], [8].

II. LITERATURE REVIEW

IDS research has changed a lot over the past twenty years. Early work mostly used traditional machine learning techniques tested on datasets like KDD Cup 1999. This dataset has problems though. It contains duplicate records and outdated attack patterns. That makes it less useful for evaluating modern IDS [1]. Researchers created the NSL-KDD dataset to fix these issues. It removed duplicates and improved class distribution. But even with these changes, NSL-KDD doesn't capture how complex and diverse today's network traffic and cyberattacks really are [2].

More recent datasets like UNSW-NB15 and CIC-IDS2017 were developed to better match real-world attack scenarios. These datasets include realistic network traffic and cover more attack types. That makes them better for testing modern IDS models [3], [4]. But problems remain. High-dimensional feature spaces and class imbalance still make model training and evaluation difficult. Researchers have studied traditional machine learning algorithms extensively. These include Decision Trees, Support Vector Machines, Random Forests, k-Nearest Neighbours, and ensemble methods like XGBoost. These methods work well for known attack patterns. But they often fail to catch complex, evolving, and rare attacks [5], [6].

IDS research has shifted toward deep learning techniques in recent years. Methods like Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Autoencoders can automatically extract features and learn temporal patterns. This has led to major improvements in detection accuracy [7]–[9]. Hybrid IDS approaches are also gaining attention. These combine machine learning classifiers with deep learning-based feature extraction. They've shown better performance when dealing with class imbalance and detecting stealthy attacks. But IDS models still have problems. Real-time deployment, scalability, computational overhead, and model explainability remain challenging. Solving these issues is a key focus for future IDS research [10]–[12].

III. PROPOSED METHODOLOGY.

This paper also presents a conceptual framework for hybrid intrusion detection. This framework brings together Machine Learning, Deep Learning, and other hybrid IDS approaches that researchers have explored. The goal isn't to prove it works through experiments or introduce a new algorithm.

Instead, we want to create a systematic plan showing how Machine Learning and Deep Learning techniques can work together effectively. Machine Learning and Deep Learning are well-established fields. This framework shows how to combine them. The framework addresses problems found in earlier research. These include situations where certain attack types are rare. They also include situations where we need to understand how things change over time. The framework also needs to handle complicated and changing cyberattack patterns. It must deal well with class imbalance and temporal dependency modelling. It needs to be resilient against evolving cyberattack patterns.

A. Data Acquisition

Network traffic data comes from benchmark datasets like UNSW-NB15 and CIC-IDS2017, or from operational networks. The collected data usually includes packet-level details or flow-level statistics. These show what normal and malicious network traffic looks like. These datasets contain different types of traffic and attack scenarios. This makes them useful for testing and developing network intrusion detection systems.

B. Data Pre-processing

Data pre-processing matters a lot for making intrusion detection work better. We need to prepare the data properly. That's why we use standard techniques to transform it. This includes scaling all numerical values so they're on the same scale. We use Min-Max scaling or Z-score normalization for this. This ensures all features are consistent.

We also handle categorical data by converting it into numerical form using One-Hot Encoding. And we try to remove noise and unnecessary features using Correlation Analysis or Principal Component Analysis. This helps reduce features that aren't important. Data pre-processing is a crucial step. It includes these techniques to improve intrusion detection performance and make sure it can detect things correctly.

C. *Deep Feature Extraction (Autoencoder)*

The Autoencoder is used as a features extractor from network traffic. It does this by finding a way to describe complicated traffic data. The Autoencoder looks at the traffic data that goes in and the traffic data that comes out and tries to make them match. This helps to get rid of information, ignore random noise and find the basic patterns in the traffic.

D. *Temporal Modeling (Bi-LSTM)*

To understand what happens in network traffic over time, we use a Bidirectional Long Short-Term Memory network, or Bi-LSTM. This Bi-LSTM network is different from similar networks because it processes information in two directions: forward and backward. This means the Bi-LSTM network can learn how things are related to each other over time in both directions. The Bi-LSTM network is really good at finding patterns that happen over time or that are hard to see. This makes it very good at detecting cyberattacks that are slow and try to hide.

E. *Classification (XGBoost Algorithm)*

For the final decision-making stage, the framework uses XGBoost, a gradient-boosted decision tree classifier. We chose XGBoost because it has high classification accuracy, built-in regularization that prevents overfitting, and it handles class imbalance well.

