Advancements in IoT Anomaly Detection: Leveraging Machine Learning for Enhanced Security

Rajesh Rajaan[1], Baldev Singh[2], Nilam Choudhary[3]

[1,2]Department of Computer Science and Engineering, Vivekananda Global University, Jaipur, India
[3]Swami Keshvanand Institute of Technology, Management &Gramothan, Jaipur, India
Corresponding Author Email: raaj0028@gmail.com

Abstract. The internet of things(IoT) is one of the fastest-growing technologies that has disrupted industries by allowing devices to connect without disruption. Nevertheless, the adoption of IoT devices has posed new security risks as it is hard to distinguish between normal and anomalous behaviors indicative of cyber-attacks or system issues. This paper aims at discussing the present progress of IoT anomaly detection; this has the argument that most of the solution relies on machine learning (ML). ML has been shown to be very influential in this field thanks to its ability to unravel sophisticated patterns from large, diverse datasets, something which many conventional approaches are incapable of. This paper investigates various approaches used in anomaly detection in IoT using ensemble learning, deep learning, and a combination of both. An insight into their strengths and weaknesses is given. Important directions for further research are the use of ensemble learning methods that combine several classifiers for better detection ability and deep learning models such as CNN and LSTM for temporal and spatial data analysis. A discussion also ensues to show how to overcome class imbalance in IoT datasets using SMOTE method. Also, it highlights that there is a demand for immediate dynamic anomaly detection for IoT systems since the nature of IoT is unsteady and emphasizes the utilization of the explainable AI (XAI) to enhance the users' trust in models. We also find gaps in the literature: first, insufficient exploration of the use of ensemble methods; second, high demand for comparing different approaches; third, the necessity to develop low-power models for IoT devices. This is an area that should receive further attention when it comes to prospective work because creating better, more flexible, and finer weight structures to protect IoT network appearances against new threats should be the goal of future work. Based on the findings in this paper, there is need for continual research in the IoT anomaly detection where such concerns as performance, flexibility and interpretability need to be met to establish secure IoT systems.

Keywords: IoT anomaly detection, machine learning, cybersecurity, ensemble learning, deep learning, class imbalance, SMOTE, real-time detection, explainable AI, adaptive models, IoT security, data preprocessing, feature extraction, lightweight models, hybrid detection methods.

© The Author(s) 2025
N. Vyas et al. (eds.), *Proceedings of the International Conference on Advancements in Computing Technologies and Artificial Intelligence (COMPUTATIA 2025)*, Advances in Intelligent Systems Research 189,
https://doi.org/10.2991/978-94-6463-700-7_30

1. Introduction

Information and technology known as the Internet of Things (IoT) is one of the most innovative networks invented that connects billions of devices among industries and fields such as healthcare, manufacturing, houses, and transportation. The adoption of IoT can provide opportunities to improve data collection, analysis, and decision making in real time that improves operations and creates opportunities for cross industry innovations. But the large and diverse structure of IoT networks has become vulnerable to cyber threats because of their complexities. This has resulted in a growing demand for critical and effective anomaly detection methods for IoT systems safety against potential risks.

The primary concern of the research found in the context of IoT anomaly detection concerns the design of approaches that improve detection accuracy, exhibit resilience to data imbalance, and are applicable to various IoT contexts. In contrast to conventional networks, IoT systems feature a diverse nature and complex pattern of data and processing capabilities of connected devices and the frequency of their interactions. Another problem is that 'the IoT devices are constantly streaming data,' and since both the nature and scale of IoT operations are unique, 'the continuous data flow complicates effective anomaly detection even more.' This scenario clearly shows the need for rich techniques to detect anomalies that may call for disruption of IoT networks in order to protect their reliability and security.

Several works have adopted machine learning as one of the most suitable approaches for developing detection systems for IoT settings. The key feature that makes ML techniques suitable for IoT anomaly detection is their capability to process massive amounts of original data to spot emerging patterns that may characterize anomalies. While detection systems such as rule-based systems utilize attacks' signatures, the ML algorithms are more productive in learning from data making them perfect to current threats. The Figure 1 gives anidea of a essence of internet of things in the current day applications. One major field of study was the development of ensemble machine learning methodologies, which integrate several weak classifiers that for a strong detection system. Some of the techniques like bagging, boosting, and stacking hold have been reported to make the accuracy and reliability of the anomaly detection systems high. Since ensemble process combines the outputs of individual classifiers, it is less prone to variance and bias hence out-performs the individual classifier. This is important especially within the IoT anomaly detection problem since the data is often skewed and the type of anomaly differs considerably.

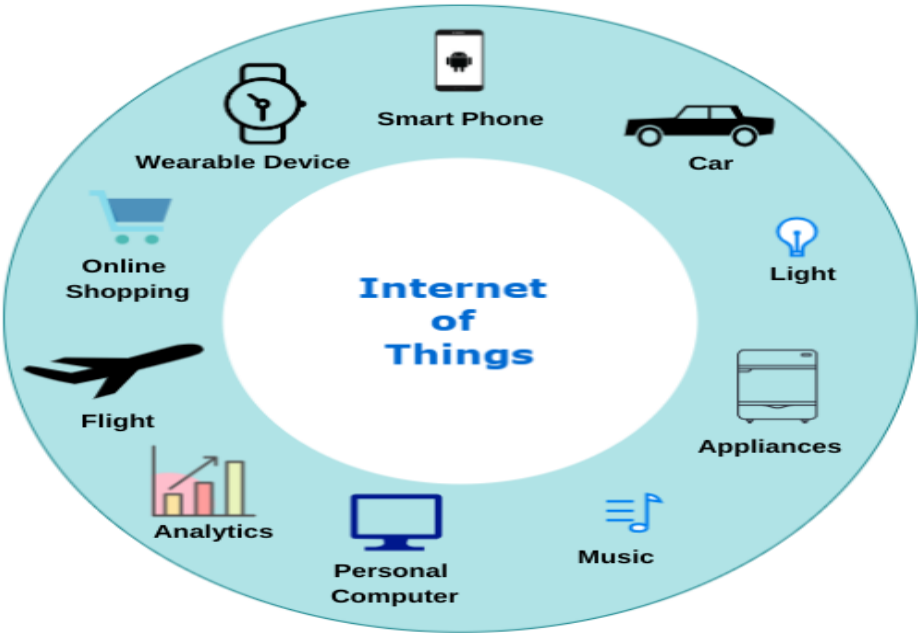

Figure 1. Diffirent Aspects of IOT in Modern Day Applications

Another crucial issue in the IoT anomaly detection is class imbalance. In datasets of IoT, it is common that normal data is much more than anomalous data making training biased and affect the performance of machine learning algorithms. Alleviating imbalance has been the subject of investigation including Synthetic Minority Oversampling Technique (SMOTE) and other derivatives. The raw data distribution is improved by using SMOTE to synthesize samples of the minority class and use them to train the model. Although these techniques have been used in other research domains, their application in the context of IoT seems to be more challenging research direction since some of the peculiarities of the IoT data sources which include real-time data analysis and the restricted computational capabilities of IoT devices need more study. While the primary consideration of those specialties is improving the accuracy of IoT systems, while paying attention to data imbalance considerations, and promoting versatility within disparate IoT arrangements, another factor about which the evolution of better approaches to anomaly detection is improving is the usage of increasingly refined techniques. Using basic concepts of machine learning and its intrinsic capability of learning from data it has emerged as a vital tool in this pursuit. However, there is some limitation such as; it faces some challenge like the data heterogeneity, there is also an issue of class imbalance and the researcher is also faced with limitation of computation resources.

