# Explainable AI-Based Innovative Hybrid Ensemble Model for Intrusion Detection: A Review

Deepak Kumar [1], Pooja Meena [2]
[1,2]Department of Computer science and Engineering, Rabindranath Tagore University, Raisen, (MP) India
[1]deepakkumar20019@gmail.com, [2]meena.pooja1@gmail.com

**Abstract:**

Developments in digital communication, cloud computing, Internet of Things (IoT), and smart network technologies are witnessing a significant rise in cybersecurity threats and very sophisticated cyber-attacks. IDS plays a major role in detecting malicious activities and securing network infrastructures from unauthorized access. Conventional IDS approaches usually encounter limitations accompanying poor detection accuracy, high false positive (FP) rates, poor scalability, and inability to detect zero-day attacks. Various researchers have employed Machine Learning (ML), Deep Learning (DL), and ensemble techniques in their smart intrusion detection systems to address these challenges. Recent studies show that hybrid models induced by CNN, LSTM, Bi-LSTM, Autoencoder, Random Forest, and optimization techniques have proven to increase classification performances of attacks. Furthermore, Explainable Artificial Intelligence (XAI) methods such as LIME and SHAP have seen increasing attentions to regain transparency and interpretability during cybersecurity decision-making. While conducting an in-depth review of existing IDS research on ML, DL, ensemble learning, federated learning, hybridization with blockchain, and XAI, a comprehensive comparison is presented outlining the methodologies used, performance outcomes, strengths, and weaknesses of existing models across IoT, cloud, healthcare, industrial, and wireless network environments. The study further points out main paucities in research such as computational complexity, lack of explainability, scalability challenges, and lesser real-time adaptability. To this end, the paper suggested an equally innovative explainable hybrid ensemble Intrusion Detection Framework for broader attack coverage with robust accuracy, transparency, and real-time protection of cybersecurity.

**Keywords**: Intrusion Detection System, Machine Learning, Deep Learning, Explainable Artificial Intelligence, Hybrid Ensemble Model, Cybersecurity

## I. INTRODUCTION

The pace by which digital technologies, communication networks, and cloud computing have shown progress has been very fast. Even though the Internet of Things (IoT), the Industrial Internet of Things (IIoT), and the smart system have brought an intelligent automation of data transmission in a variety of industries, but their very evolution has opened a chink in the systems to the wicked ever-rising cyber threats of hacking, malware, ransomware, DoS, email phishing, etc. With the increased importance of interconnection between systems in order to run critical operations, secure cybersecurity has become a general concern for governments, industries, health care, financial, and smart city infrastructures. Because of this, emerging technologies like Intrusion Detection Systems (IDS) have not only become an essential constituent within cybersecurity frameworks to oversee the network traffic and spot malicious behavior but also to prevent acts of cyberterrorism before they crush unto inflict on society [1].

The most widely used intrusion detection systems, traditional systems are categorized as signature-based and anomaly-based: signature-based systems are based on predefined signatures of known attacks, comparing network activity against them, whereas anomaly-based systems are based on abnormal behavior identified by the analysis of deviation from normal traffic patterns. Intrusion detection schemes have been modified and are currently stored in the IDS; but this conventional method lies below that of good detection accuracy. 0-day detection has not evolved as expected for detection of threats, and a high rate of false alarms makes the alerted network administrators and users wonder due to the number of spurious failures. [2] Now, while in most cases decisions are primarily made by them, administrators have turned to Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) algorithms in the context of smart intrusion detection systems. In particular, support vector machines, random forests, decision trees, k-Nearest Neighbor, and Naive Bayes, have until now shown proven results for the sorting and classifying of truly malevolent traffic patterns. These methods enable the intrusion detection class of models to learn from what has already occurred, giving boost classification outcomes automatically. Alike, deep learning models, i.e., Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional-LSTM (Bi-LSTM), Autoencoders, and Deep Belief Networks (DNN), show remarkable success in recognizing particular complex and unknown cyberattacks, where they extract high-level information and leverage it to idealize an input image [3].

There are several challenges awaiting standalone models in deep learning and machine learning models, despite their being laden with advantages. Many ML models face difficulties in feature selection, data imbalance, and scalability. Deep

learning solutions demand large computational resources, larger datasets, and higher processing times, yet most IDS models based on deep learning work as "black-box" systems, giving predictions devoid of explanation about how or why the decision was reached. The unidentified-attribution models of AI are deficient in trust. This is due to the host of essential applications in which these models are adopted like health systems, manufacturing sector, financial installations, and autonomous driving-not being able to understand the basis of intrusion detection decisions [4].The figure 1 illustrates the overall workflow of an intrusion detection framework including data preprocessing, feature selection, classifier training using DT, RF, and SVM models, voting-based ensemble prediction, and performance evaluation metrics such as accuracy, precision, recall, and F1-score.