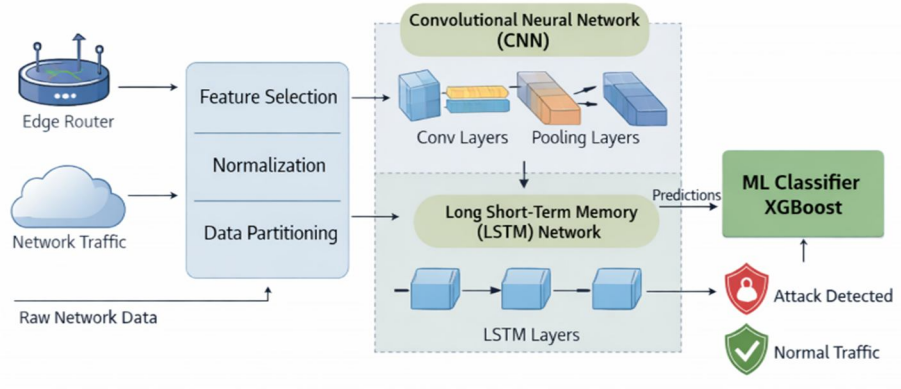


Fig.1 Proposed Conceptual hybrid intrusion detection framework

The system first takes the traffic and extracts important parts, makes sure everything is on the same scale, and divides it into smaller pieces. This ensures the system only uses the right information during training and testing. This follows approaches used in previous intrusion detection studies [1], [2]. The prepared features then go into a model that combines a Convolutional Neural Network with a Long Short-Term Memory network.

The Convolutional Neural Network processes the network flows. It finds connections between nearby features and extracts important characteristics. It does this using convolutional layers and pooling layers. This helps the Convolutional Neural Network learn automatically, which is useful for intrusion detection systems, as shown in previous work [3], [4]. The learned features then go to the LSTM network. This network is good at seeing what happens over time and finding patterns in attacks that occur sequentially in network traffic sessions. The LSTM network understands how things connect even when they happen far apart in time.

The features extracted by the CNN-LSTM module go to the XGBoost classifier for final decision-making. The XGBoost classifier makes the final decision. The XGBoost classifier is good at dealing with complex relationships and imbalanced traffic patterns. This makes the whole system more stable and effective. The XGBoost classifier examines the output and then decides if the network traffic is normal or malicious. By combining the spatial learning strength of CNN, the temporal modeling capability of LSTM, and the high classification performance of XGBoost, the proposed framework provides a more reliable and efficient network intrusion detection solution for modern LAN and router environments [3], [5], [7].

IV. COMPARATIVE ANALYSIS OF DIFFERENT IDS MODELS

The functionality of Intrusion Detection Systems is evolving. Previously, individuals relied solely on Machine Learning to identify issues. Nowadays, there is a shift towards utilizing Deep Learning techniques. Additionally, some practitioners are employing a combination of both Machine Learning and Deep Learning approaches. By utilizing these measures for cybersecurity, their efficacy in addressing cyber threats can be evaluated.

A. *Machine Learning Based IDS*

Machine learning models like Decision Trees and Support Vector Machines are really good at finding intruders. They are also simple to understand and do not need a lot of computer power. Decision Trees and Support Vector Machines work well with network traffic data that's easy to read. This is especially true when the important features of the network traffic data are carefully picked. Machine learning models such as Random Forests and k-Nearest Neighbours are also used for this. Ensemble learning techniques are used too. Machine learning models are good, at intrusion detection because they are simple and easy to understand. They do not need a lot of computer power to work. When we talk about machine learning techniques some methods really stand out. Ensemble classifiers are one of them and a great example is XGBoost. XGBoost is very good at what it does because it uses a lot of models called weak learners and combines them to make a strong one. It does this by using gradient boosting and regularization. This is explained in detail in reference number 7. XGBoost and other ensemble classifiers like it are really good, at making predictions. Machine learning based systems also do not work well when they are looking at a lot of normal computer traffic and just a little bit of bad traffic. This is because the bad traffic gets lost in all the traffic and the system has a hard time finding it as we see in things, like [3] and [4]. Machine learning based systems that find things have trouble with this.

B. *Deep Learning Based IDS*

Deep learning techniques have significantly advanced the study of Intrusion Detection Systems by enabling the automatic identification and hierarchical learning of features. Specific types of networks, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Autoencoders, excel at detecting patterns in network traffic across both spatial and temporal dimensions. These networks play a crucial role in research related to Intrusion Detection Systems. CNNs are particularly adept at analysing spatial relationships among traffic features, while LSTM and Bi-directional LSTM networks are effective in capturing long-term dependencies. This capability is vital for recognizing attack behaviours that occur sequentially and may be challenging to identify.