Due to these challenges, the further development of ensemble learning techniques, hybrid models and adaptive learning approaches looks very promising. While the literature is still expanding its horizons and proposing more sophisticated approaches for IoT anomaly detection, the overarching objective is to design intelligent IoT anomaly detection systems that are not only robust, with high accuracy and low computational overhead, but also able to learn and evolve as more theoretical and real-world applications are developed in the IoT domain. This constant advancement is necessary for the principle of developing IoT technologies and, therefore, making innovations safe and useful in various industries and for the improvement of people's lives worldwide.

The following are the general contributions of this research towards the future development of IoT anomaly detection: The future of IoT anomaly detection research has to be characterized by sustained improvement and evolution to reflect practice relevance and applicability and the capability to handle large datasets. Using the findings of the existing work and analysing new opportunities, scholars and practitioners can create a stable and safe IoT environment which can stand against numerous threats that appear with the growth of connectivity and data sharing.

2. Related Works

The research on this broad area of IoT anomaly detection aims to enhance accuracy, manage inconsistencies, and provide assurance of flexibility in different IoT settings [1]. Huge and diverse data streams have been generated due to the IoT devices, requiring novel means for detecting and mitigating such threats [2]. Another interesting area is ensemble machine learning methods that use several weak classifiers to form a strong detection system to enhance the accuracy and reliability of the anomaly detection [3]. This method enhance the performance and reduce the false positives since it incorporated several detection skills. Class imbalance in IoT datasets is another problem where more classical approaches often fail Class imbalance. Such concepts like the Synthetic Minority Oversampling Technique (SMOTE) where new samples for the minority class are generated artificially and the model is trained on the unbalanced data set [4,5].

Comparative studies are equally important in IoT anomaly detection because when there is a new model, one can be able to tell how much effective it's compared to the existing ones, as well as pointing out some potentials for improvement as suggested in [6,7]. These works assess the effectiveness of the model as a whole and also its performance at detecting a range of threats – frequent risks and rare but highly potentially hazardous ones [8,9]. As IoT networks are dynamic and fast evolving, the models used must be accurate while at the same time being adaptive to capture new evolving anomalous patterns [13,14]. In light of this, the use of adaptive models, which on many occasions involves online learning algorithms since they update their detection accuracy as data patterns and network behaviors evolve.

In many applications of IoT, including smart homes and industrial automation,

practical usage of anomaly detection in different models defines scalability and cybersecurity consequences [12,13]. These practical applications examine the capability of the models in covering a variety of situations, which provides significant useful data about their advantageous and potential disadvantageous circumstances [14,15]. Besides contributing to expansion of knowledge repository in academia, anomaly detection has relevant applications which will further spur advancement of sophisticated and secure IoT network [16,17].

This paper also identifies several problems with IoT anomaly detection that are not related to the amount of data. It is difficult to apply the time-consuming algorithms as the IoT data differ in types, and the operational IoT devices are constrained by the physical and computational capabilities [18,19]. Thus, the importance is stressed upon the development of models that are as scalable as possible, efficient, and as lean as possible. Also, because of the constantly evolving nature of the IoT environment, which continually releases new devices into the ecosystem with new communication interfaces, as well as developing new threat types, the techniques for anomaly detection require constant enhancement and improvement [20,21].

Any machine learning models that are aimed at detecting anomalies must have a proper understanding of how device communicates and behaves in a network. This drives the importance of offering solutions that can solve complex relations experienced in IoT ecosystems [22,23]. The requirement of keeping the security and safety of IoT networks increases as they expand. This has encouraged further work to build more sophisticated anomaly detection systems [24,25]. High level anomaly detection is required to safeguard such settings since IoT technology is growing and comes with opportunities and risk" [26,27].

Here, research extends beyond enhancing the sophistication of the tools and techniques used for anomaly detection to contribute to creating a more secure environment online where IoT devices operate optimally and securely [28,29]. To further explore IoT security to its maximum, researchers and prac-titioners are needed because the outcomes of researches are translate into practical usable technologies in concrete challenging cases [30,31]. While IoT anomaly detection research continues, the goal remains to create systems that can both find an appropriate number of anomalies and adapt to the changing IoT landscape to protect the security and stability of networks [32,33].

Therefore, studies in the area of IoT anomaly detection are evolving in response to the development and diversification of IoT networks [34,35]. The concern is on the kind of detection procedures that can effectively detect anomalies while at the same time have the versatility to adapt to the IoT systems' environment [36,37]. The work of such researchers is underneath different areas, which in turn enhances the organizational efficiency and security of the IoT systems preparing for future advancements and safe implementation of IoT in our everyday lifestyles [38,39]. These prove the richness of the field and the importance of the knowledge in shaping the further development of cybersecurity in the context of the IoT [40,41] such as in the never ceasing search for novelties and improvements in the area of anomaly detection.

This field's research is aimed at creating an IoT ecosystem where the common attributes are intelligence, connectivity, safety and resistance to the many risks that accompany the increase in connections and the sharing of data [42,43]. Thanks to this knowledge derived from such studies, all the users can embrace IoT technologies in the safest and most secure way possible since the research shapes enhanced safety solutions for deployment in real-life conditions [44,45]. The current innovation in IoT anomaly detection makes way for future progression of IoT technologies to enhance people's lives with minimal security and trust – something that's paramount for its sustainability [46,47].

This means that new approaches for anomaly detection in IoT must be created, the existing ones need to be enhanced, and their performance in different IoT contexts needs to be tested periodically [48,49]. Besides, fulfilling the aim to enhance the understanding of IoT security risks this journey also influences society's using and benefiting from IoT, enabling a proper usage of its opportunities while almost eradicating the potential dangers [50,51]. As long as this research works, it is set become an integral aspect of efforts to fully unlock the potential of IoT technology for the benefit of achieving a future with highly secured IOT devices that will help promote innovation.

