International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 13, Issue 9, September 2024)

Review of Intrusion Detection based on Machine and Deep Learning Technique for Smart Home IOT Application

[1]Mohini Mishra, [2]Prof. Nitesh Kumar

[1]Research Scholar, Dept. of Electronics and Comm. Engineering, Sagar Institute of Research & Technology, Bhopal, India,
[2]Assistant Professor, Dept. of Electronics and Comm. Engineering, Sagar Institute of Research & Technology, Bhopal, India

***Abstract*— With the rapid expansion of Internet of Things (IoT) devices in smart home environments, security threats have become a pressing concern. Intrusion detection systems (IDS) are crucial for safeguarding these interconnected devices from cyberattacks. In recent years, machine learning (ML) and deep learning (DL) techniques have gained prominence in enhancing the capabilities of IDS. This review provides a comprehensive analysis of existing intrusion detection approaches based on ML and DL, specifically tailored for smart home IoT applications.**

***Keywords*— IOT, Smart Home, IDS, Security, ANN.**

I. INTRODUCTION

The rise of smart home technology, fueled by the proliferation of Internet of Things (IoT) devices, has transformed how people interact with their living environments. From connected thermostats and smart lighting to voice-activated assistants and surveillance systems, IoT devices offer unprecedented convenience, automation, and control[1]. However, as smart home networks become more complex and interconnected, they also become prime targets for cyberattacks. Hackers can exploit vulnerabilities in IoT devices to gain unauthorized access, steal personal data, or disrupt household functions, raising significant privacy and security concerns [2]. In this context, the need for robust and efficient intrusion detection systems (IDS) has become critical.

Traditional security measures, such as firewalls and antivirus software, are often inadequate in protecting smart home IoT environments due to the diverse nature of IoT devices and the unique traffic patterns they generate. Many IoT devices are constrained by limited processing power [3], memory, and bandwidth, making it challenging to implement conventional security mechanisms. Additionally, the sheer number of connected devices and the variety of communication protocols used (e.g., Wi-Fi, Zigbee, Bluetooth) create a highly dynamic and heterogeneous network, complicating the detection of malicious activity. To address these challenges, researchers have turned to machine learning (ML) and deep learning (DL) techniques as promising solutions for enhancing the accuracy and efficiency of intrusion detection in smart homes[4].

Machine learning, a subset of artificial intelligence, involves the development of algorithms that allow systems to automatically learn and improve from experience without being explicitly programmed. When applied to IDS, machine learning algorithms can analyze vast amounts of network data, recognize patterns, and classify network traffic as normal or suspicious [5]. Supervised learning models, such as decision trees, support vector machines (SVM), and k-nearest neighbors (KNN), are commonly used in IDS to detect known attack signatures and anomalies. These algorithms rely on labeled training data to build a model that can distinguish between benign and malicious traffic. However, the effectiveness of these techniques is often limited by the availability of comprehensive and accurately labeled datasets [6][7].

Unsupervised learning, which does not require labeled data, has also been explored in the context of intrusion detection. Algorithms such as k-means clustering and principal component analysis (PCA) are employed to identify abnormal patterns in network traffic without prior knowledge of attack signatures. While unsupervised methods are useful for detecting previously unknown or zero-day attacks, they often struggle with high false-positive rates, as normal traffic fluctuations can be misclassified as intrusions [8]. This limitation highlights the need for more advanced learning techniques capable of adapting to the continuously evolving threat landscape in IoT environments.

Deep learning, an advanced branch of machine learning, has garnered significant attention for its ability to automatically extract complex features from raw data, making it particularly well-suited for intrusion detection in smart home IoT networks. Deep learning models, such as artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), can learn

hierarchical representations of data, allowing them to detect subtle anomalies and improve detection accuracy. Unlike traditional ML models that require manual feature extraction, DL models learn directly from the raw data, capturing intricate relationships that might be missed by human-designed features. As a result, deep learning-based IDS have demonstrated superior performance in detecting both known and novel attacks compared to traditional methods.

In smart home applications, deep learning techniques offer significant advantages over traditional machine learning approaches. Smart home environments are characterized by continuous streams of data generated by IoT devices, ranging from sensors to cameras. Deep learning models, particularly those employing RNNs and long short-term memory (LSTM) networks, are capable of analyzing time-series data, which is crucial for identifying intrusions that unfold over time. Moreover, deep learning models can be trained on large, diverse datasets, allowing them to generalize well to unseen attacks. However, deep learning-based IDS face challenges in terms of computational complexity and resource constraints. Many IoT devices in smart homes lack the processing power and memory needed to run deep learning models locally, necessitating the use of cloud-based or edge computing solutions.

II. LITERATURE REVIEW

F. He et al., [1] An increasingly popular and life-improving technology, the Internet of Things (IoT) has recently been pivotal in a number of contexts. However, security flaws in IoT devices are becoming worse as their number increases and their defence systems aren't up to par. Within the context of smart home IoT networks, this study proposes a bi-layer intrusion detection system (IDS) that relies on behaviour profiling of devices. An innovative rule set module and a supervised learning model make the system work, allowing the IDS to do detection tasks more quickly and correctly.

N. Panneerselvam et al., [2] Internet of Things (IoT) platforms have grown into an international powerhouse in the last decade, permeating any aspect of human life with its limitless smart services and so enriching our existence. This change happened all around the world. The ever-increasing need for smart devices and networks, coupled with the Internet of Things' (IoT) relative accessibility, has led to security problems of a magnitude never seen before. The internet of things may be protected using existing security measures.

R. Bolleddula et al., [3] As a result of rapid development and spread of intelligent schemes and autonomous, energy-aware sensing plans, the IIoT has skyrocketed and obstructed almost all demands. Despite the limitations of IIoT devices in terms of connection, storage, and computation, botnet attacks based on IIoT have been on the rise. In order to combat this danger, a system is needed that can map out harmful occurrences across IoT networks. We provide an ensemble learning model for machine learning that examines the features of IoT networks in order to identify abnormal traffic generated by infected IoT nodes; this model can then be used to detect botnet assaults in IIoT networks. In order to establish our Internet of Things (IoT) botnet detection approach, we additionally evaluate four separate machine learning techniques: XGBoost, Multi-Layer Perceptron (MLP), Generalised Additive Models (GAMs), and Random Forests (RF).

M. A. Muhammad et al., [4] There are currently no network access control systems that can detect or identify anomalous behaviours depending on the kind of device, and these systems may include unanticipated interactions between different device models, different network protocol levels (e.g., TCP, UDP, and ICMP), hardware, and effects related to clock skew. An additional complication is that intrusion detection and prevention systems are becoming more ineffective in detecting unlawful and unauthorised access to devices inside company networks due to the proliferation of security threats, vulnerabilities, and dangers brought about by the "bring your own device" policy. The results can be catastrophic. The authors of this study thus provide a simple clustering method that can distinguish between typical and atypical network traffic patterns in order to identify these problems.

R Mills et al., [5] The health, energy, manufacturing, and transportation industries are just a few that stand to benefit from the convergence of the Internet of Things (IoT) with developments in communications, Big Data, and networked systems. Because of the way manufacturers and ICT operators are now doing business, there is very little security when it comes to deploying IoT devices across different networked infrastructures. This leaves a lot of room for new types of attacks. Because they are based on previously established attack signatures, the traditional rule-based intrusion detection methods used by network management systems cannot detect novel assaults. On the other hand, when it comes to profiling typical network activity, anomaly detection systems sometimes have significant false positive rates since ground truth data isn't statistically validated enough.

S. Ho et al.,[6] The internet has become the central hub for the majority of people's day-to-day activities as a result of the vast quantity of data and services that have been transferred online to users in recent years, as well as the massive amount of digital privacy information that has been shared. The growing use of the internet, on the other hand, entails an increase in the number of potential targets for cyberattacks. If an efficient protection system is not put into place, the internet will become significantly more susceptible, which will in turn increase the likelihood that data will be stolen or compromised. The model has been assessed for its overall accuracy, as well as its rate of attack detection, its rate of false alarms, and its training overhead.

V. K. Navya et al.,[7] Because of the exponential rate at which technology is advancing, the threat of invasion is one that must be considered on a regular basis. The purpose of this research is to identify such breaches by utilizing certain algorithms that fall under the umbrella of machine learning. The development of IDS that can identify and categorize cyberattacks in a timely and automated manner at both the network level and the host level is making extensive use of machine learning methods. This can be fairly difficult to do effectively due to the fact that there are many distinct kinds of invasions happening on a wide scale all at once.

Y. A. Farrukh et al.,[8] Because of their reliance on information and communication technology, modern smart grid systems are vulnerable to cyberattacks because of their dependence on these technologies. In recent years, there has been a rise in the number of cyberattacks, which has led to significant damage being caused to power infrastructure. Techniques for cyber security, control, and detection are increasingly becoming necessary in order to provide a dependable and stable functioning. It is difficult to automate the detection of cyberattacks with a high degree of precision. In order to solve this problem, we have developed a two-layer hierarchical machine learning model that has an accuracy of 95.44% and can significantly increase the detection of cyberattacks. The initial layer of the model is responsible for differentiating between the two modes of operation, which can either be a regular state or a cyberattack. The state is then categorized into the various kinds of cyberattacks using the second layer of analysis.

S. Thirimanne et al., [9] Over the course of the past several years, numerous novel kinds of incursions that are distinct from those already known have been discovered. In addition, because cyberattacks are always evolving, the datasets that machine learning algorithms use need to be regularly updated so that they include the most current breaches. The primary objective of this study is to identify the most effective machine learning algorithm for intrusion detection that can be trained on the NSL-KDD and the UNSW-NB15 datasets. Additionally, this study will conduct a comparative analysis of six different machine learning algorithms that can be categorized as supervised, semi-supervised, or unsupervised learning. According to the findings of this research, the performance of supervised and semi-supervised machine learning algorithms outperformed the performance of unsupervised machine learning algorithms for both datasets.

T. T. Nguyen et al.,[10] The number of systems that are linked to the internet has grown significantly, and as a result, they are more vulnerable than ever before to being attacked by malicious software. Because of the complexity and fluidity of cyberattacks, protective measures need to be able to respond quickly, adapt to changing circumstances, and scale up as needed. Methods based on machine learning, and more especially DRL, have received a lot of attention as potential solutions to these problems. DRL is extremely capable of tackling complicated, dynamic, and especially high-dimensional cyber protection challenges since it incorporates deep learning with classical RL. An overview of DRL techniques that have been developed for cyber security is provided in this article. We discuss a variety of essential features, some of which are as follows: DRL-based security approaches for cyber-physical systems; autonomous intrusion detection techniques; and multiagent DRL-based game theory simulations for defensive tactics against cyberattacks.

R. Tiwari et al.,[11] It has been suggested that cognitive radio, or CR for short, be used as a method to increase the efficiency of spectrum utilization. This is accomplished by providing unlicensed users with opportunistic access to vacant or underutilized spectrum. Coordinated multipoint joint transmission (JT), also known as joint transmission between several sites, is an additional potential technique for improving the performance of cognitive radio networks. We present a coordinated multipoint JT technique as a part of a CR system in this study. An analytical model for the received signal-to-noise ratio at a CR is developed in order to determine the energy detection threshold and the smallest number of samples required for energy detection-based spectrum sensing in a CR network (CRN) with the CoMP JT technique. This is done in order to determine the energy detection threshold.

A. Sonkar et al.,[12] Blockchain innovation has as of late drawn in a great deal of interest as a potential major advantage for various applications. By the by, there are as yet significant

snags with blockchain organizations' presentation and versatility. AI approaches give possible solutions to improving and advancing blockchain network activities around here. There is an additional opportunity for supporting the proficiency of blockchain networks: choice tree calculations. These calculations are prestigious for being basic, interpretable, and adaptable. This study dives into choice tree-based AI procedures that are planned in light of blockchain networks.

N. K. Gupta et al.,[13] The reason for this examination is to give a clever strategy to foreseeing sluggishness by utilizing electroencephalogram (EEG) signal examination related to a mixture AI model. There are a few unique fields where sluggishness identification is vital, including medical services and transportation wellbeing. There is many times an absence of accuracy and strength in customary strategies. With the end goal of this examination, we utilize the broad transient data that is accumulated by EEG signals to plan a drowsiness forecast framework that is more reliable. The methodology that we use integrates approaches for highlight extraction with a cross-breed AI model that joins profound learning and conventional AI calculations.

D. Park et al.,[14] Traditional intrusion detection systems are finding it increasingly challenging to identify sophisticated cyberattacks since these attacks diverge from the patterns, they have previously stored because cyberattacks are becoming cleverer. In order to address this issue, a model for an intrusion detection system that is based on deep learning has been developed. This model examines sophisticated attack patterns by learning from data. However, deep learning models have the drawback of needing to relearn every time a new cyberattack technique is discovered. This can be a time-consuming process.

III. CHALLENGES

Intrusion detection in smart home IoT environments using machine learning (ML) and deep learning (DL) techniques offers promising solutions but is not without its challenges. Several key obstacles must be addressed to ensure the effective deployment of these systems in real-world scenarios:

1. Data Collection and Labeling

One of the most significant challenges in applying ML and DL techniques to intrusion detection is the collection of high-quality and representative data. Smart home IoT devices generate massive amounts of heterogeneous data from sensors, cameras, and other connected devices, but obtaining well-labeled datasets that cover all potential attack types is difficult. The lack of publicly available IoT datasets makes it challenging to train models that can generalize to diverse environments. Additionally, labeling such data requires extensive domain expertise and resources, as labeling must be accurate to avoid model bias.

2. Resource Constraints on IoT Devices

Many IoT devices in smart homes have limited computational resources, such as low processing power, memory, and storage capacity. Running complex ML or DL models locally on these devices is impractical, especially for deep learning models, which require significant computational power. The resource constraints also hinder real-time detection and model updates. Solutions such as edge computing or cloud-based IDS have been proposed to offload processing tasks, but these approaches introduce latency and privacy concerns.

3. Real-Time Detection and Latency

Intrusion detection systems for smart homes must operate in real-time to quickly detect and respond to potential threats. However, ML and DL models, particularly deep learning models, can be computationally intensive and may introduce latency during inference. Achieving a balance between detection accuracy and computational efficiency is critical. Systems must be optimized to ensure that they process data rapidly without compromising detection rates, as delays in identifying an attack could result in significant damage to the smart home infrastructure.

4. False Positives and False Negatives

The issue of false positives (benign activity incorrectly flagged as malicious) and false negatives (malicious activity going undetected) remains a major concern in IDS. False positives can overwhelm users with unnecessary alerts, reducing trust in the system, while false negatives expose the network to undetected attacks. ML and DL models must be carefully fine-tuned to minimize both types of errors, but achieving this balance is difficult due to the dynamic nature of IoT networks and traffic patterns.

5. Adversarial Attacks

Adversarial attacks pose a significant challenge to ML and DL-based intrusion detection systems. In such attacks, malicious actors can manipulate input data to deceive the model, tricking it into misclassifying malicious traffic as normal. This is particularly concerning in smart home environments, where subtle changes to network traffic may be difficult to detect. Defending against adversarial attacks requires advanced strategies such as adversarial training, which can significantly increase the complexity of model development and deployment.

6. Heterogeneous Smart Home Environments

Smart homes consist of a wide variety of IoT devices, each with different communication protocols, standards, and functionalities. This heterogeneity complicates the development of a one-size-fits-all IDS solution. ML and DL models must be adaptable enough to handle diverse traffic patterns and device behaviors while maintaining high detection accuracy across different smart home setups. Developing models that generalize well to various environments without requiring extensive retraining remains a challenge.

7. Privacy Concerns

Intrusion detection systems must analyze vast amounts of network traffic and device data to detect malicious activities. However, this raises privacy concerns, as sensitive user information could be inadvertently exposed or misused. The use of cloud-based or edge computing for intrusion detection also involves transmitting data to external servers, which can further compromise privacy. Ensuring that IDS solutions respect user privacy while providing robust security is a delicate balance that requires careful consideration of data protection protocols.

8. Evolving Attack Vectors

The threat landscape for smart homes is constantly evolving, with new attack vectors and sophisticated intrusion techniques emerging regularly. ML and DL models trained on historical data may struggle to detect novel attacks or zero-day exploits. Continuous retraining of models is necessary to keep up with the changing nature of cyber threats, but this process can be resource-intensive and time-consuming. Developing models that are adaptable and capable of self-learning from new data without requiring frequent manual intervention remains a key area of research.

IV. CONCLUSION

While machine learning and deep learning techniques offer significant potential for enhancing intrusion detection in smart home IoT environments, numerous challenges must be addressed to fully realize their benefits. These include the need for high-quality data, overcoming resource constraints, managing false positives and negatives, and defending against adversarial attacks. Additionally, the heterogeneous nature of IoT devices and evolving cyber threats require adaptable and scalable solutions. By addressing these challenges through innovative approaches such as privacy-preserving mechanisms and advanced learning techniques, the future of intelligent intrusion detection systems for smart homes holds great promise for securing connected living environments.

REFERENCES

1. F. He, F. Tong and Y. Zhang, "A Bi-Layer Intrusion Detection Based on Device Behavior Profiling for Smart Home IoT," *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, Denver, CO, USA, 2022, pp. 373-379, doi: 10.1109/MASS56207.2022.00060.
2. N. Panneerselvam and S. Krithiga, "Enhancing the Security of Iot Service Using Sem Model Based Machine Learning Technique," *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, Chennai, India, 2023, pp. 1-12, doi: 10.1109/ICCEBS58601.2023.10449310.
3. R. Bolleddula, M. Balakrishna, C. S. Vyshnavi, A. Harshitha, B. S. Nithisha and B. Vineetha, "Routing attack Detection using Ensemble Artificial Intelligence Model for IIoT," *2023 2nd International Conference on Futuristic Technologies (INCOFT)*, Belagavi, Karnataka, India, 2023, pp. 1-6, doi: 10.1109/INCOFT60753.2023.10425527.
4. M. A. Muhammad, F. Caraffini, A. Fasanmade, O. Ishola, K. Mohammed and J. Morden, "Data-Driven Design for Anomaly Detection in Network Access Control Systems," *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, Dubai, United Arab Emirates, 2023, pp. 1-10, doi: 10.1109/ICBATS57792.2023.10111130.
5. R. Mills, A. K. Marnerides, M. Broadbent and N. Race, "Practical Intrusion Detection of Emerging Threats,"

in *IEEE Transactions on Network and Service Management*, vol. 19, no. 1, pp. 582-600, March 2022, doi: 10.1109/TNSM.2021.3091517.

6. S. Ho, S. A. Jufout, K. Dajani and M. Mozumdar, "A Novel Intrusion Detection Model for Detecting Known and Innovative Cyberattacks Using Convolutional Neural Network," in IEEE Open Journal of the Computer Society, vol. 2, pp. 14-25, 2021, doi: 10.1109/OJCS.2021.3050917.
7. V. K. Navya, J. Adithi, D. Rudrawal, H. Tailor and N. James, "Intrusion Detection System using Deep Neural Networks (DNN)," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-6, doi: 10.1109/ICAECA52838.2021.9675513.
8. Y. A. Farrukh, Z. Ahmad, I. Khan and R. M. Elavarasan, "A Sequential Supervised Machine Learning Approach for Cyber Attack Detection in a Smart Grid System," 2021 North American Power Symposium (NAPS), 2021, pp. 1-6, doi: 10.1109/NAPS52732.2021.9654767.
9. S. Thirimanne, L. Jayawardana, P. Liyanaarachchi and L. Yasakethu, "Comparative Algorithm Analysis for Machine Learning Based Intrusion Detection System," 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), 2021, pp. 191-196, doi: 10.1109/ICIAfS52090.2021.9605814.
10. T. T. Nguyen and V. J. Reddi, "Deep Reinforcement Learning for Cyber Security," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2021.3121870.
11. R. Tiwari and K. Mishra, "Performance Analysis of Spectrum Sensing over Cognitive Radio Network with Joint Transmission," 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), Bhopal, India, 2022, pp. 1-6, doi: 10.1109/CCET56606.2022.10080214.
12. A. Sonkar, S. K. Sahu, A. Nayak, D. Sahu, P. Verma and R. Tiwari, "An Efficient Privacy-Preserving Machine Learning for Blockchain Network," *2024 4th International Conference on Intelligent Technologies (CONIT)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/CONIT61985.2024.10627061.
13. N. K. Gupta, P. Verma, R. Verma, T. Gajpal, J. Patel and R. Tiwari, "Enhanced Drowsiness Prediction through EEG Signal Analysis using Hybrid Machine Learning Model," *2024 4th International Conference on Intelligent Technologies (CONIT)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/CONIT61985.2024.10627591.
14. D. Park, S. Kim, H. Kwon, D. Shin and D. Shin, "Host-Based Intrusion Detection Model Using Siamese Network," in IEEE Access, vol. 9, pp. 76614-76623, 2021, doi: 10.1109/ACCESS.2021.3082160.