C. *Hybrid ML-DL IDS Models*

Hybrid systems that integrate Machine Learning and Deep Learning offer effective solutions to the limitations of using either approach in isolation. In these hybrid models, the Deep Learning components are typically responsible for feature extraction and analyzing temporal changes, while the Machine Learning components focus on decision-making. This integration allows us to leverage the strengths of both methodologies: the deep analytical capabilities of Deep Learning and the efficiency and straightforwardness of Machine Learning. Recent research indicates that hybrid approaches, which integrate methods such as Autoencoders with XGBoost, CNNs alongside BiLSTM, or a combination of Autoencoders with both BiLSTM and Random Forest, produce superior outcomes compared to using individual techniques in machine learning and deep learning. These integrated approaches have proven to be effective on multiple datasets, such as UNSW-NB15 and CIC-IDS2017. They excel in accurately detecting threats by minimizing false positives and improving the identification of atypical attack patterns. Moreover, models like Autoencoders with XGBoost, CNNs paired with BiLSTM, and Autoencoders combined with BiLSTM alongside Random Forest are particularly adept at processing diverse data types. They skillfully handle distracting or irrelevant information, which makes them highly appropriate for extensive and intricate network settings.

Table1 Comparative Analysis of traditional and Hybrid Intrusion detection Algorithm

Ref. No.	Algorithm	Accuracy in %	Latency	Key contribution towards proposed problem
[1]	Decision Tree(DT)	86	Low	Suitable for small LAN environments
[2]	Random Forest(RF)	88	Moderate	Robust to noise, better feature learning than decision tree
[3]	Support Vector Machine(SVM)	87	High	Good Margin based separation. Limited performance on large scale.
[4]	K Nearest Neighbour	85	High	Simple and effective but computationally expensive for real time IDS
Proposed Methodology	Hybrid CNN + LSTM+XGBoost	96	Optimized	Robust Boosting classification for high detection accuracy and better LAN and Router attacking handling

Table 1 shows us how traditional machine learning-based intrusion detection algorithms compare to the hybrid CNN+LSTM+XGBoost framework. We look at classifiers like Decision Tree and Random Forest and Support Vector Machine and K-Nearest Neighbour. These traditional machine learning-based intrusion detection algorithms work okay they are correct, about 85 to 88 percent of the time. The hybrid CNN+LSTM+XGBoost framework and these traditional machine learning-based intrusion detection algorithms also have latency characteristics. Decision Tree and Random Forest are good at making decisions and dealing with noisy traffic but they do not do well when they have to handle complex attack patterns in big local area networks.

Traditional ML models like Decision Tree and Random Forest struggle with complex, high-dimensional attacks, while SVM and KNN, though accurate, are too slow for real-time intrusion detection. The proposed hybrid CNN-LSTM-XGBoost framework achieves higher accuracy (96%) with optimized latency, offering superior performance, scalability, and reliability for LAN and edge-router environments. Overall, intrusion detection systems must handle increasing network complexity and evolving threats efficiently. Datasets like UNSW-NB15 and CIC-IDS2017 help but suffer from class imbalance, causing difficulties in detecting rare attacks. Deep learning models excel at capturing attack patterns and temporal dependencies (e.g., via LSTM/Bi-LSTM), yet are computationally expensive, limiting their use in resource-constrained settings like IoT.

Hybrid ML-DL approaches combine deep feature extraction with robust classifiers, improving detection rates and reducing false alarms. However, they often act as “black boxes,” making explainability a key challenge for trust and adoption. Scalability and real-time deployment remain open issues, demanding lightweight models that balance accuracy with efficient resource use.

V. RESEARCH GAPS AND FUTURE DIRECTIONS

Despite advances in ML, DL, and hybrid IDS, several critical research gaps remain.

A. Explainable Hybrid ML-DL Models

Hybrid IDS achieve high accuracy but lack transparency. These hurts trust in security-critical applications. Existing explainability methods are limited. They're computationally intensive and not optimized for real-time use (they need to be under 100ms). Developing lightweight, real-time, model-specific explainability for hybrid architectures is essential.