The relevance of this topic is not just theoretical, it influences business processes, legislation and the everyday functioning of enterprises and individuals who rely on IoT [54,55]. The fact that IoT anomaly detection research is still ongoing is a demonstration of how, in addition to creating new knowledge, research may also be laying the groundwork for advances that arise in the reliability and protection of incrementally more essential technologies [56,57]. Thus, in order to protect IoT systems from cyber attacks and leverage the benefits of interconnection, the findings for IoT anomaly detection will be critical for future improvement and innovation of IoT systems [58,59]. In conclusion, anomaly detection for IoT is a very important aspect in combating cybersecurity challenges and ensuring that IoT solutions are safe and efficient to use across the globe [60].

3. AnalysisandFindings

Significant progress in IoT anomaly detection has been made, according to the examined literature, thanks to creative machine learning applications and the creation of sophisticated detection techniques. Important insights that show the potential of machine learning, dynamic behavior-based detection, and protocol-specific techniques in recognizing and neutralizing botnet threats have been provided by researchers like [35,36], and [44]. Notwithstanding these successes, there are still a number of research gaps that point to directions for further study to increase the resilience and versatility of IoT anomaly detection systems. Table 1 provides an overview of the literature review findings, emphasizing key observations and areas for further investigation to enhance IoT anomaly detection methodologies.

Table 1. Analysis of Methodologies Related to Different Approaches

Aspect	Findings	Analysis
Ensemble Machine Learning	Limited exploration in IoT anomaly detection; potential to improve accuracy by combining weak classifiers [2][3].	Ensemble methods like Random Forest and XGBoost show promise for boosting model robustness and precision, but their application in IoT remains underexplored. This gap presents research opportunities.
Traditional ML Techniques	Algorithms like Decision Trees, SVM, and k-NN have demonstrated basic efficacy in detecting IoT anomalies [35][36].	While these models provide foundational insights, they often struggle with high-dimensional IoT data and do not generalize well without extensive tuning.
Deep Learning Models	CNNs and LSTMs have been effective in capturing complex spatial and temporal patterns in IoT data [44].	Deep learning enhances anomaly detection by learning intricate patterns, but models can be resource-intensive, making them less suitable for devices with limited processing power.
Class Imbalance Solutions	Approaches like SMOTE and ADASYN have been used to address skewed data distributions [4][5].	These techniques can improve the identification of rare anomalies by balancing datasets. However, their impact on IoT applications requires further empirical validation to understand performance trade-offs.
Comparative Studies	Few studies benchmark new models against traditional ones, limiting a clear understanding of their relative strengths [6][7].	Comprehensive comparative analyses are essential to measure the true advancement of newer models. A lack of such studies creates ambiguity in determining their actual benefit over existing methods.
Real-Time Adaptability	The dynamic nature of IoT data requires models that can learn and adapt on the fly [10][11].	Online learning algorithms and incremental learning in neural networks are promising approaches but need more refinement to balance accuracy with real-time processing capabilities.
Protocol-Based Detection	Protocol-specific anomaly detection (e.g., SVELTE for CoAP) shows potential in targeted security [51].	These methods enhance detection for particular IoT protocols, but their broader applicability across diverse IoT ecosystems is limited, necessitating more versatile solutions.

Behavioral Analysis	Behavioral models such as HADES-IoT use dynamic monitoring to detect botnet activity effectively [44].	Real-time behavioral analysis is powerful for identifying evolving threats but often requires integration with system-level data, which may increase overhead and complexity.
Hybrid Detection Methods	Combining signature-based and anomaly-based detection offers comprehensive coverage for known and unknown threats [52][53].	Hybrid systems can provide better anomaly coverage, but scalability and processing speed can become challenges in large IoT networks.
Explainable AI (XAI)	Limited use in IoT anomaly detection, but offers potential to make models more transparent and user-friendly [28][29].	XAI can build trust in anomaly detection systems by clarifying decision-making processes, essential for user adoption and practical implementation in sensitive IoT applications.
Energy Efficiency	Resource-intensive models pose challenges for IoT devices with limited computational power.	The development of lightweight and energy-efficient models is crucial for real-world deployment, especially on edge devices.
Emerging Trends	Research is moving towards federated learning and transfer learning for improved data privacy and model generalizability [30][31].	Federated learning enables collaborative training across distributed IoT devices without centralizing data, promising enhanced security and privacy. Transfer learning can aid in training models on related IoT tasks.