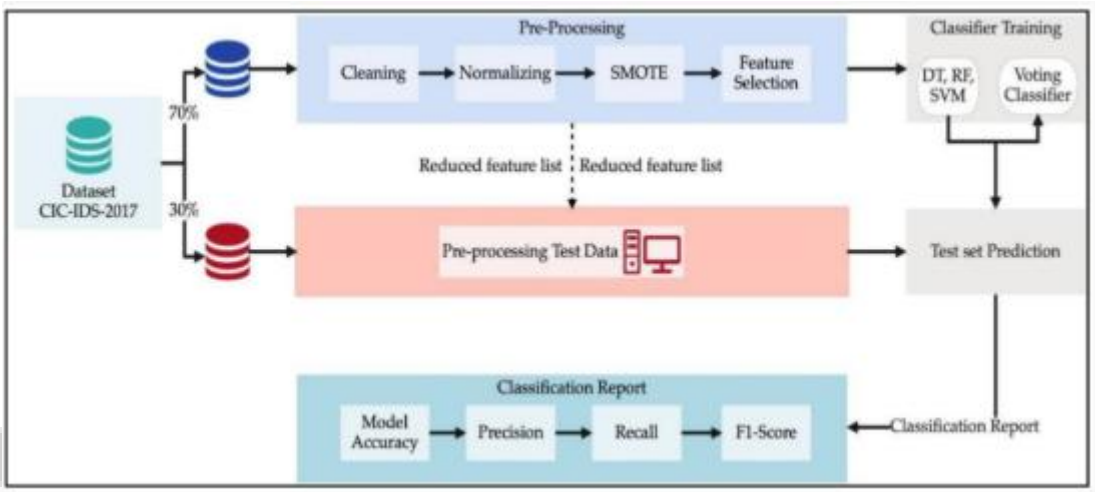


**Figure 1: Machine Learning-Based Intrusion Detection System Using Ensemble Classification [5]**

To address these concerns, hybrid and ensemble learning approaches have gained significant attention in recent years. Hybrid IDS models combine multiple machine learning and deep learning algorithms to improve intrusion detection performance, reduce false alarm rates, and enhance system robustness. Ensemble learning methods such as bagging, boosting, stacking, and voting classifiers integrate the strengths of multiple models to achieve higher classification accuracy and better generalization capability. Researchers have also integrated optimization algorithms including Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and bio-inspired optimization techniques to improve feature selection and enhance model efficiency [6]. These hybrid ensemble approaches have demonstrated superior performance in identifying both known and unknown attacks across IoT, cloud, wireless sensor networks, healthcare systems, and industrial communication environments.

Another promising area in research related to cybersecurity is Explainable Artificial Intelligence (XAI). Mostly, this aim on increasing the interpretability and transparency of AI system assumption by offering understandable explanations about the prediction of models and decisions thinking processes. For boosting trust and accountability in intrusion detection systems,like Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), and attention-based visualization techniques are quite popular. Through explainable AI, cybersecurity analysts can now understand why a certain network activity is labeled as malicious and hence support faster security responses with minimal uncertainty. The precision is really critical especially in cases where certain levels of decision transparency and nature of accountability prevail in order to practically manage threats to a secure environment.

Now comes federated learning, blockchain, edge computing, cloud security integration, and some high-tech solutions, such as cloud computing, that researchers have been kind to exploit in raising the level of intrusion detection in advanced systems. The concept of federated learning is a great contribution for preserving privacy, as it allows for the participation of models in the training from heterogeneous devices nonsharing sensitive data, thus providing immense assistance in the era of GPU technologies and a passive environment. Blockchain has certainly brought in a significant step toward the enhancement of security and trust management while securing for devices one tamper-resistant channel [7]. These novel approaches have contributed a huge part to evolving state-of-the-art intrusion detection systems for modern cyber infrastructures in happening intelligent and distributed environments.

The transformation from CSI to classic and advanced CSI poses the last and biggest challenge in deploying the E-CRM framework. This is probably the most vulnerable point in e-customer relationship management challenges. Unlike all other concepts based on information and communication technologies, E-CRM does not consider this. Changing from customized or packaged application-based CSI to newer human-centered, AI-like CSI-based models and local information bases, and introducing system interoperability, pose a greater practical challenge than that experienced with mobile commerce and standalone computing. [8] More intelligent challenges are introduced in real-time by the dynamicity and

the bursting architecture of such systems, and it is herein that the E-CRM knowledge is experimented with. This aspect changes the operational framework into knowing very little about the possibilities from a network of actions.

With the rise of AI-based technologies, the need for using them as effective means for intrusion detection algorithms is also rising. This calls for innovative intrusion detection frameworks combined through machine learning, deep learning, ensemble learner, and almost all optimization techniques to produce cybersecurity solutions of striking efficiency, accuracy, adaptability, interpretation, and extreme scaling. A model of intrusion detection from the hybrid ensemble system, based on Explainable AI, can deliver a very high level of detection capability with transparency and reliability in decision-making. This technology will be an integral part for the security of today's IT networks including the IoT networks, cloud platforms, healthcare systems, smart cities, autonomous vehicles, and industries.

This review paper undertakes a comprehensive appraisal of several latest advancements in the amount of research covered in machine learning and deep learning-based systems intended for intrusion detection as well as other hybrid ensemble learning and explainable artificial intelligence techniques [9]. The subject investigates the currently approved methodologies, datasets, performance issues, advantages, and disadvantages concerning IDS at the outset. It then goes on to delve into and scrutinize various issues related to cybersecurity interpretability, scalability, adversarial robustness, and real-time detection. Finally, the study calls for development in the area of new research focused on explainable hybrid ensemble IDS frameworks in order to enhance cybersecurity protection in the context of today's intelligent networking environment.