B. Privacy-Preserving Federated Learning

Current IDS rely on centralized data. This raises privacy concerns. Federated learning shows promise but existing work is limited. It focuses mainly on homogeneous networks. Research needs to focus on non-IID data, diverse architectures, and effective communication protocols. This will help create collaborative intrusion detection systems that respect privacy.

C. Semi-supervised Learning for IoT

Supervised learning dominates IDS research. Semi-supervised approaches remain scarce despite their suitability for resource-constrained edge and IoT devices with limited labeled data. Future work should develop semi-supervised models. These should balance accuracy (over 90%) with tight memory (under 50MB) and power (under 1W) constraints.

D. Graph Neural Networks for Network-Aware IDS

Networks have inherent graph structures that IDS often overlook. GNN-based IDS research is limited and small-scale. New work should focus on scalable GNNs that model large, dynamic network topologies and integrate temporal information for detecting multi-stage attacks.

E. Taxonomy-Driven Research Opportunities

Ultra-low-power IoT settings require models that work with minimal resources. Hybrid ML-DL and Federated Learning: No existing works combine these promising paradigms for privacy-preserving and accurate IDS.

Semi-supervised Learning and Edge Computing: These areas are relatively unexplored but very important for real-world applications where labeled data is scarce.

F. Additional Priorities

Intrusion detection systems should improve adversarial resilience. This should focus not only on evasion attacks but also on overall adversarial robustness. Additionally, there's a need to develop transfer learning techniques. These should allow adaptation of IDS models from one organization to another while using less labeled data. There's also a need for real-time adaptive IDS that can adapt to concept drift.

VI. CONCLUSIONS

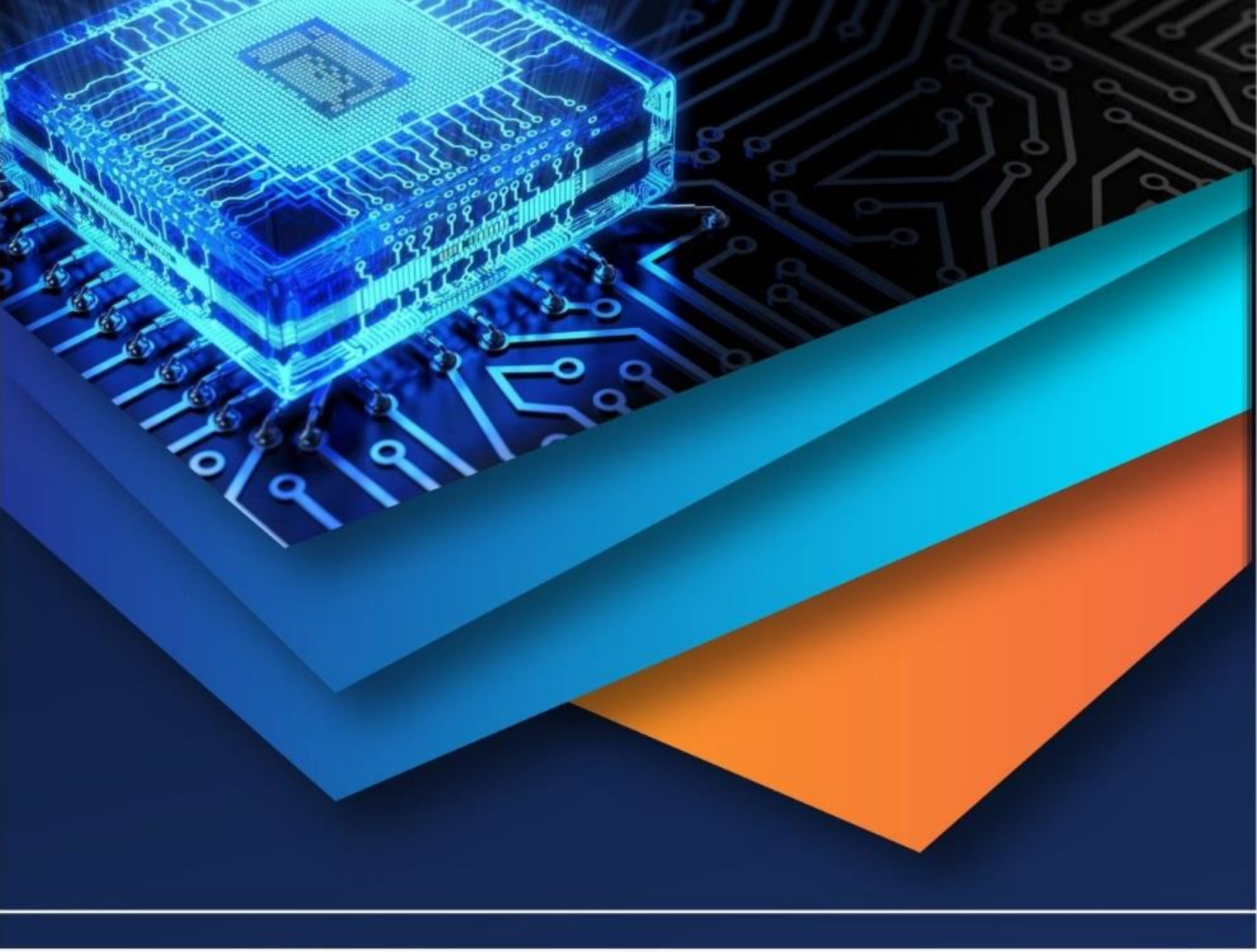
This paper reviews machine learning, deep learning, and hybrid ML, DL approaches for intrusion detection in modern cybersecurity settings. It highlights the shortcomings of traditional signature-based IDS in dealing with complex and changing threats. Machine learning models offer efficient and understandable solutions, but they rely heavily on manually crafted features. Deep learning methods, like CNNs, LSTMs, and Autoencoders, enhance detection accuracy by automatically learning features, though this comes with increased computational demands. This review shows that hybrid ML, DL frameworks successfully combined and enhanced both the approaches. They achieve better detection accuracy, reduced false positives, and greater robustness across benchmark datasets. This makes it promising option for scalable intrusion detection systems.