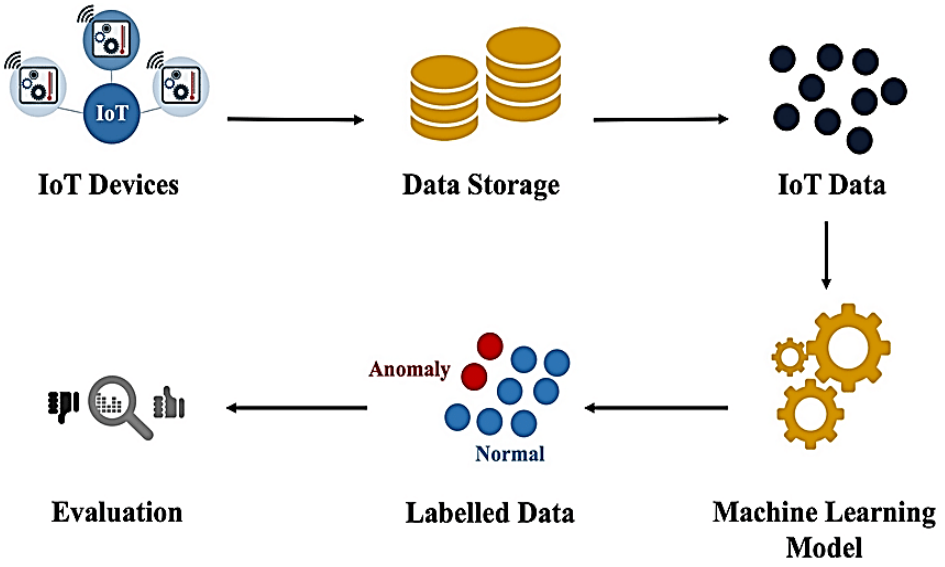

Figure 2. Basic Methodological Framework for IOT Anomaly Detection

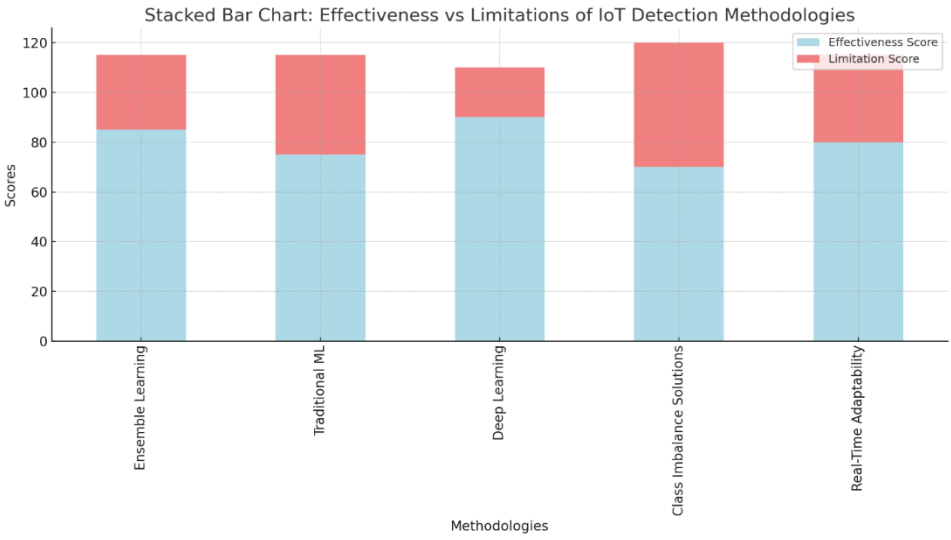

Figure 3. Analysis of Effectiveness vs Limitations

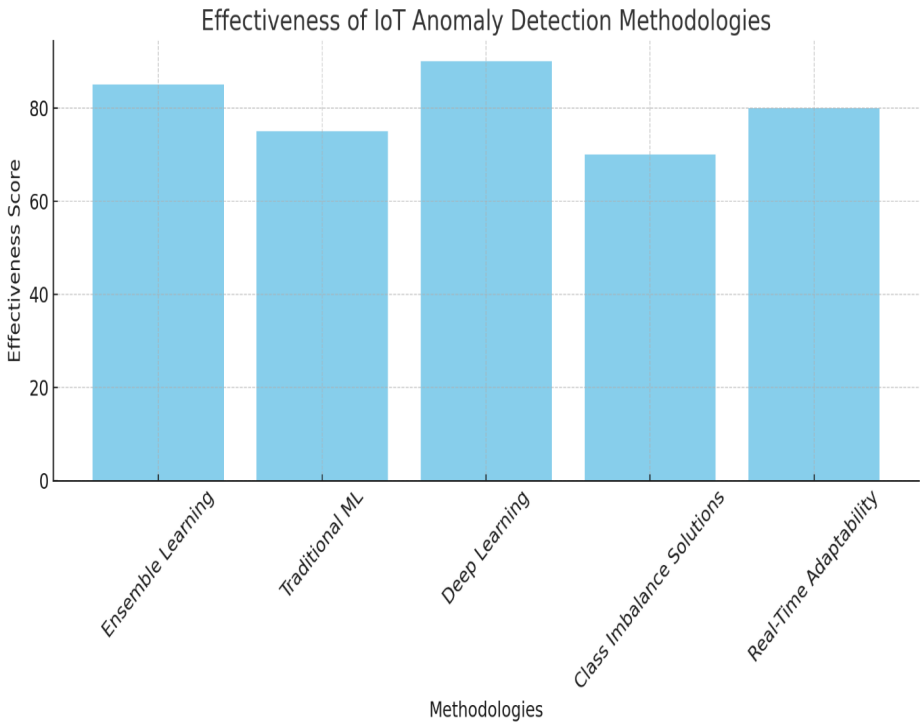

Figure 4. Analysis of Effectiveness Score Based on Review

Figure 2 present a basic methodological framework for IOT anomaly detection and an analysis between effectiveness and limitations shows in Figure 3, while Figure 4 provide an analysis of effectiveness score based on review. Firstly, although other studies have presented effective models of ML for the detection of the anomalies in IoT systems, there are limited number of investigations of the use of the ensemble ML methods for IoT anomaly detection proposed enhancements by SVM models combined with better feature sets as well as by by presenting uniform models of intrusion detection [35, 36]. However, there is little literature on how ensemble methods that involve the use of multiple weak classifiers to derive more comprehensive detection systems can be incorporated. These ensemble approaches could, therefore, enhance the detection accuracy and reliability in a way which fortifies the safeguard against complicated and evolving IoT threats.One of the highlighted challenges in the literature that needs an urgent address is the disparate

distribution of class in IoT datasets. It is evident from the reviewed research that majority of them classify machine learning models as being generally effective; however, there is little detail on how outcomes of the anomaly detection is impacted on by skewed class distributions. This imbalance may hold the key to explaining models that are not good at detecting extremely rare but very valuable outliers. One direction that is missing in the current research is underscored by a paucity of research that specifically examines this problem. The imbalance of the prepared model training and performance of anomaly detection could be enhanced using a range of methods; for instance, improved hybrid balancing techniques or the Synthetic Minority Over-sampling Technique (SMOTE) [4,5].An area that requires improvements is the lack of comprehensive comparative studies. The problem here is that while numerous articles often focus on presenting new IoT anomaly detection frameworks as highly effective, there is scant research that directly compares these novel approaches to more traditional, conventional strategies. In particular, comparative research is necessary to determine the effectiveness of new approaches and identify the segments that require additional work. In this way, the assessments of the present IOS anomalies would help outline future developments to achieve a higher understanding of the strengths and weaknesses of current models [6, 7].

Additionally, although data balancing methods such as SMOTE have proven effective in other domains, their use in IoT anomaly detection has not been thoroughly investigated. Such methods have a great deal of potential to address the issue of class imbalance in IoT security, but their efficacy in this particular setting has not yet been thoroughly confirmed by study. The effects of these techniques on the performance and training of anomaly detection models customized for IoT data require more in-depth research [4,5].

The field is moving in new directions, as seen by behavior-based techniques, like HADES-IoT, as noted by [44], and domain-specific detection methodologies, like DNS traffic analysis. But it's still unclear how effective these strategies will be in practical situations, especially when dealing with the resource limitations that characterize IoT devices. The wider application of such solutions across various IoT contexts requires more research, even while HADES-IoT and comparable systems make use of real-time threat detection capabilities through sophisticated monitoring.

The work of [53] using SVELTE is an example of a hybrid anomaly detection system that combines signature-based and anomaly-based techniques and has demonstrated promise in addressing a wider spectrum of threats. However, it is still unknown to what degree these techniques may be expanded and modified to fit more complex and varied IoT networks. Another aspect of the research difficulties that need to be addressed is the importance of lightweight and energy-efficient models for Internet of Things devices with constrained processing capabilities.

For this reason, in addition to focusing on increasing the detection accuracy as networks of IoTs evolve, the research also emphasizes the need for learning algorithms that can update themselves as soon as new data patterns appear. So we could speak about “future proof” IoT security systems only when models in the way are ready to learn on the occurrences of new kinds of abnormalities [10,11]. Lastly, while a great deal has been accomplished in the development of XAI methods for traditional machine learning models,

there is an absence of application for these techniques to IoT anomaly detection. It also makes use of real-time knowledge of the why when it comes specifically to the anomaly that a given model detected and can help to boost the confidence of the user and assist in a higher degree of decision-making. Increased understanding of how IoT security systems function might be achieved through usage of XAIenabled solutions for detecting anomalies.

The research gaps that have been ascertained in this study reveal the areas that are most important and deserve more emphasis: First, ensemble learning has been investigated sparingly; second, class imbalance has received insufficient attention; third, comparative studies are scarce; and, fourth, data balancing methodologies are poorly validated. Closing these gaps may lead to, more flexible, extensive, and more resistant IoT anomaly detection frameworks. The future work can more proactively confront with the complex and evolving IoT threats through enhancing the existing strategies and exploring other possibilities to enhance the IoT security and resilience.

4. Materials and Methodologies

1. The analysis process of IoT anomaly detection is divided into several phases, namely data acquisition, data cleaning, feature extraction and learning algorithm development, to establish a reliable system that could detect and mitigate the abnormal activities. Due to the fact that IoT networks encompass heterogeneous devices from basic sensors to data smart home systems, such framework calls for materials and methodologies that address diverse data types, meet computational restrictions, and are capable of real time data processing and Figure 5 present a detection framework for different types of data.

1. Data Collection and Datasets IoT anomaly detection system development is incomplete without proper data collection. There is a strong positive correlation between the quality, variety, and utility of data incorporated into the system and the efficiency of the detection model. IoT networks engender large and heterogenous datasets as a result of the vastly connected object and the complex interconnectivity between devices in the Internet. These datasets often include:

- **Network Traffic Data**: Captured using tools like Wireshark or tcpdump, providing detailed packet-level information about device communications.
- **Sensor Logs and Device Data**: Generated by IoT sensors and devices, offering context-specific information such as temperature, motion, or light levels.
- **Benchmark IoT Datasets**: Commonly used datasets include the Bot-IoT, UNSW-NB15, and N-BaIoT datasets. These datasets are curated to reflect real-world IoT scenarios and incorporate various anomaly types, such as Distributed Denial of Service (DDoS), data exfiltration, and spoofing.

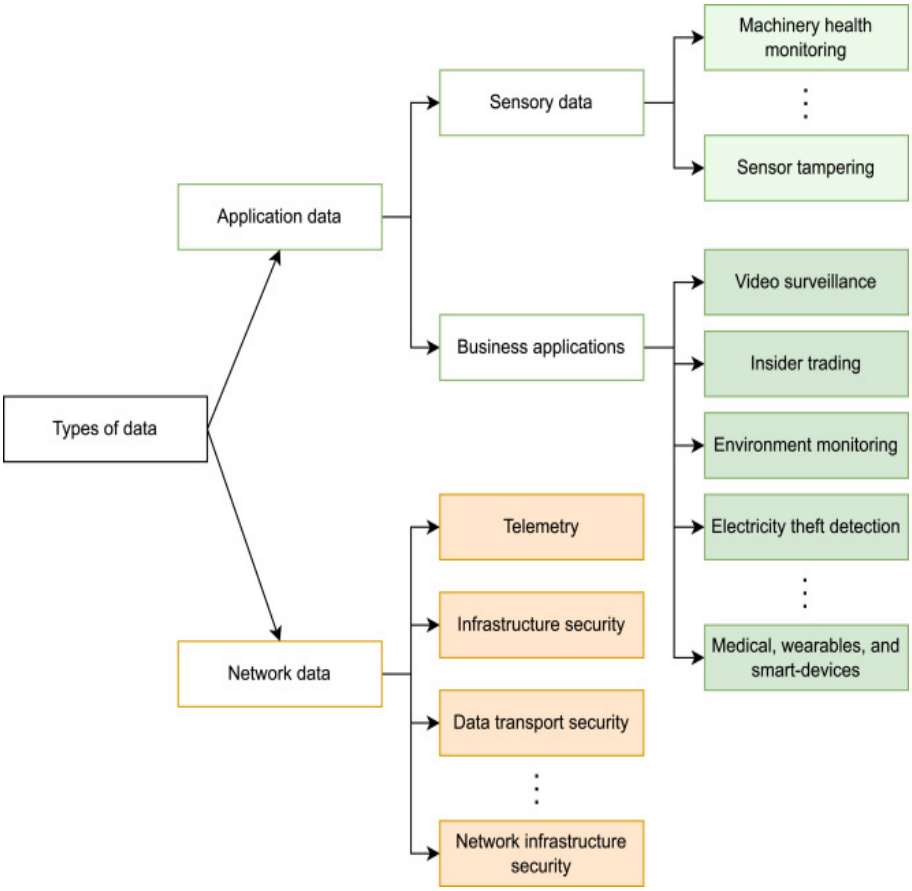

Figure 5. Different Types of Data for Detection Framework

2. Preprocessing Technique- Unfortunately, IoT data is usually vast and diverse, and there is always a need to pre-process the data before it can be acted upon. Preprocessing techniques include:

- **Data Cleaning:** Process of cutting down the data to a level that will feed the ML models by either deleting the useless data, wrong data or data that is similar to other data kinds.

- **Normalization and Standardization:** It checks the feature range of data and ensures that it is within an appropriate range so that many of the ML algorithms can converge.
- **Handling Missing Data:** There are several techniques in handling missing data entries, inclusively mean or median imputation method can be used.
- **Feature Engineering:** Packet size, source and destination IP address, protocol, and time stamp, additional features that can improve the learning capability of the Model learning algorithms.

3. Feature selection It is one of the important steps with the goal to reduce the size of the data set and enhances the models performance. Therefore, it can be seen that selection of features has direct influence on the precision, speed, and scalability of the process of anomaly detection in the model. Common techniques for feature selection include:

- **Filter Methods:** Features based on statistical relevant features selection techniques like mutual information and chi- square tests.
- **Wrapper Methods:** other techniques such as Recursive Feature Elimination (RFE) that work in cycles whereby it is only able to delete the least useful features.
- **Embedded Methods**: Part of model training, it was designed with models such as Lasso (L1 punishment that diminishes less significant features).

4. Anomalies in IoT environment can be detected based on the ability of ML algorithm to learn from data and identify characteristics of normal and abnormal behavior. It is most effective to divide the methodologies into classical machine learning algorithms, boosted techniques, and deep learning strategies.

4.1 Traditional Machine Learning Algorithms

- **Decision Trees**: Simple yet effective for initial experiments. They split the data based on feature thresholds, making them easy to interpret but prone to overfitting if not controlled.

- **Support Vector Machines (SVM)**: Known for their high performance in binary classification tasks. SVMs aim to find the hyperplane that best separates different classes. However, they can struggle with large, unbalanced datasets.

- **k-Nearest Neighbors (k-NN)**: A non-parametric method that classifies data points based on the majority class of their k-nearest neighbors. Although intuitive, k-NN can be computationally expensive for large datasets.

4.2 Ensemble Learning Techniques Ensemble methods combine multiple learning algorithms to improve the overall performance and robustness of anomaly detection systems.

- **Random Forest**: An ensemble of decision trees, each trained on a bootstrap sample of the data. The final prediction is based on the majority vote from all trees, reducing overfitting and enhancing generalizability.
- **Boosting Techniques (e.g., AdaBoost, XGBoost)**: Sequentially train models, where each subsequent model attempts to correct the errors of its predecessor. XGBoost, in particular, is known for its scalability and high accuracy in large datasets.
- **Stacking**: Combines the outputs of multiple base classifiers using a meta-classifier that learns from their predictions, potentially improving predictive power.

4.3 Deep Learning Models With the complexity and high-dimensionality of IoT data, deep learning models have shown significant promise.

- **Convolutional Neural Networks (CNNs)**: Traditionally used in image processing, CNNs can be adapted to work with structured data by treating feature matrices as input grids. Their ability to capture local dependencies and patterns makes them suitable for detecting specific anomaly patterns in IoT traffic data.
- **Recurrent Neural Networks (RNNs)**: Particularly Long Short-Term Memory (LSTM) networks are effective for sequential data analysis. IoT data streams often involve temporal dependencies, making LSTM ideal for learning from past behaviors to predict future anomalies.
- **Autoencoders**: Unsupervised deep learning models that learn to compress data into a latent space and reconstruct it. They are useful for anomaly detection because anomalies generally have a higher reconstruction error compared to normal data points.

5. Handling Class Imbalance A significant challenge in IoT anomaly detection is the imbalance between normal and anomalous data. Anomalies often constitute a very small portion of the dataset, which can bias models toward predicting the majority class. Solutions include:

- **Synthetic Minority Oversampling Technique (SMOTE)**: Generates synthetic data points for the minority class by interpolating between existing data points, effectively balancing the dataset.
- **Adaptive Synthetic Sampling (ADASYN)**: An extension of SMOTE that focuses on generating samples in areas where learning is more difficult.

- **Cost-Sensitive Learning**: Modifies the loss function of a model to penalize misclassifications of the minority class more heavily, thus encouraging the model to pay more attention to anomalies.

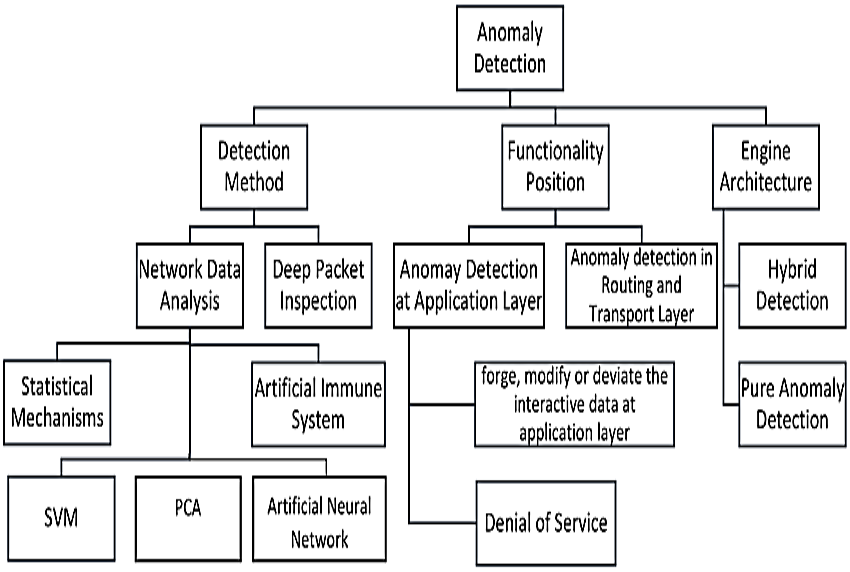

Figure 6. Compilation of Approaches

6. Evaluation Metrics for Anomaly Detection Performance evaluation is essential to validate the efficacy of IoT anomaly detection models. Common metrics include:

- **Accuracy**: The proportion of correctly classified instances, although not reliable for imbalanced datasets.
- **Precision, Recall, and F1 Score**: Precision measures the proportion of true positive detections out of all positive predictions, while recall (or sensitivity) measures the ability to identify true anomalies. The F1 score is the harmonic mean of precision and recall.
- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**: Represents the model's ability to discriminate between positive and negative classes.
- **Confusion Matrix Analysis**: Provides detailed insight into true positives, true negatives, false positives, and false negatives.

7. Real-Time Adaptation and Model Deployment The dynamic nature of IoT networks requires models that can adapt and update based on new data. Approaches to achieve this include:

- **Online Learning Algorithms**: Enable models to learn incrementally from new data without the need for complete retraining. Examples include online versions of gradient boosting and deep learning frameworks.
- **Incremental Learning in Neural Networks**: Techniques such as continual learning allow deep learning models to integrate new information without forgetting previously learned data.
- **Edge Computing Integration**: Deploying lightweight models on edge devices ensures faster processing and reduced latency, which is crucial for real-time anomaly detection.

8. Protocol-Based and Behavioral Analysis Techniques

- **Protocol Analysis**: Involves examining the specific protocols used in IoT communication (e.g., MQTT, CoAP) to identify deviations from expected behavior. Protocol-specific IDS (Intrusion Detection Systems) like SVELTE and lightweight models for Constrained Application Protocol (CoAP) have shown promise in efficiently detecting attacks.
- **Behavioral Analysis**: Monitors device behavior patterns and flags activities that deviate from established baselines. Advanced methods like HMMs (Hidden Markov Models) and graph-based models analyze sequences of actions, offering insights into anomalies stemming from botnet activities or DDoS attacks.

9. Hybrid Models and Future Directions Hybrid anomaly detection models combine multiple detection techniques to leverage their respective strengths. For example:

- **Signature-Based and Anomaly-Based Hybrid Systems**: Use known attack signatures for quick detection while simultaneously employing anomaly-based approaches to identify new, unknown threats.
- **Multi-Stage Detection Pipelines**: Employ lightweight models at the edge for initial filtering and deeper analysis in the cloud or centralized systems for complex pattern recognition.

10. Future methodologies may involve:

- **Federated Learning**: Allows collaborative model training across multiple devices while preserving data privacy, which is essential in IoT environments with sensitive data.

- **Explainable AI (XAI)**: Focuses on creating transparent models where decision-making processes are clear, enhancing trust and usability in real-world applications.
- **Transfer Learning**: Facilitates knowledge transfer from related domains, enabling models trained on one type of IoT data to be adapted for another with minimal retraining.

A compilation of approaches for IoT anomaly detection provided in Figure 6. The material and methodologies used applied in IoT anomaly detection are very vast and are being developed to cater for the emerging complex, large, and dynamic IoT networks. This paper argues that IoT anomaly detection should address data preprocessing, feature selection, and readiness of the latest machine learning algorithms and models, together with deep learning, focused approaches to handling class imbalance and evaluation metrics. Therefore, there is a need for future research with developing adaptive, scalable and interpretable approaches according to the actual problems and limitations of the IoT systems. Such a continuous refinement will help in creating more secure IoT networks, which would be more efficient at mitigating an increasing number of cyber threats.

5. Conclusion

IoT anomaly detection has further expanded greatly due to the improvements in machine learning and data analysis techniques. Finally, this review also emphasise the importance of using different materials and techniques in order to construct powerful detections that can respond to the specificities of IoT space. IoT networks, due to their largescale and heterogeneity, need solutions other than the conventional detection models. Hence, ensemble learning, as well as deep learning, has become an essential approach to ensure the creation of comprehensive solutions where and when an IoT system can detect different kinds of anomalies. Even the conventional techniques like Decision Trees, SVMs, and k-NN have created the basics of the initial anomaly detection frameworks; unfortunately, they bear limitations relating to the intricate structures of high-dimensionality data that solicit more civilized techniques. Area of robustness is produced by using a combination of classifiers known as ensemble methods, examples being Random Forests and Boosting algorithms for instance XGBoost. Artificial neural networks such as CNN and LSTMs are suited to learning more complex patterns and temporal dependence for IoT data streams because behavior patterns may change over time. However, several problems still remain unresolved; for instance, class imbalance issue that hampers the proper functioning of the model and renders the method incapable of detecting peculiar but potentially highly dangerous variants of anomalies. These approaches like SMOTE or ADASYN have given solutions to this if they have reduced the number of samples but approaches like SMOTE have shown that the issue can also be

exploitable in improving the representation of samples In the context of the IoT environment, it is still unexplored in terms of its usability with resource constrained IoT boards. However, deeper and adaptive models are again quite necessary to improve the real-time application models and systems where time latency and energy consumption are critical. This emphasizes the importance of approaches to learning that take place on the Web and non-conventional learning, which makes changes to models in motion without requiring a great deal of recalibration. Specific signature-based and protocol-based anomaly presence approaches, as well as behavioral analysis, also looked promising in improving the detection of complex attacks, including botnets and DDoS mostly experienced in IoT. Thus, the authors propose a combination of signature-based and anomaly-based models, which provides a complete set of methods for protection against known and unknown threats. These methodologies are the future of IoT anomaly detection as they aim at integration to build powerful and efficient systems ideal for handling IoT network increasing complexity. More future work should be dedicated to applying explainable AI and increasing the trust in solutions, federated learning for model cooperation with data sharing being kept to the minimum, and transfer learning for solving multiple tasks across various domains. Closing the current research gaps will help make IoT more secure in the future to protect users and key infrastructure from new forms of threats.

REFERENCES

1. E. Irmak and M. Bozdal, “Internet of things (iot): The most up-to-date chal- lenges, architectures, emerging trends and potential opportunities,” Foundation of Computer Science, 2018.
2. F. Armenio, H. Barthel, P. Dietrich, J. Duker, C. Floerkemeier, J. Garrett, M. Har- rison, B. Hogan, J. Mitsugi, J. Preishuber-Pfluegl, et al., “The epcglobalarchitec-ture framework-epcglobal final version 1.3. gs 1,” 2009.
3. D. Evans, “The internet of things: How the next evolution of the internet is changing everything,” CISCO white paper, vol. 1, no. 2011, pp. 1–11, 2011.
4. M. Fagan, K. N. Megas, and D. Lemire, “Summary report for the virtual workshop addressing public comment on nist cybersecurity for iot guidance,” 2021.
5. D. Anstee, D. Bussiere, G. Sockrider, and C. Morales, “Worldwide infrastructure security report,” Arbor Netw., Burlington, MA, USA, Tech. Rep, vol. 9, 2014.
6. E. ENISA, “Baseline security recommendations for iot in the context of critical information infrastructures,” 2017.
7. N. Miloslavskaya and A. Tolstoy, “Internet of things: information security chal- lenges and solutions,” Cluster Computing, vol. 22, no. 1, pp. 103–119, 2019.
8. N. Ye, Y. Zhu, R.-c. Wang, R. Malekian, and Q.-m. Lin, “An efficient authentica- tion and access control scheme for perception layer of internet of things,” 2014.

9. Z. Liu, P. Longa, G. C. Pereira, O. Reparaz, and H. Seo, “Four q on embedded devices with strong countermeasures against side-channel attacks,” IEEE Trans- actions on Dependable and Secure Computing, vol. 17, no. 3, pp. 536–549, 2018.
10. P. Pongle and G. Chavan, “A survey: Attacks on rpl and 6lowpan in iot,” in 2015International conference on pervasive computing (ICPC), pp. 1–6, IEEE, 2015.
11. N. Elsayed, Z. ElSayed, and M. Bayoumi, “Iot botnet detection using an economicdeep learning model,” arXiv preprint arXiv:2302.02013, 2023.
12. M. Ammar, G. Russello, and B. Crispo, “Internet of things: A survey on the security of iot frameworks,” Journal of Information Security and Applications, vol. 38, pp. 8–27, 2018.
13. P. Sethi and S. R. Sarangi, “Internet of things: architectures, protocols, and applications,” Journal of Electrical and Computer Engineering, vol. 2017, 2017.
14. R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, “Internet of things (iot) security: Current status, challenges and prospective measures,” in 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST),pp. 336–341, IEEE, 2015.
15. H. Mrabet, S. Belguith, A. Alhomoud, and A. Jemai, “A survey of iot security based on a layered architecture of sensing and data analysis,” Sensors, vol. 20, no. 13, p. 3625, 2020.
16. H. Cai, B. Xu, L. Jiang, and A. V. Vasilakos, “Iot-based big data storage systems in cloud computing: perspectives and challenges,” IEEE Internet of Things Journal,vol. 4, no. 1, pp. 75–87, 2016.
17. R. J. Alzahrani and A. Alzahrani, “A novel multi algorithm approach to identify network anomalies in the iot using fog computing and a model to distinguish be- tween iot and non-iot devices,” Journal of Sensor and Actuator Networks, vol. 12, no. 2, p. 19, 2023.
18. G. Nayak, A. Mishra, U. Samal, and B. K. Mishra, “Depth analysis on dos &ddosattacks,” Wireless Communication Security, p. 159, 2023.
19. J. Shareena, A. Ramdas, and H. AP, “Intrusion detection system for iot botnet attacks using deep learning,” SN Computer Science, vol. 2, no. 3, p. 205, 2021.
20. S. Behal, A. S. Brar, and K. Kumar, “Signature-based botnet detection and prevention,” in Proceedings of International Symposium on Computer Engineering and Technology, pp. 127–132, 2010.
21. C. Liu, J. Yang, R. Chen, Y. Zhang, and J. Zeng, “Research on immunity- based intrusion detection technology for the internet of things,” in 2011 Seventh International conference on natural computation, vol. 1, pp. 212–216, IEEE, 2011.
22. M. Alhanahnah, Q. Lin, Q. Yan, N. Zhang, and Z. Chen, “Efficient signature generation for classifying cross-architecture iot malware,” in 2018 IEEE conference on communications and network security (CNS), pp. 1–9, IEEE, 2018.
23. I. Apostol, M. Preda, C. Nila, and I. Bica, “Iot botnet anomaly detection using unsupervised deep learning,” Electronics, vol. 10, no. 16, p. 1876, 2021.
24. G. Liu, W. Quan, N. Cheng, H. Zhang, and S. Yu, “Efficient ddos attacks mitiga- tion for stateful forwarding in internet of things,” Journal of Network and Com-puter Applications, vol. 130, pp. 1–13, 2019.