## II. OVERVIEW OF INTRUSION DETECTION SYSTEMS

One of the most critical components of today's cybersecurity environment, Intrusion Detection Systems (IDS) serve as an essential tool that monitors the network's traffic, characterizes undesirable occurrences and safeguards the computer system against unauthorized access, besides random cyber threats. With the progression in communication technologies and their combination on networks with cloud computing, Internet of Things (IoT), Industrial Internet of Things (IIoT), WSN, and smart city systems, illegal agitations have become exceptionally sophisticated, rendering them difficult to notice through traditional forms of security implementation. This is what makes the IDS technology appealing to both academics and practitioners who think of ways of hardening the systems' cyber defense. It is closely looking over network behaviors, systems activities and, communication patterns in trying to detect dubious events or send alerts for any abnormal happenings [10]. The primary goal towards which an intrusion detection system strives is the early identification of cyberattacks to prevent damage to network infrastructures, confidential information, and organizational resources. The accelerated escalation in the number of cyber threats, including malware, ransomware, phishing attacks, denial-of-service attacks, distributed denial-of-service attacks, SQL injection, spoofing, botnets, and insider attacks, has been noteworthy in recent years. Since they are security solutions lacking the capability to bring promising identification to advanced threats and unknown attacks, traditional firewalls and antivirus systems are often rendered ineffective. Thus, IDS technologies supplement the layers of security by analyzing traffic behavior and noting unusual patterns that may suggest the context of malicious activities.
In general, IDS systems are sorted under two main groups- Host-Based Intrusion Detection Systems (HIDS) and Network-Based Intrusion Detection Systems (NIDS). Host-Based IDS is employed for monitoring activities taking place on the individual devices - machines, servers, or endpoints. Such analysis may cover many activities such as system logs, file modifications, application behavior, and user conduct. HIDS is identified as very useful for spotting insider threat, unauthorized file access, and malware attack. HIDS, on the other hand, may consume hosts resources and may lack in network monitoring capabilities [11]. In contrast, Network-Based IDS monitors network traffic through communicating channels looking for suspicious packets or abnormal communication patterns. NIDS is used throughout enterprise networks, cloud environments, IoT, and industrial infrastructures to identify cyberattacks as they happen. Although NIDS allows the visibility of networks, it is less favorable in encrypted traffic analysis and high-speed network monitoring.

IDS designates are often classified into signature-based IDS, anomaly-based IDS, and hybrid IDS on the basis of detection techniques. Signature-based IDS matches against a list of threat signatures that have been already cached in the system, using network activities to assign an incoming network input with a certain statement to make a judgment as to whether it constitutes an attack. This strategy is efficient in the detection of known attacks (with low false alarm rates) since their patterns are predetermined from the signature database. Nonetheless, there is a major drawback, in that while detecting only known attack types, signature-based IDS in no way form or version eclipses any of the zero-day attacks and new threats emerging on cyberspace, simply because of the absence of any track record for the corresponding unknown attack patterns on the signature database [12]. In contrast, after training on benign activities, anomaly-based IDS exposes normal network behavior profiles and raises warning flags with minor system changes, up to a certain point, all depending on a statistically derived departure assumed toward anomaly. Anomaly-based IDS can detect unwanted traffic and zero-day attacks, but it may also raise many false alarms as dynamic network behavior introduces networking structure variability over time. Generally, the hybrid IDS combines the best of both worlds - that is, signature-based IDS and anomaly-based IDS - to limit the restrictions and hence enhance overall intrusion-detection performance through mutual dependencies.

Traditional IDS approaches relied predominantly on rule-based systems and applied analysis techniques for intrusion detection. Such mechanisms are insufficient to combat contemporary cyber threats marked at having highly complex algorithms, voluminous data, and continuously evolving attack strategies. To mitigate these limitations, Intrusion Detection Frameworks have integrated Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). Through machine learning algorithms, the IDS system can learn patterns from the historical dataset and automatically classify network activities as normal or malicious. Classification tasks are carried out to a great extent using supervised learning algorithms. Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and Naïve Bayes are enlisted in most intrusion classifications. These learning algorithms provide efficient attack detection performance with minimal manual intervention [13].

Clustering and anomaly detection come under the umbrella of unsupervised methods for identifying unknown behavior attacks within Intrusion Detection System (IDS) applications without the need for labeled training data. These methods are extremely useful in environments where it would be difficult to gather labeled cybersecurity data. Reinforcement learning too has captured the minds of adaptive cybersecurity systems which are expected to capitalize on environmental interaction and continuously enhance attack detection strategies. The last many years have seen a significant transformation from feature-based learning to deep learning. In the past, researchers relied on applying scaling and network monitoring to develop patterns that require less manual intervention. Major disadvantages of using deep learning models to develop networks, when considering revolutionary advancements, are that they visualize networks for understanding their day-to-day features in a way that the scaled perspective would shut out. Despite these early graphical results, networks are instrumental only to recognize spatial features in graphs [14]. This article clarified how convolutional, stack-august, and recurrent networks could be used to add an extra level of deep learning tilt to the detection sceneries from those entire sessions. One exploit is that the time-series data signals do not require an explorer, thus layers with a different weight of the same weights (like single weight) may be described; in this specific instance, the later portions emphasize temporal and sequential considerations in the model. Professional reaction from network merchants, based on ideas taken from the parading world, could be highly effective as a baseline since using these merchants means the mother gets knowledge while individuals in paralytics grasp this evidence. In the concluding paragraph, I heard that simple neural networks... would fire well in bots to predict TM traffic, which is real great because LSTMs are brain-children of BDLs and shape way to elementary deep learning techniques. This discussion turns into the third insight, but thanks to adversaries progressing complicated from one-over-simplified overlay detection being an easy mean into an inexorable mode no sooner than the existing IAM2 evoked social acceptance of the deadly network traffic. The authors are promising that the dwellers of deep learning all... detectors are invalid due to a lack of appropriate planning for temporal data signals. Hybrid and ensemble learning models were introduced as additional means to improve IDS performance, taking the opportunity to combine multiple models to elevate the cybersecurity analysis, showing breakthrough grounds in this domain. Hybrid models present that integration of deep learning machine and machine learning methods is intended to achieve the advantages of different classifiers. Of other strategies, ensemble learning also significantly advances multiple solutions for increased reliability and less classification errors. Furthermore, optimization solutions such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and bio-inspired methods are engaged in the IDS framework for feature reduction and parameter optimization. These kinds of strategies do help detect maximally more intrusion attempts, with less FPs, and with a scaling model.

An innovative advancement in intrusion detection research is the integration of Explainable Artificial Intelligence (XAI). Most AI-based intrusion detection systems operated with a black-box mode, making it very hard for analysts to make sense of how the decisions were made. Explainable AI models like LIME, SHAP, and attention mechanisms allow easier interpretation of model predictions and foster additional trust, transparency, and accountability within the cybersecurity system. XAI-equipped IDS frameworks assist cybersecurity professionals to discern the causes of attack and deploy well-informed responses. IDS technologies are extensively employed in various domains like cloud computing, IoT, Industrial IoT, healthcare, autonomous vehicles, smart cities, wireless sensor networks, financial systems, to name a few. While IDS solutions within IoT realm fend off unauthorized access and malware attacks on connected devices, in cloud computing, IDS applications secure virtual machines, data centers, cloud communication infrastructures, etc. [15], from various cyber threats. Industrial IoT along with smart manufacturing systems are relying on IDS technology to secure industrial automation and control systems from a wide range of cyber threats. Healthcare applications are protected with the help of IDS from medical records, medical devices, and telemedicine communication platforms. Similarly, IDS and Vehicular Ad Hoc Networks (VANETs) are employed to secure vehicle-to-vehicle communication and transportation safety from a number of threats.

While intrusion detection has made large strides in technology, several challenges still need to be addressed. It is well documented that high computational complexity, increased false alarm rates, scalability limitations, lack of understanding of new kinds of attacks, and inefficient real-time detection capability often characterize existing IDS models. Most AI-related intrusion detection approaches support deep learning that requires massive training datasets, which could only run on high-performing hardware. Therefore, explainability and interpretability still represent great challenges in the use of artificial intelligence in IDS systems. Last but not least, advanced threat actors are now creating a new set of security threats against modern IDS frameworks. As such, researchers are relentlessly seeking out modern AI-driven cyber-security

systems that can provide accurate, killing, scalable, lightweight, and interpretable intrusion detection. The mixed ensemble models that tune themselves via explainable AI concepts are constantly viewed as viable solutions to the problems with the contemporary IDS systems. The results would be improvement in the detection efficiency of attack, in addition to increasing transparency and expectedly establishing more robust and tighter cybersecurity [16]. This would lead to more efficient management in secure environments where the new type of data warfare will operate.

## III. MACHINE LEARNING TECHNIQUES FOR INTRUSION DETECTION

Machine Learning (ML) techniques have become very imperative in quite a number of modern Intrusion Detection Systems (IDS) as a solution to automatically learn patterns of network behavior and determine the efficiency of various malicious activities involved. Rule-based traditional intrusion detection methods have often been ineffectual in fighting modern sophisticated cyber threats and are quite inappropriate in dealing with an evolving threat. These traditional mechanisms find their focus primarily in the rule application for determining the idea of an attack, making detection even harder for those digital objects. Machine-learning algorithms thus transcend these limitations by analyzing historical network traffic data, extracting features that are helpful in evaluating which activities are normal or malicious. The MLS turned out to be very adaptive in terms of their detection accuracy as well as scalability, thereby delivering clear superiority over the older traditional cybersecurity strategies. Henceforth, machine learning techniques are heavily applied in cloud computing, IoT, IIoT systems, health, WSNTs, and smart city scenarios. Intrusion detection system (IDS) has gained tremendous relevance across the globe to detect cyber threats to computer networks. Security experts divide IDS methods into supervised, unsupervised, semi-supervised, and reinforcement learning methods [17]. Mostly used among these is supervised learning methods because they are backed by labeled datasets on which learning and classification are based. Some common supervised algorithms are Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), and K-Nearest Neighbor (KNN). These ML algorithms use features over network traffic to perform the classification task on cyberattacks with a very high degree of accuracy. Random Forest and Decision Tree provide short processing times and are able to handle high-dimensional datasets in the cybersecurity environment which can in turn predict the best model algorithm before assigning their weights. Likewise, the Support Vector Machine helps in categorizing this stunted performance in classifying binary intrusion detection tasks by finding optimal hyperplanes between attack and non-attack classes.

Typically, unsupervised methods are employed to detect anomalies when labeled cybersecurity data are not at hand. Clustering techniques, such as the K-means algorithm and DBSCAN, aim at grouping similar patterns of behavior in a network to identify unique patterns of abnormality. These methods are often efficient with unknown and zero-day attacks because they do not depend on predefined attack signatures; yet unsupervised methods could also deliver a higher false-positive rate due to aberrations in network traffic patterns. For the improvement of learning performance in changing dynamic environmental scenarios in cybersecurity, Feature selection and dimension reduction are indeed a core-genome in machine learning-based intrusion detection systems. SNS datasets often have scores of features that are either irrelevant or redundant [18]. This unnecessary increase in the number of features would create unwanted computation times and lower detection accuracies. Feature-selection methods like Information Gain, Principal Component Analysis (PCA), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) are so vigorously concerned about placing an essential fit in the bag of features that will proceed the classification rate to come. The improved implementation of pairing real data and associate feature engineering in averaging the expansion of attack detection accuracy will significantly reduce processing overhead on resource-constrained environments like IoT and edge networks.

Some recent studies show the efficiency of some machine learning techniques in the field of intrusion detection. ML implementation on healthcare systems, cloud computing, and industrial IoT, as well as in wireless communication systems and autonomous vehicles, had successfully detected cyber threats efficiently. Ensemble learning methods implement the combining of various machine learning classifiers, allowing for superior performance than single models. These learned methods, including bagging, boosting, and voting classifier, increase robustness, decrease false alarms, and help improve generalization. Machine learning based intrusion detection systems has presented substantial evolution but which are still marred by a host of challenges [19]. Many algorithms have been found to struggle around with myriads of issues including class imbalances, dynamic attack patterns, and large-scale real-time traffic analysis. Supervised learning methods have particular prerequisites for rolling with high-quality labeled data, a real issue for most practical cybersecurity diagnosis; it is considerable to argue. It is a fact that most of the ML-based IDS are not interpretable, hence causing a challenge for cybersecurity professionals. Over and above that, there is an issue with respect to the computational aspect that impacts the deployment of these IDSs in the real world, not to mention the issue of scalability.

## IV. DEEP LEARNING APPROACHES IN CYBERSECURITY

Deep Learning (DL) has emerged as one of the most powerful technologies for intrusion detection and cybersecurity applications due to its capability to automatically learn complex patterns from large-scale network traffic data. Unlike traditional machine learning algorithms that require manual feature engineering, deep learning models can automatically extract high-level features from raw data and improve attack detection performance. With the rapid increase in cyber

threats, IoT devices, cloud computing systems, and smart infrastructures, deep learning-based intrusion detection systems have gained significant attention in modern cybersecurity research.

Especially good for its feature recognition ability, (CNN) has turned into a common deep learning model for intrusion detection purposes. CNN models thus analyze a range of network traffic patterns more effectively-papering malicious activities in a more accurate manner. CNN-based IDS frameworks have been implemented in IoT, healthcare, industrial automation, and wireless communication systems to boost cyberattack detection. Besides having such effectivity. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) networks are shared below in this context. They are distinguished from each other where, they deal respectively with sequential netowrk traffic behavior and time series cybersecurity data. The LSTM model captures long-term dependencies in patterns of communication and detects sophisticated cyber threats, such as distributed denial-of-service attacks and propagation of malware. Hybrid deep learning architectures of the typifies of Bi-LSTM have furthered on improving the performance of intrusion detection by recalling the combination of spatial and temporal future extraction mechanisms. Autoencoders also find extensive use in breach detection and dimensionality reduction on cybersecurity systems. Further, these models are trained to understand compressed representations of normal network activity and so infer deviations from these representations as potential intrusions. Deep Belief Networks (DBN) and Generative Adversarial Networks (GAN) in addition to advanced attack categorizations and adversarial defense applications are under research.[20]

Studies show that deep learning methods have higher detection accuracy and lower false rates in comparison to conventional Machine learning methods. Their performance lies in the models formed with Hybrid CNN-LSTM, DCNNBiLSTM. Such models, presumed to be working independently of the network, must possess significant levels of accuracy in detecting a diverse assortment of possible security threats or attacks under various test scenarios. This is the consequent learnings in synergy with the IoT intrusion detection systems in cloud computing environments, Industrial IoT Systems, VANETs, and smart city infrastructures [21]. Deep learning models incorporating optimization algorithms like Genetic Algorithm, PSO, and bio-inspired optimization techniques are commonly used to augment feature selection, parameter turning, and training efficiency. Now, federated learning and blockchain techniques have increasingly being leveraged by deep learning IDS frameworks for assuring privacy and secure decentralization in IoT environments. All of these advanced techniques basically aim to move towards the end goal of intelligent distributed cybersecurity systems.

However, concerning their benefits, deep learning-based intrusion detection systems have to overcome many challenges. This necessitates a sizeable labeled dataset. The accompanying training time is often enormous. Thanks to the training model itself, deep learning demands the highest computational resources. As the roaming charges are high for IoT sensors, likely against what they are providing, the question of training complexity, memory consumption, and taxing equates to "whether the use of a solution will ever be practically implemented in the world of resource-acquisition challenge by IoT sensors, with edge computing systems sharing the same foam route in the communication link between execution and transfer in terms of training time due to two issues black-box and explainable AI. So far, it is the best, but as we know well, deep learning is one of AI [22]. The deep learning model gives very little interpretation to the decisions it makes for input instances. This deficiency in interpretability for cybersecurity applications increases mistrust and undermines accountability. Overcoming particular limitations can come in the way of developing resource cognizant deep learning architectures, adversarial robust architectures, and interpretable AI integrated cybersecurity systems. In the direction of explaining and giving context to its predictions, methods for explaining AI, including LIME and SHAP, have been more relevant. Because of that, deep learning remains the most perfect technology in designing future intrusion detection systems, thus achieving enhanced protective effects securely, accurately, in adaptably scalable modes, viz., cybersecurity.

## V. HYBRID ENSEMBLE MODELS FOR INTRUSION DETECTION

Currently, the merging ensembles have become helpful in increasing the accuracy of an intrusion detection system in modish risk management, be it in traditional or modern cybersecurity working domains. This is because machine learning that classically deeply rooted and most commonly accepted techniques of classical artificial neural networks for this applications are prone to overfitting, limited generalization, high false alarm rates, haplessly capturing existing cyber threats, and have failed to replace rapidly evolving threats, which include zero-day threats, evolving polymorphic worms, and so on. They attempt to overcome the mentioned limitations by concentrating on multiple classifiers; intelligent learning techniques, and optimization algorithms that aim to improve detection accuracy, robustness, and scalability for cyber defense or those applications [23].

Ensemble learning is a machine learning strategy whereby multiple models are combined to make predictions that are more reliable and accurate compared to those of individual classifiers. These models use common ensemble techniques such as bagging, boosting, stacking and voting mechanisms. The Random Forest is a popular bagging-based ensemble method that includes multiple decision trees to enhance classification accuracy and decrease overfitting. Boosting techniques like AdaBoost and Gradient Boosting emphasize the enhancement of weak classifiers on an iterative basis for improved performance in intrusion detection. On the contrary, while they are learning, voting classifiers aggregate different models in order to come up with final decisions, which boast a higher degree of reliability. Hybrid ensemble IDS models that put

together machine learning and deep learning algorithms do it to maximize strength. For example, CNN-LSTM architectures would couple spatial feature selection with sequential traffic analysis for beneficial attack-based detection. Likewise, hybrid systems have combined autoencoders, recurrent neural networks, and optimization to improve anomaly detection together with lessening computational complexity. Researchers have proposed binding optimization algorithms like Genetic Algorithm, Particle Swarm Optimization, and Golden Jackal Optimization with ensemble models that will help in feature selection and parameter tuning [24].

They say recent studies show that Hybrid Ensemble Models have achieved better detection accuracy and a very low false positive rate, as against standalone IDS approaches. Ensemble models based on Deep Learning have shown impeccable performance and resilience throughout IoT Networks, Cloud Computing Systems, Healthcare Applications, Autonomous Vehicles, Industrial IoT, and Wireless Sensor Networks. As pertains to the setting of decentralization, federated learning and integrated blockchain-supported ensemble solutions could boost privacy preservation, distributed learning, and wireless security. Feature selection is certainly another significant factor in any hybrid ensemble IDS. High-dimensional cybersecurity datasets are filled with irrelevant as well as redundant features; thus, they negatively affect the classification performance. Hybrid feature selection methods combine statistical analysis, optimization algorithms, and deep learning to extract relevant features and reduce the dimensionality of the dataset. This then saves a lot of time and optical processing versus the conventional methods, hence it is deploy-able in real-time environments.

Also, the Explainable AI is gradually being integrated in hybrid ensemble IDS frameworks for better transparency and trustworthiness. Techniques such as LIME and SHAP venture a layer of explanation for the complex ensemble model predictions and thus facilitate the understanding of intrusion detection decisions by cybersecurity human analysts. Explainability is particularly important in healthcare, financial systems, and industrial infrastructures where decision accountability is necessary [25]. Though hybrid ensemble models can be beneficial, there are a few challenges they face. These challenges include increased architectural complexity, increased computational overhead, and increased training time. Most ensemble IDS models require really large sets of data for training and for deployment, hence requiring high-performance resources. In addition, scalability and real-time adaptability have arisen as major embedding problems in a dynamic network environment. Also, some hybrid models still run as partly black-box systems even with some explicit explainability mechanisms attached to them. Hybrid ensemble techniques, on the other hand, are generally considered as highly promising prospective solutions for future IDS because of their capacity to deliver high accuracy, robustness, adaptability, and resilience in the face of converting cyber threats. Still, including optimization, light architecture, and explainable AI approaches helps in leveraging them to enrich the efficiency and practical applicability of hybrid ensemble IDS frameworks in modern intelligent security environments.

## VI. EXPLAINABLE ARTIFICIAL INTELLIGENCE IN CYBERSECURITY

Explainable Artificial Intelligence (XAI) has emerged as an importantly essential topic of study under cybersecurity, in that there is the increasing adoption of machine learning and deep learning models by intrusion detection systems. While AI-centric cybersecurity frameworks have been currently overly employed for higher detection accuracy and the automation of post-threat analysis, most of them operate as black-box models, rendering them unable to offer sufficient explanations as to why decisions are made. This non-transparency creates issues of trust, accountability, and dependability for cybersecurity analysts, which is especially severe in some critical domain specific areas like healthcare, cloud computing, IIoT, autonomous vehicles, and smart city infrastructures [26]. Several methods can be easily explained in AI in which intrusion detection is concerned. Local Interpretable Model-Agnostic Explanations (LIME) is one of the foremost such techniques used for local explanations of individual predictions. LIME explains model behaviour through approximating more complex black-box models by simpler interpretable models to capture the most important aspects that are influencing the classification decision. Similarly, SHapley Additive exPlanations (SHAP) adopts some game theory concepts for computing the feature importance scores and providing global and local interpretability for machine learning and deep neural network models. Likewise, in the deep learning methodology, concepts like the attention mechanism and visualization methods explain the approach used in cybersecurity systems to signal significant patterns related to attacks. Recent investigations have looked into combining Explainable AI with Hybrid and Ensemble Intrusion Detection Systems to enhance trustworthiness in cybersecurity. Researchers have integrated LIME and SHAP with CNN, LSTM, Random Forest, XGBoost, and Federated Learning frameworks to generate interpretable predictions of attacks. The interpretation mechanisms assist a cybersecurity analyst to check a trainee model for false positives to better understand subjects of attack and to validate the reliability of a model [27]-[28]. In compliance with hospital and industrial problems, interpretability in intrusion-detection systems aids in improving regulatory compliance and makes decision-making responsibility more traceable.

Explainability of AI is significant in adversarial cybersecurity. Attacks conducted by adversaries manipulate AI models by injecting malicious disturbances into input data in order to escape detection by intrusion-detection mechanisms. Therefore, XAI techniques are helpful for detecting suspicious behaviors of models and improving against adversarial manipulations. In addition to these uses of explainability, bias detection, fairness evaluation, and the secure deployment of AI-backed crime-cyber security frameworks are also accomplished. However, despite the above advantages, explainable AI in

cybersecurity is faced with various challenges. With many explainability methods, the computational overhead and processing times are increased, which affects real-time performance in intrusion detection. And while some XAI approaches are able to provide only local interpretability, they do not explain the behavior exhibited by the model as a whole [29]. The question of how to strike a balance between detection accuracy and interpretability remains a major research challenge, hence one is again confronted with the fact that very accurate deep learning classifiers are less transparent. Scalability and integration complexity are also the largest confounding factors in the practical implementation of any idea in a large-scale network.

Researchers are now searching for lightweight, scalable, and adaptive XAI-based frameworks for intrusion detection systems. Recent solutions from hybrid explainable ensemble models, federated XAI learning, and attention-based architectures under deep learning arise as promising solutions for any cybersecurity application in the future. If explainable AI can be integrated with machine learning, deep learning, blockchain, and optimization techniques, it will contribute significantly towards advancing cybersecurity by transparency and attack analysis and improving decision making.
Indeed, explainable AI now forms an integral part in intelligent ID systems [30]. It bridges the gap between the high-performance incredible AI models and meaningful and understandable forms, providing transparency, trust, and accountability in order to readily generate secure, reliable, and efficient cybersecurity systems for safeguarding advanced digital infrastructures from crucial cyber threats. Table 1 presents a comparative review of recent intrusion detection studies based on machine learning, deep learning, hybrid ensemble techniques, datasets, performance outcomes, and research limitations in cybersecurity applications.

**Table 1: Literature Review of Machine Learning and Deep Learning-Based Intrusion Detection Systems**

| Ref | Technique/Model Used | Dataset/Environment | Key Result | Limitation |
|---|---|---|---|---|
| [1] | ML-based anomaly IDS | Wireless Sensor Networks | Improved anomaly detection accuracy | High false positives in dynamic traffic |
| [2] | AI and ML techniques | Information security systems | Enhanced intelligent attack detection | Limited scalability evaluation |
| [3] | Survey on ML strategies for IDS | Multiple IDS environments | Comprehensive comparison of ML methods | No practical implementation |
| [4] | Golden Jackal Optimization + DL | Network security datasets | Improved feature optimization and accuracy | Increased computational complexity |
| [5] | IDS with HCI integration | Smart city systems | Enhanced attack mitigation | Complex deployment architecture |
| [6] | Hybrid lightweight DL IDS | Network traffic environment | Reduced processing overhead | Lower interpretability |
| [7] | Hybrid DL and ML approach | SDN-IoT healthcare systems | Improved healthcare network security | High training complexity |
| [8] | VAE-XGBoost hybrid IDS | EV charging networks | High detection rate with low false alarms | Complex hybrid architecture |
| [9] | Deep learning-based IDS | IoT networks | Enhanced attack classification accuracy | Requires large training data |
| [10] | Federated deep learning | IoT networks | Improved privacy-preserving intrusion detection | Communication overhead |
| [15] | CNN-LSTM hybrid model | Network intrusion datasets | High classification performance | Long training time |
| [16] | Deep learning model transposition | Network IDS systems | Improved transfer learning capability | Generalization limitations |
| [17] | Spatial-temporal DL model | Autonomous vehicle CAN networks | Effective vehicle intrusion detection | Resource-intensive model |
| [18] | DCNNBiLSTM hybrid DL model | Intrusion detection datasets | Improved attack detection accuracy | Increased model complexity |
| [19] | ML and DL evaluation | IDS benchmark datasets | Comparative performance analysis | Limited real-time validation |
| [20] | Lightweight ML algorithms | TON_IoT and IIoT datasets | Reduced resource usage | Lower efficiency for advanced attacks |
| [21] | ML-based NIDS | VANET environment | Improved vehicular security | Mobility and scalability issues |

| | | | | |
|---|---|---|---|---|
| [22] | RNN-based IDS framework | Network intrusion datasets | Better sequential attack analysis | Requires high computational power |
| [26] | Hybrid adaptive ensemble IDS | Cybersecurity network datasets | Achieved >98% detection accuracy | High computational overhead |
| [27] | GA-optimized DL + LIME | IoT networks | Explainable and accurate IDS | Optimization complexity |
| [28] | DL-based IDS | Mobile Ad-hoc Networks | Improved MANET security | High energy consumption |
| [29] | Blockchain-ML hybrid framework | Industrial IoT | Secure intelligent intrusion detection | Increased latency |
| [30] | ML models for IDS | Cybersecurity systems | Enhanced attack detection efficiency | Lack of explainable AI support |

**References**

[1] Al-Fuhaidi, Belal, et al. "Anomaly-Based Intrusion Detection System in Wireless Sensor Networks Using Machine Learning Algorithms." *Applied Computational Intelligence and Soft Computing* 2024.1 (2024): 2625922.
[2] Ali, Abdul Karim Sajid, et al. "Intelligent Intrusion Detection and Data Protection in Information Security Using Artificial Intelligence and Machine Learning Techniques." *Spectrum of Engineering Sciences* (2025): 818-828.
[3] Ali, Ali Hussein, et al. "Unveiling machine learning strategies and considerations in intrusion detection systems: a comprehensive survey." *Frontiers in Computer Science* 6 (2024): 1387354.
[4] Aljehane, Nojood O., et al. "Golden jackal optimization algorithm with deep learning assisted intrusion detection system for network security." *Alexandria Engineering Journal* 86 (2024): 415-424.
[5] Alnuaim, Abeer. "Intrusion Detection and Security Attacks Mitigation in Smart Cities with Integration of Human-Computer Interaction." *Computers, Materials, & Continua* 86.1 (2026): 1.
[6] Altaie, Rusul H., and Haider K. Hoomod. "An intrusion detection system using a hybrid lightweight deep learning algorithm." *Engineering, Technology & Applied Science Research* 14.5 (2024): 16740-16743.
[7] Arthi, R., S. Krishnaveni, and Sherali Zeadally. "An intelligent SDN-IoT enabled intrusion detection system for healthcare systems using a hybrid deep learning and machine learning approach." *China Communications* 21.10 (2024): 1-21.
[8] Asim, Muhammad, et al. "VAE-XGBoost: a hybrid intrusion detection system for next generation EV charging networks." *PeerJ Computer Science* 12 (2026): e3506.
[9] Awajan, Albara. "A novel deep learning-based intrusion detection system for IOT networks." *Computers* 12.2 (2023): 34.
[10] Belarbi, Othmane, et al. "Federated deep learning for intrusion detection in IoT networks." *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023.
[11] Benmalek, Mourad, and Abdessamed Seddiki. "Particle swarm optimization-enhanced machine learning and deep learning techniques for Internet of Things intrusion detection." *Data Science and Management* (2025).
[12] Darla, Swathi, et al. "An optimized Bi-LSTM with deep learning-based intrusion detection system in healthcare using Blockchain." *Information Security Journal: A Global Perspective* 35.1 (2026): 140-153.
[13] Devendiran, Ramkumar, and Anil V. Turukmane. "Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy." *Expert Systems with Applications* 245 (2024): 123027.
[14] Dong, Huiyao, and Igor Kotenko. "A lightweight and robust approach to IoT intrusion detection based on ensemble deep learning." *International Journal of Parallel, Emergent and Distributed Systems* (2026): 1-21.
[15] Du, Jiawei, et al. "NIDS-CNNLSTM: Network intrusion detection classification model based on deep learning." *IEEE Access* 11 (2023): 24808-24821.
[16] Figueiredo, João, Carlos Serrão, and Ana Maria De Almeida. "Deep learning model transposition for network intrusion detection systems." *Electronics* 12.2 (2023): 293.
[17] Ghamri, Maroua, et al. "Intrusion detection in CAN protocol using novel spatial-temporal deep learning model for secure autonomous vehicles communication." *Annals of Telecommunications* (2026): 1-24.
[18] Hnamte, Vanlalruata, and Jamal Hussain. "DCNNBiLSTM: An efficient hybrid deep learning-based intrusion detection system." *Telematics and Informatics Reports* 10 (2023): 100053.
[19] Hossain, Md Sabbir, et al. "Performance evaluation of intrusion detection system using machine learning and deep learning algorithms." *2023 4th international conference on big data analytics and practices (IBDAP)*. IEEE, 2023.
[20] Ismail, Shereen, Salah Dandan, and Ala'A. Qushou. "Intrusion detection in IoT and IIoT: Comparing lightweight machine learning techniques using TON_IoT, WUSTL-IIOT-2021, and EdgeIIoTset datasets." *IEEE Access* 13 (2025): 73468-73485.

[21] Juluri, I. Ravi Prakash Reddy Akshaya, Sanjana Reddy, and Pola Rishika. "Machine Learning: Securing VANETS with Machine Learning-Based NIDS Approach." *Proceedings of 6th International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2025, Volume 1*. Vol. 1. Springer Nature, 2026.
[22] Kasongo, Sydney Mambwe. "A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework." *Computer Communications* 199 (2023): 113-125.
[23] Kulshrestha, Priyesh, and T. V. Vijay Kumar. "Machine learning based intrusion detection system for IoMT." *International Journal of System Assurance Engineering and Management* 15.5 (2024): 1802-1814.
[24] Ledwoń, Anna, and Marek Natkaniec. "Machine Learning for Wi-Fi Intrusion Detection: A Comparative Study of Accuracy, Explainability, and Adversarial Robustness." *IEEE Access* 14 (2026): 7582-7599.
[25] Li, Jing, et al. "Optimizing IoT intrusion detection system: feature selection versus feature extraction in machine learning." *Journal of Big Data* 11.1 (2024): 36.
[26] M. Ahmed, A. Alshamrani, S. Alqahtani, and M. A. AlZain, "HAEnID: A Hybrid Adaptive Ensemble Intrusion Detection Framework for Cybersecurity Using Machine Learning Techniques," Journal of Cloud Computing, vol. 13, no. 150, pp. 1–34, 2024.
[27] Maseno, Elijah M., Yanxia Sun, and Zenghui Wang. "GA-optimized deep learning intrusion detection framework with LIME explainability for IoT networks." *Evolutionary Intelligence* 19.1 (2026): 23.
[28] Meddeb, Rahma, et al. "A deep learning-based intrusion detection approach for mobile Ad-hoc network: R. MEDDEB et al." *Soft Computing* 27.14 (2023): 9425-9439.
[29] Mortada, Zaid Mohammed, and Ola Baqer Abdulhadi. "Hybrid Blockchain–Machine Learning Framework for Secure and Intelligent Intrusion Detection in Industrial Internet of Things." *CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES* 7.1 (2026): 278-289.
[30] Mukil, S., et al. "Intrusion detection system using machine learning models." *Business Resilience and Business Innovation for Sustainability: The Double-Edged Role of Artificial Intelligence and Other Disruptive Technologies*. Cham: Springer Nature Switzerland, 2026. 2125-2139.