VII. ACKNOWLEDGMENT

The authors wish to acknowledge the contributions of researchers in the field of intrusion detection systems whose work has been reviewed and analyzed in this comprehensive literature review.

REFERENCES

[1] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, “Cost-based modeling for fraud and intrusion detection: Results from the JAM project,” DARPA Information Survivability Conference and Exposition, 2000.
[2] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.
[3] N. Moustafa and J. Slay,“UNSW-NB15: A comprehensive data set for network intrusion detection systems,”Military Communications and Information Systems Conference (MilCIS), 2015.
[4] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,”International Conference on Information Systems Security and Privacy (ICISSP), 2018.(CIC-IDS2017)
[5] T. Chen and C. Guestrin,“XGBoost: A scalable tree boosting system,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.
[6] A. L. Buczak and E. Guven,“A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
[7] Li, P. Yi, Y. Wu, L. Pan, and J. Li, “A new intrusion detection system based on KNN classification algorithm in wireless sensor network,”Journal of Electrical and Computer Engineering, 2014.
[8] M. A. Aydin, A. H. Zaim, and K. G. Ceylan, “A hybrid intrusion detection system design for computer network security,” Computers & Electrical Engineering, Elsevier, 2009.
[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.
[10] C. Yin, Y. Zhu, J. Fei, and X. He,“A deep learning approach for intrusion detection using recurrent neural networks,” IEEE Access, vol. 5, pp. 21954–21961, 2017.
[11] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, “Long short-term memory recurrent neural network classifier for intrusion detection,” International Conference on Platform Technology and Service (PlatCon), IEEE, 2016.
[12] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai,“Kitsune: An ensemble of autoencoders for online network intrusion detection,”Network and Distributed System Security Symposium (NDSS), 2018.
[13] Z. Wang, Y. Zeng, Y. Liu, and D. Li, “Deep belief network integrating XGBoost for intrusion detection,” IEEE Access, vol. 6, pp. 1908–1918, 2018.
[14] H. Hindy et al.,“A taxonomy and reminder of intrusion detection systems,” IEEE Communications Surveys & Tutorials, 2020.
[15] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies, 2016.
[16] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,”IEEE Transactions on Emerging Topics in Computational Intelligence, 2018.
[17] A. Ahmad et al.,“A deep learning based intrusion detection system for IoT networks,”IEEE Internet of Things Journal, 2021.
[18] E. Hodo et al.,“Threat analysis of IoT networks using artificial neural network intrusion detection system,”IEEE Transactions on Emerging Topics in Computational Intelligence, 2022.
[19] S. Yaras et al.,“Hybrid deep learning and metaheuristic-based intrusion detection system for IoT,”Electronics, MDPI, 2024.
[20] A. Pinto et al.,“A comprehensive survey on machine learning-based intrusion detection systems,”IEEE Access, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY