IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A THOROUGH ANALYSIS OF THE IOT NETWORK'S INTRUSION DETECTION SYSTEM

Megha.M. Shete, Anuradha

Department of Computer Science and Engineering,
P.D.A College of Engineering, Kalaburagi, Karnataka, India

Abstract— The rapid growth of Internet-connected devices and modern communication networks has significantly increased cybersecurity threats. Traditional Intrusion Detection Systems (IDS) are often ineffective against sophisticated and zero-day attacks due to their dependence on static signatures and predefined rules. Recent developments in machine learning, deep learning, and reinforcement learning have significantly improved the capability of intrusion detection systems to identify both known and unknown network attacks. This literature review presents a comprehensive analysis of recent research works related to network intrusion detection systems, emphasizing machine learning, deep learning, explainable AI, ensemble learning, and reinforcement learning approaches. The review identifies current challenges, strengths, and research gaps that motivate the development of a Multi-Agent Reinforcement Learning-based Network Intrusion Detection System.

Keywords— Internet, cybersecurity, Intrusion Detection Systems (IDS), Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and Reinforcement Learning (RL)

I. INTRODUCTION

An extensive growth of the IOT which includes industrial IOT is ensuring connected systems are now part of almost all crucial infrastructure, cities in smart ways and people's everyday life while maintaining digital connectivity with growing numbers of connected devices; this global change has created extensive opportunities to access and compromise online and digital systems via connectivity allowing growth of cybercrime opportunities in unprecedented ways. Traditional perimeter protections are not capable of preventing many of the increasingly advanced sophisticated cyberattacks targeting networks worldwide. IDS systems (intrusion detection systems) have become key solution technologies for continuously identifying the actions of malicious actors on networks. Recent evolution has transitioned from fixed (signature-based systems) to adaptive Intelligence based formally defined solutions supporting adaptive AI. New systems now utilize ML, DL, and multiple heuristics for detecting anomalous network traffic" patterns. As the use of complex AI developed intelligent models will help to build large amounts of "high dimensional data generated from the network over time; however, as these much larger and complex AI based models process more data using new techniques or methodologies to identify anomalies, currently, they operate with minimal visibility (A "black box") so that users do not know why an alert was generated if the computer model was correct or not.

The primary objective of this synthesis is to track and critically evaluate the evolutionary trajectory of machine learning and deep learning architectures deployed within both standard Internet of Things (IoT) and Industrial IoT (IIoT) ecosystems. Specifically, we examine state-of-the-art methodologies engineered to optimize detection metrics, alleviate heavy computational overhead, and demystify historically opaque algorithms. The analysis further explores the diverse integration of advanced optimization routines,

defense-in-depth frameworks, and dimensionality reduction strategies within modern IDS implementations. Concluding this evaluation is an assessment of the paradigm shift toward Explainable AI (XAI). We investigate how these interpretability layers are woven into predictive models to foster institutional trust, establish operational accountability, and deliver human-centric diagnostics—all while preserving the raw accuracy of the core detection engine.

The research done in this review will cover literature on advanced academic research on Network Security through 2021-2026. This “literature has been categorized into” three main areas. A review of single-layer and multi-layer Deep Learning (DL) architectures using Nature Inspired Optimization (NIO) algorithms, ensemble learning techniques to increase the effectiveness of handling high-dimensional data in networks, and the creation of Explainable Artificial Intelligence (XAI) frameworks to produce comprehensible complex models are some of these areas. This research focuses on the particular operating characteristics and environment of IoT, IIoT, and Edge Computing settings, which face severe resource limitations (e.g., limited memory, processing power, bandwidth). The evaluation structure of this study will be accomplished by meeting three research objectives:

To determine whether hybrid optimization techniques (algorithms) combined with multi-layer structure configurations improve the accuracy of IoT/IIoT systems for detecting anomalies, to investigate how using sophisticated feature selection methods in conjunction with ensemble learning frameworks reduces high signal density brought on by high network traffic volumes, hence lowering the frequency of false positives and to evaluate how an intrusion detection system (IDS) framework can use explainable artificial intelligence (XAI) to create a balance between transparency and predictive modeling. The IoT era, it is extremely important to secure network traffic for both global economic stability and personal data privacy. Current IDS have high false positives and latency issues, putting critical networks at risk of zero-day attacks and sophisticated advanced persistent threats. This study will look at how XAI, deep learning, and ensemble optimization can address this cybersecurity engineering issue. Resolving these issues will enable the efficient implementation of transparent, lightweight, and extremely accurate defense mechanisms. In the end, this study will provide useful suggestions for safeguarding ecosystems with limited resources and establishing the necessary human-in-the-loop trust for protecting next-generation digital infrastructure.

II. Literature Review

The Internet of Things' (IoT) rapid development has raised a number of new security and safety concerns. IoT devices lack sufficient information sharing capabilities and the capacity to employ conventional security measures. As a result, there are several sophisticated assaults taking place within IoT ecosystems. To address this issue, numerous researchers are working to create Intelligent Intrusion Detection Systems (IDS) using Machine Learning (ML) and Deep Learning (DL) techniques.

Author [1], developed an IDS that is based on Deep Learning, combined with Enhanced Transient Search Optimization (ETSO) for feature selection and classification purposes. This system produced better than average detection accuracies and fewer false alarms, which illustrates the usefulness/effectiveness of using optimization techniques to assist in deep learning methods.

An intelligent two-layer IDS architecture was developed by author [2], especially for Internet of Things settings. The second layer categorizes assault kinds, while the first layer detects anomalies. This combination improved detection and increased the system's scalability.

A group of researchers led by author [3], proposed a network intrusion detection system that used an ensemble-based framework. Their technique employed multiple classifier models for classification and showed considerable improvement in detecting intrusions with reliable results.

Using ensemble learning approaches and dimensionality reduction strategies, by author [4], created an Industrial Internet of Things (IIoT) edge computing intrusion detection system with Random Forest classifiers. In comparison to numerous other techniques, their results also showed better detection outcomes.

Author [5], investigated the application of ensemble learning techniques to improve network security by optimizing several classifiers. They discovered that the false-positive rate was significantly lower when multiple classifiers were combined than when they were used separately.

Synthetically, these investigations confirm that when integrated alongside complementary paradigms, ensemble learning methodologies fundamentally enhance both the baseline reliability and overall robustness of contemporary intrusion detection systems, while drastically dampening generalization variance. Yet, despite the stellar detection accuracies achieved by standard machine learning and deep learning models, their practical deployment is routinely hindered by a persistent lack of algorithmic transparency. This pervasive "black box" limitation has driven the recent, intense research focus on

Explainable AI (XAI) frameworks as a vital mechanism for decoding and auditing complex IDS classifications.

In [6], authors examined how XAI techniques can be applied to analyze and explain the outputs of the IDS. The study showed that explainable AI increases trust, accountability and the transparency of the model's results.

For IoT networks, author [7], developed an IDS based on machine learning. According to the study, supervised machine learning algorithm classifiers can effectively and efficiently detect fraudulent activities without requiring a significant amount of computing power. The combination of many classifiers using ensemble learning has been shown to be an effective means of enhancing the performance of intrusion detection systems.

Author [8], addressed these visibility gaps by pioneering a hybrid ensemble architecture infused with explainable AI specifically tailored for network anomaly detection. Their framework not only yielded robust classification accuracy but also delivered human-interpretable rationale directly to security personnel. Building upon a similar philosophy, author [9], developed a hybrid IDS engineered to simultaneously heighten detection performance and model transparency. Their empirical findings demonstrated that the inclusion of XAI prominently equips security analysts with the contextual insights needed to dissect complex attack signatures and rigorously audit operational defense decisions.

Author [10], extended this area by providing a human-based framework for explainable deep intrusion detection systems. Their research focused on interpretable user outcomes and improving security through transparent decision-making methods.

Researchers have also explored the use of a multi-layered architecture for progressively filtering threats. The authors in [11], proposed an intelligent two-layer IDS that decouples the detection tasks

Recent studies further support that a distributed ensemble architecture based on fog computing technology will provide closer proximity to the network edge for threat detection capabilities, as evidence continues to emerge through the work of paper [12], the authors of paper [13] and [14] with a much broader analysis compiled around the use of ensemble averaging techniques and ensemble deep neural networks for various network traffic types and heterogeneous botnets.

Although deep and ensemble models provide high levels of accuracy, the ability to apply them to critical infrastructures presents challenges related to trust and deployment due to their "black-box" nature. As a result, the most recent paradigm shift is towards transparency through Explainable AI (XAI). A comprehensive survey of the opportunities and architectural designs available for implementing XAI into cybersecurity, a major research initiative conducted a thorough examination of how XAI frameworks decode complex Intrusion Detection Systems (IDSs) internal decision-making processes [15], was overseen by several authors in recent year. Practical implementations of hybrid learning and XAI have progressed rapidly over recent years.

The author [16], contributed to the work of previous author, with regards to this baseline, by providing an example of how refined ML-related technology could promote a high level of security relative to the traditional boundaries of IoT networks.

With the continuing sophistication of cyber threats, there has been a growing interest among researchers in utilizing Deep Learning (DL) methods to model complex, non-linear attack patterns.

The work proposed by author [17], integrated hybrid ensemble learning with XAI frameworks to improve transparency for detecting abnormalities in networks. Similarly, author of paper [18], successfully integrated XAI into their hybrid IDS providing both increased classification accuracy and enhanced model transparency. Most recently, while research of paper [19], developed a human-centered XAI deep intrusion detection framework to create a collaborative environment for security analysts to easily interpret and respond to automated deep learning alerts. Recently, the paper [20], extended the concept of sequential feature extraction by integrating Pearson correlation with a hybrid Deep Neural Network (DNN) and Transformer architecture, demonstrating improved performance when modelling long-range dependencies in IIoT traffic.

Table 1.Summary of Literature Survey Analysis

Ref No	Year	Title	Authors	Methodology (Dataset + Approach)	Accuracy / Results
[1]	2020	Scalable Machine Learning-Based Intrusion Detection System for IoT-Enabled Smart Cities	M.A. Rahman et al.	IoT-Botnet 2020 Dataset + Machine Learning (DNN, RF, SVM, etc.)	99.95% detection rate for abnormal flows; 96.085% overall accuracy; 0.57–2.6% increase in model accuracy vs. other methods
[2]	2021	IoT Intrusion Detection System Using Deep Learning and Enhanced Transient Search Optimization	A. Fatani, M. Abd Elaziz, A. Dahou, M.A.A. Al-Qaness, and S. Lu	KDD-CUP & NSL-KDD Datasets + D eep Learning (CNN) + Enhanced Transient Search Optimization (TSO)	99.47% accuracy (CNN with TSO on Bot-IoT dataset); Enhanced accuracy on KDD-CUP & NSL-KDD
[3]	2022	Dependable Intrusion Detection System for IoT: A Deep Transfer Learning-Based Approach	S.T. Mehedi et al.	Custom IoT Dataset + Deep Transfer Learning (P-ResNet)	87% overall accuracy; 88% precision; 86% recall; 86% F1-score; 83% ROC AUC; Low computational complexity
4	2021	A Distributed Ensemble Design Based Intrusion Detection System Using Fog Computing to Protect the Internet of Things Networks	P. Kumar, G.P. Gupta, and R. Tripathi	NSL-KDD, UNSW-NB15, AWID Datasets + Distributed Ensemble + Fog Computing + Stacked Autoencoders + Transformer-CNN-LSTM	99.7% accuracy (NSL-KDD multi-class); 99.5% (NSL-KDD binary); 99.16% (UNSW-NB15); 99.9% (AWID); 99% precision, recall, F1
5	2021	Ensemble Detection Model for IoT IDS	A. Alhowaide, I. Alsmadi, and J. Tang	IoT Network Dataset + Ensemble Learning	Not explicitly stated in reference; Ensemble-based detection model for IoT
6	2022	A New Ensemble-Based Intrusion Detection System for Internet of Things	A. Abbas et al.	NSL-KDD, UNSW-NB15 Datasets + Ensemble Learning (Stacking: KNN, GNB, RF) + Double Feature Selection	Up to 99.96% accuracy; 0.007% false alarm rate; Detection time <13 seconds on CIC-IDS2017 & CIC-DDoS2019
7	2023	Random Forest-Based IDS for IIoT Edge Computing Security Using	M. Mohy-eddine, S. Benkirane, A.	NF-UNSW-NB15-v2, Bot-IoT Datasets + Random Forest + Isolation Forest +	99.30% accuracy (PCC-IF strategy); 99.18% accuracy (IF-

		Ensemble Learning for Dimensionality Reduction	Guezzaz, and M. Azrour	Pearson Correlation Coefficient (PCC)	PCC strategy) on NF-UNSW-NB15-v2
8	2023	An Adaptable Deep Learning-Based Intrusion Detection System to Zero-Day Attacks	M. Soltani et al.	CIC-IDS2017, CSE-CIC-IDS2018 Datasets + Deep Learning (Autoencoder + GMM + Gate CNN) + Adaptive Thresholding	98.68% overall accuracy (adaptive threshold); 87.47% detection rate for zero-day attacks at 0.19% FPR; 97.69% accuracy (zero-day); 98.37% (integrated cascade)
9	2023	Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions	N. Moustafa et al.	NSL-KDD, UNSW-NB15, TON-IoT Datasets + LSTM + SPIP Framework (SHAP, PFI, ICE, PDP)	High detection accuracy with interpretability; SPIP framework for global & local explanations; Improved transparency in cyber defense
10	2023	A Hybrid CNN+LSTM-Based Intrusion Detection System for Industrial IoT Networks	H.C. Altunay and Z. Albayrak	UNSW-NB15, X-IIoTID Datasets + Hybrid CNN+LSTM	93.21% accuracy (UNSW-NB15 binary); 92.9% (UNSW-NB15 multi-class); 99.84% (X-IIoTID binary); 99.80% (X-IIoTID multi-class)
11	2023	An Intelligent Two-Layer Intrusion Detection System for the Internet of Things	M.M. Alani and A.I. Awad	NSL-KDD, UNSW-NB15 Datasets + Two-Layer System (Packet-based + Flow-based)	99.15% accuracy (packet-based); 99.66% accuracy (flow-based)
12	2023	Ensemble-Based Approach for Efficient Intrusion Detection in Network Traffic	A.M.T.H. Al-Hayani et al.	Network Traffic Dataset + Ensemble Learning	Efficient intrusion detection via ensemble approach; Performance metrics not explicitly stated in reference

III. RESEARCH GAP

According to the study conducted by authors in [1], their work tackles scalability in IoT-enabled smart cities with machine learning. Still, there's room for improvement. Standard ML models tend to lag when handling big, fast-changing data streams in real-time think massive smart city setups. Also, relying on a central ML framework poses issues can create a single failure point which means shuffling tons of raw data between the cloud and edge devices, which further adds latency. To handle these hiccups, the authors could move toward a decentralized or federated learning setup. By integrating lightweight anomaly detection right at the edge like on IoT gateways we can slash down that data processing time. Plus, adding a global model aggregator keeps everything scalable without sacrificing performance. This way, the system processes data quicker, reduces that pesky communication delay, and overall gets more dependable.

Author [2], found that using deep learning along with Transient Search Optimization improves how well features are selected and how accurately things are detected. However, the downside of it is that these optimization methods require a lot of computing resources during training and tuning. Because of this, these

systems can't keep up with real-time changes, which means they aren't suitable for IoT devices that need fast updates when new threats show up. To fix this, researchers suggest tweaking the optimization procedure. They propose a multi-stage approach where the heavy lifting—metaheuristic search—is done offline with historical data. Then, for real-time tasks, the system relies on speedy methods like gradient-based or lightweight heuristics. Also, adding early-stopping rules in the TSO loop saves comp time by skipping unneeded cycles, especially when dealing with not-so-crucial features.

Using deep transfer learning for securing IoT networks as observed from work of author [3], lets models use info from one area in another yet, a big issue is negative transfer. It strikes up when the starting and ending IoT areas don't line up well enough. Then, the pre-trained parts create a hurdle in the network's performance on the novel things, because of all those different device rules and actions. To fix it, experts suggest adding adversarial domain adaptation methods or domain-adversarial neural networks (DANN). These techs tweak the model to focus on finding things that work across both domains. Essentially, the system figures out how to match what's essential in both areas, keeping the good data aligned and helpful. So, it transfers properly without causing problems.

While studying [4], we observed that operating an ensemble-based Intrusion Detection System in a distributed fog computing environment beefs up security and helps balance network loads. But it also creates complex sync and orchestration problems. Managing deployment, updating individually, and managing voting mechanisms for various models across different fog nodes is tough. It could cause high operational costs and inconsistencies in threat situations that require speedy responses. To fix this, we can use a lightweight container orchestration method like K3s or Edge Kubernetes tailored for fog nodes. Also, an asynchronous ensemble voting protocol helps. This let’s fog nodes send instant local alerts while asynchronously syncing their ensemble weights with nearby nodes. That way, the system stays quick and responsive.

A study conducted by author [5], talks about ensemble detection models in IoT security. While these models combine multiple base classifiers for better performance, they come with big drawbacks—high computation and latency issues. This means that using many machine learning models on each network packet in IoT doesn’t work well because IoT devices have limited power and need to respond quickly. To fix this, the authors suggest creating a smarter ensemble system. It would include a quick preliminary check by a simpler model or a gating network. This initial layer filters out straightforward cases and only flags tricky packets for the full ensemble to handle which means the resource-intensive analysis will only be activated when really needed, making it both efficient and effective.

Author [6], talks about a new way to protect IoT networks using an ensemble method. While this approach does boost detection rates, it still uses fixed combos of base learners. Since IoT networks add new gadgets all the time, a static setup struggles with changes in traffic patterns and user behavior over time. To fix this issue, they suggest the ensemble method needs to learn and change dynamically. Changing classifiers on the fly and including online adaptive learning could be beneficial. The idea is to watch how each classifier performs with rolling checks of network activity. This way, the system can tweak voting weights and swap in fresher models when needed.

In [7], we learned that Ensemble learning for dimensionality reduction, paired with a Random Forest classifier, works well for Industrial IoT edge computing. Yet, leaning solely on random forests and basic methods for reducing dimensions might miss out on the complex patterns in sensor data. These patterns include long-term dependencies that are key to detecting multi-step cyberattacks. To solve this, we could switch out or combine the classifier part with something that understands sequences—like a GRU or TCN. The ensemble method would still handle feature selection to keep things lightweight on the edge devices. After that, passing the filtered features to one of these sequence models ensures not only efficiency but also improved accuracy in spotting the more complex, multi-stage threats in industry settings.

Designing a deep learning system to tackle zero-day attacks is super important, but it's tricky. It is observed that highly sensitive model’s increases false positives were learned from [8]. Normal IoT network changes get caught in the net meant for nasty new threats causing unnecessary downtime. There’s a potential fix though – combine self-supervised contrastive learning with an “human-in-the-loop” active learning process. This approach lets the model cluster unusual activities by behavior rather than a strict binary cutoff. So, similar zero-day oddities bunch together easier. Then, when weird stuff happens, security teams can look at it quickly and add the clear-cut harmless things to the whitelist. With this method, we keep the noise down without sacrificing our ability to catch real dangers.

While deep learning models offer exceptional accuracy in detecting network anomalies, their intrinsic opacity leaves cyber defenders unable to trust, verify, or act upon their outputs during critical incidents is explained by author [9]. This comprehensive review systematically analyzes the intersection of cyber

defense and interpretability to map out how transparency can be introduced into IoT security frameworks and evaluating Explainable Artificial Intelligence (XAI) frameworks, opportunities, and mathematical solutions that translate complex neural decisions into actionable, transparent security insights but fundamental limitation highlighted in the text is the performance-explainability trade-off, where injecting high transparency can often require simplifying the core detection models, thereby degrading absolute threat-classification accuracy.

The research of author [10], focuses on creating an organized vetting process for network packets through architectural partitioning and addresses the underdeveloped ability of current intrusion detection systems (IDS) to effectively deal with the dual challenge of detecting coarse-grained or large-scope attacks (like DDoS) and smaller-scope or highly targeted exploitation attempts. The two-layered IDS described in this dissertation assists with splitting responsibilities between both layers (where one will only identify and filter coarse-grained anomalies while the other will only identify and isolate any specific or subtle exploitation profiles) and therefore makes it easier for those types of attacks to be identified and mitigated as a result of this dual-tiered architecture; however, the potential for a loss of accountability (via a single point of failure) exists at the first layer of the IDS architecture. If an advanced threat can bypass the first layer of detection and evasion (FADE), there is a possibility that the advanced threat will evade detection altogether at the second or deeper level of IDS inspection and evaluation.

In [11], we observed that two-layer Intrusion Detection System is great for dividing up the workload, like sorting known attacks from anomalies or handling different network layers. Yet, if the first layer has high false-negatives or lag issues, it messes up the whole system. Errors just go straight through to the second layer without any chance to fix them. To deal with this, we could make the system smarter by adding a way for the layers to talk back and forth, or use reinforcement learning. If the second layer spots something fishy that it thinks is a mistake, it could adjust the rules of the first layer. This would turn a plain step-by-step process into one that learns and fixes itself on the fly.

While the approach of [12], makes network traffic processing more efficient, the emphasis is mostly on selecting features and pruning models. Sometimes, this means accuracy takes a hit when dealing with those sneaky, rare attacks. Also, there's no special design for handling very imbalanced data where attacks make up less than 1% of traffic. To fix this, we could use advanced synthetic data generation methods like CTGAN or Borderline-SMOTE, but just for the offline training part of the base learners. By also using an ensemble method that prefers cost-sensitive classifiers, we keep the model sharp at catching those rare anomalies without slowing down operations.

The author [13], observes the link between ensemble learning and network security. Yet, it mostly uses static, historical data. The big issue is that it doesn't look at how well the ensemble handles tricky, changing networks and sneaky attacker tactics. Intruders may rearrange packet headers or timing to slip avoid the system's checks. To improve this, the researchers should use adversarial training during development. They could employ tools like the Adversarial Robustness Toolbox (ART) to create fake attack scenarios. The final system improves its ability to identify actual evasion attempts by adjusting network traffic and allowing the ensemble to learn from these examples.

An ensemble averaging DNN as studied in [14] does a great job smoothing out variance for botnet detection across different IoT devices. It's true though, simple ensemble averaging gives the same weight to every model, which kind of misses the point that some DNNs kick serious botnet family butt better than others. To make things more efficient, we could swap out the regular averaging scheme for something smarter like a "stacking" classifier or a dynamic meta-learner. This meta-learner learns how to weigh the predictions of each base DNN based on what they're really good at, boosting the ensemble's ability to nail down those tricky botnets according to the unique traffic signs they leave behind.

The work of author [15], describes how Explainable AI is used in intrusion detection. However, it misses one big thing: its heavy focus on technical explanations that SOC analysts can't use during a breach. You know, stuff like raw SHAP feature values doesn't help someone dealing with a real-time, cyber-attack to fix this issue, we could add a semantic translation layer. It’d sit above the XAI engine and change those complicated metrics into easy-to-understand, helpful info. For example, rather than just seeing a high value in "destination port," the system tells the analyst something like, "A potential brute-force attack is happening on the SSH port. Temporarily isolate IP address X to stop it." Making these changes would really boost how effective SOC teams are at handling threats.

The work of author [16], helps us understand the basic machine learning framework for IoT network security, but there’s a problem. Standard ML-based intrusion detection systems struggle in today's diverse IoT networks because of their performance nosedive with a multitude of different device types. To make matters worse, the current model isn’t modular; changing the network means totally retraining the whole

system, which is costly. To fix this Researchers could upgrade the setup to a modular, cluster-based design. By clustering similar IoT devices—think smart home sensors and IP cameras—and assigning micro-classifiers for each group, we can simplify things. Under a central controller, these clusters handle their own training and retraining, making network changes way easier and less expensive

Combining hybrid ensemble learning with XAI produces accurate and transparent network traffic analysis was observed in [17]. Still, a big issue arises: the mismatch in computing needs. Hybrid ensembles demand a lot of power, and adding post-hoc XAI computations ramps up the processing load, slowing things down too much for speedy enterprise networks. To fix this, we can use a decoupled, parallel processing approach. So, the hybrid ensemble focuses on an optimized data plane for fast threat detection and blocking. Meanwhile, a copy of the data goes to a background control plane where the XAI engine works separately, making sure it doesn't slow down the network. This way, transparency isn't compromised while still keeping high throughput.

The proposed system of [18], combines machine learning methods like Support Vector Machine and Random Forest with deep learning techniques such as CNN, along with SHAP and LIME, to find a balance between performance and interpretability using the CICIDS 2017 dataset. It achieved high accuracy on statistically, but not applicable in real time. Specifically, LIME or SHAP often generate inconsistent or contradictory explanations for similar network events, leaving security analysts to confused. To fix this, researchers could incorporate a consistency-enforcing regularization loss function during training or add an explanation-stabilization wrapper. Connecting the local explanations to a trusted domain-knowledge base would keep the XAI explanations steady, making them reliable for security pros dealing with slight traffic tweaks.

The author [19], uses cutting-edge framework uses a human-centred UI combined with CNN-LSTM feature extraction and SHAP for security pros. Trust and usability get a boost because of this, but there's a catch. While interactive, the "human-centred" design mainly keeps humans on the passive side. The system shows explanations to analysts yet doesn't let them actively correct or tweak the deep learning model dynamically. To fix this, we could add an interactive Explainable Machine Learning (X ML) framework. Security analysts should be able to tweak feature weights or mark problematic explanations right through the UI. These adjustments then get turned into gradients that instantly tweak the CNN-LSTM layers. This way, humans and machines work together more effectively in real-time, creating a strong collaborative cyber defence.

Using Pearson correlation for feature selection with Deep Neural Networks and Transformer architectures really boosts performance in Industrial IoT was learned from author [20]. But the hidden drawback is that Pearson only looks at linear relationships. This means it might overlook important non-linear interactions that cyber threats use to hide themselves. To fix this, we could swap out the linear Pearson method for something more flexible like non-linear mutual information or kernel-based methods such as Spearman's rank and Maximal Information Coefficient. These options pick up on those tricky, non-linear connections. By doing this, we make sure the feature matrices sent to the DNN-Transformer are richer and more representative of the actual data behaviour. This trick helps the whole system spot those sneaky cyber threats better.

IV. DISCUSSION

The literature on IDS for both IoT and IIoT demonstrates tremendous advancement from conventional machine learning methods, to advanced ensemble learning and XAI technologies as it relates to this area.

The basic research of author. [1] demonstrated the effectiveness of traditional ML classifiers for analyzing standard anomaly detection in very large-scale distributed networks such as smart cities. However, one major limitation of these methods was that they relied primarily on shallow feature extraction. Consequently, although the use of computationally efficient and highly scalable statistical models on their own do not perform well on 'novel' and/or 'zero-day' threat variants from a false-positive perspective, they still represent an important advancement in the field.

The author [2], used a combination of Deep Architectures and Transient Search Optimization in order to automate hyperparameter tuning while author [4], recognised that the execution layer for massive cloud-based data-pipelining would place a heavy burden on the network bandwidth and, thus, shifted the execution layer to a Distributed FOG framework

In addition to the significant amount of literature on Deep Learning there has also been an extensive focus on Ensemble Learning as a method of improving the generalization boundary by aggregating learners' results. Earlier ensemble integrations of research by author [5] and [6], have shown that by aggregating a

number of models together, the resulting averages are able to dampen the impact of the variance that can lead to a False Positives.

Author [7] used a combination of Random Forest engines and Ensemble Reduction filters specifically for Edge Resources.

The most significant change in recent literature is the emphasis on model transparency instead of solely raw accuracy numbers. As previously discussed regarding the use of hybrid, deep and ensemble frameworks. The biggest historical limitation of these types of frameworks has always been their black box nature, which has made them very high risk both legally and operationally for critical infrastructure. The previous year study [9], discusses structural requirements for explainability within the context of IDS systems. With this conceptual framework as a guide, current generations of IDS architectures have leveraged Explainable AI frameworks. In support of this conclusion, multi-authored paper to dissect the mathematics of how continuous auditing can be achieved within IDS models of XAI systems

When it comes to optimizing the features for those multi-model clusters and reducing their associated computational footprints, Metaheuristics and Dimensionality Reduction have been used.

The combination of ensembles with local processing has been extensively validated, as evidenced by the works of researcher [12], [13] and [14]. Collectively, these studies have demonstrated that ensemble averaging can securely scale across very heterogeneous and highly geographically distributed Botnet Attack vectors.

Author [16] demonstrated the effectiveness of traditional ML classifiers for analyzing standard anomaly detection in very large-scale distributed networks such as smart cities. However, one major limitation of these methods was that they relied primarily on shallow feature extraction. Consequently, although the use of computationally efficient and highly scalable statistical models on their own do not perform well on ‘novel’ and/or ‘zero-day’ threat variants from a false-positive perspective, they still represent an important advancement in the field.

Immediate architectural integration occurred shortly thereafter as study of paper [17], unified hybrid ensemble architectures with explainable, localized frameworks, and Hemalatha et al. Using a dual-objective method, [18] has produced evidence that increasing the level of transparency in the model does not negatively impact classification accuracy. Finally, the study of paper [19], provide the state-of-the-art with the introduction of a human-centered, explainable AI deep intrusion framework; thereby fundamentally changing the security paradigm from fully autonomous, uninterpretable security blocks to a collaborative and transparent security system that presents timely and intuitive risk explanations based on specific features to security analysts within a defined timeframe.

V. CONCLUSION

Recent advancements in machine learning, deep learning, ensemble learning, explainable AI, and reinforcement learning have transformed intrusion detection systems from static signature-based solutions into intelligent adaptive security platforms. Existing studies demonstrate significant improvements in attack detection accuracy and anomaly identification; however, challenges such as false positives, zero-day attack detection, dynamic adaptation, and explainability remain unresolved which has to be taken as future work for enhancing the security of IDS Network .The literature indicates that integrating Multi-Agent Reinforcement Learning with machine learning and explainable AI provides a promising direction for next-generation intrusion detection systems capable of real-time, adaptive, and transparent cybersecurity defense.

REFERENCES

[1] M.A. Rahman et al., "Scalable Machine Learning-Based Intrusion Detection System for IoT-Enabled Smart Cities," Sustainable Cities and Society, Vol. 61, Art. no. 102324, 2020.

[2] A. Fatani, M. Abd Elaziz, A. Dahou, M.A.A. Al-Qaness, and S. Lu, "IoT Intrusion Detection System Using Deep Learning and Enhanced Transient Search Optimization," IEEE Access, Vol. 9, 2021, pp. 123448–123464.

[3] S.T. Mehedi et al., "Dependable Intrusion Detection System for IoT: A Deep Transfer Learning-Based Approach," IEEE Transactions on Industrial Informatics, Vol. 18, No. 8, 2022, pp. 5623–5634.

[4] P. Kumar, G.P. Gupta, and R. Tripathi, "A Distributed Ensemble Design Based Intrusion Detection System Using Fog Computing to Protect the Internet of Things Networks," Journal of Ambient Intelligence and Humanized Computing, Vol. 12, No. 10, 2021, pp. 9555–9572.

[5] A. Alhowaide, I. Alsmadi, and J. Tang, "Ensemble Detection Model for IoT IDS," Internet of Things, Vol. 16, Art. no. 100435, 2021.
[6] A. Abbas et al., "A New Ensemble-Based Intrusion Detection System for Internet of Things," *Arabian Journal for Science and Engineering*, Vol. 47, No. 2,2022, pp. 1805–1819.
[7] M. Mohy-eddine, S. Benkirane, A. Guezzaz, and M. Azrour, "Random Forest-Based IDS for IIoT Edge Computing Security Using Ensemble Learning for Dimensionality Reduction," International Journal of Embedded Systems, Vol. 15, No. 6, 2023, pp: p. 467.
[8] M. Soltani et al., "An Adaptable Deep Learning-Based Intrusion Detection System to Zero-Day Attacks," *Journal of Information Security and Applications*, Vol. 76,2023, Art. no. 103516, Aug.
[9] N. Moustafa et al., "Explainable Intrusion Detection for Cyber Defenses in the Internet of Things: Opportunities and Solutions," *IEEE Communications Surveys & Tutorials*, Vol. 25, No. 3, 2023, pp. 1775–1807.
[10] H.C. Altunay and Z. Albayrak, "A Hybrid CNN+LSTM-Based Intrusion Detection System for Industrial IoT Networks," *Engineering Science and Technology, an International Journal*, Vol. 38, 2023, Art. no. 101322.
[11] M.M. Alani and A.I. Awad, "An Intelligent Two-Layer Intrusion Detection System for the Internet of Things," IEEE Transactions on Industrial Informatics, Vol. 19, No. 1, January 2023, pp. 683–692.
[12] A.M.T.H. Al-Hayani et al., "Ensemble-Based Approach for Efficient Intrusion Detection in Network Traffic," Intelligent Automation & Soft Computing (IASC), Vol. 45, No. 1, 2023, pp. 651–667.
[13] Baluguri, V. Pasumarthy, I. Roy, B. Gupta, and N. Rahimi, "Optimizing Network Security via Ensemble Learning: A Nexus with Intrusion Detection," Journal of Information Security, Vol. 15, 2024, No. 4, pp. 545–556.
[14] A.A. Wardana et al., "Ensemble Averaging Deep Neural Network for Botnet Detection in Heterogeneous Internet of Things Devices," *Scientific Reports*, Vol. 14, No. 1, 2024, Art. no. 3878.
[15] Multiple Authors, "Use of Explainable Artificial Intelligence for Analyzing and Explaining Intrusion Detection Systems," MDPI – Algorithms, Vol. 14, No. 5, Art. no. 160,2025.
[16] Y. Zhang, "A Machine Learning-Based Intrusion Detection System for Securing Internet of Things Networks," 2025 7th International Conference on Information Science, Electrical and Automation Engineering (ISEAE), Harbin, China, 2025, pp. 374-377.
[17] A. Gupta, A. K. Phulre, A. Patel and R. U. Rahman, "Hybrid Ensemble Learning with Explainable AI for Anomaly Detection in Network Traffic," 2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC), Davangere, India, 2024, pp. 1-8.
[18] A. Hemalatha, V. K. M N, F. T. Graf, A. S. I. T. M, P. Pavithra and R. Suresh, "A Hybrid Intrusion Detection System using Explainable AI for Enhanced Accuracy and Transparency," 2025 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2025, pp. 923-929.
[19] M. M. J. Ayan et al., "Human-Centered Explainable AI for Security Enhancement: A Deep Intrusion Detection Framework," SoutheastCon 2026, Huntsville, AL, USA, 2026, pp. 01-08,
[20] Y. Du, "Industrial Internet of Things Intrusion Detection Based on a Hybrid Model of Pearson-Deep Neural Network and Transformer," Engineering Applications of Artificial Intelligence, Vol. 163, Part 2, Art. no. 111289, 2026.