25. M. Nobakht, V. Sivaraman, and R. Boreli, “A host-based intrusion detection andmitigation framework for smart home iot using openflow,” in 2016 11th International conference on availability, reliability and security (ARES), pp. 147–156,IEEE, 2016.
26. A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, and A. Alazab, “A novelensemble of hybrid intrusion detection system for detecting internet of things attacks,” Electronics, vol. 8, no. 11, p. 1210, 2019.
27. E. Bertino and N. Islam, “Botnets and internet of things security,” Computer, vol. 50, no. 2, pp. 76–79, 2017.
28. M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, “A survey on the internet of things (iot) forensics: challenges, approaches, and open issues,” IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1191–1221,2020.
29. G. Kambourakis, M. Anagnostopoulos, W. Meng, and P. Zhou, Botnets: Architectures, countermeasures, and challenges. CRC Press, 2019.
30. A. Thakkar and R. Lohiya, “A review on machine learning and deep learning perspectives of ids for iot: Recent updates, security issues, and challenges,” Archives of Computational Methods in Engineering, pp. 1–33, 2020.
31. D. Oh, D. Kim, and W. W. Ro, “A malicious pattern detection engine for embed- ded security systems in the internet of things,” Sensors, vol. 14, no. 12, pp. 24188– 24211, 2014.
32. M. F. B. Abbas and T. Srikanthan, “Low-complexity signature-based malware detection for iot devices,” in International Conference on Applications and Techniques in Information Security, pp. 181–189, Springer, 2017.
33. P. Kasinathan, C. Pastrone, M. A. Spirito, and M. Vinkovits, “Denial-of-service detection in 6lowpan based internet of things,” in 2013 IEEE 9th international conference on wireless and mobile computing, networking and communications (WiMob), pp. 600–607, IEEE, 2013.
34. X. Jia, D. He, L. Li, and K.-K. R. Choo, “Signature-based three-factor authen- ticated key exchange for internet of things applications,” Multimedia Tools and Applications, vol. 77, pp. 18355–18382, 2018.
35. Y. Fu, Z. Yan, J. Cao, O. Kon´e, and X. Cao, “An automata based intrusion detection method for internet of things,” Mobile Information Systems, vol. 2017, 2017.
36. H. Wang, J. Gu, and S. Wang, “An effective intrusion detection framework basedonsvm with feature augmentation,” Knowledge-Based Systems, vol. 136, pp. 130–139, 2017.
37. W. Wang, Y. Shang, Y. He, Y. Li, and J. Liu, “Botmark: Automated botnet detection with hybrid analysis of flow-based and graph-based traffic behaviors,” Information Sciences, vol. 511, pp. 284–296, 2020.
38. R. Vinayakumar, M. Alazab, S. Srinivasan, Q.-V. Pham, S. K. Padannayil, and K. Simran, “A visualized botnet detection system based deep learning for the internet of things networks of smart cities,” IEEE Transactions on Industry Ap- plications, vol. 56, no. 4, pp. 4436–4456, 2020.

39. M. Mijwil, I. E. Salem, and M. M. Ismaeel, “The significance of machine learning and deep learning techniques in cybersecurity: A comprehensive review,” Iraqi Journal For Computer Science and Mathematics, vol. 4, no. 1, pp. 87–101, 2023.
40. M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, “Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in iot,” Sensors, vol. 17, no. 9, p. 1967, 2017.
41. D. Midi, A. Rullo, A. Mudgerikar, and E. Bertino, “Kalis—a system for knowledge-driven adaptable intrusion detection for the internet of things,” in 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 656– 666, IEEE, 2017.
42. L. Liu, B. Xu, X. Zhang, and X. Wu, “An intrusion detection method for internet of things based on suppressed fuzzy clustering,” EURASIP Journal on Wireless Communications and Networking, vol. 2018, pp. 1–7, 2018.
43. J. Ashraf, M. Keshk, N. Moustafa, M. Abdel-Basset, H. Khurshid, A. D. Bakhshi, and R. R. Mostafa, “Iotbot-ids: A novel statistical learning-enabled botnet de- tection framework for protecting networks of smart cities,” Sustainable Cities and Society, vol. 72, p. 103041, 2021.
44. D. Breitenbacher, I. Homoliak, Y. L. Aung, N. O. Tippenhauer, and Y. Elovici, “Hades-iot: A practical host-based anomaly detection system for iot devices,” in Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, pp. 479–484, 2019.
45. J. Yu, D. Mo, Z. Zhu, and X. M. Chen, “A high-order hidden markov model for dynamic decision analysis of multi-homing ride-sourcing drivers,” Transportation Research Part C: Emerging Technologies, vol. 148, p. 104031, 2023.
46. T. L. von Sperling, F. L. de Caldas Filho, R. T. de Sousa, L. M. e Martins, and
47. R. L. Rocha, “Tracking intruders in iot networks by means of dns traffic analysis,” in 2017 Workshop on Communication Networks and Power Systems (WCNPS), pp. 1–4, IEEE, 2017.
48. L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, “Exposure: A passive dns analysis service to detect and report malicious domains,” ACM Transactionson Information and System Security (TISSEC), vol. 16, no. 4, pp. 1–28, 2014.
49. T. Nagunwa, P. Kearney, and S. Fouad, “A machine learning approach for detect- ing fast flux phishing hostnames,” Journal of Information Security and Applica- tions, vol. 65, p. 103125, 2022.
50. J. Galeano-Brajones, J. Carmona-Murillo, J. F. Valenzuela-Vald´es, and F. Luna-Valero, “Detection and mitigation of dos and ddos attacks in iot-based stateful sdn: An experimental approach,” Sensors, vol. 20, no. 3, p. 816, 2020.
51. N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the devel-opment of realistic botnet dataset in the internet of things for network foren- sic analytics: Bot-iot dataset,” Future Generation Computer Systems, vol. 100, pp. 779–796, 2019.
52. H. Cheng, J. Liu, T. Xu, B. Ren, J. Mao, and W. Zhang, “Machine learning based low-rate ddos attack detection for sdn enabled iot networks,” International Journal of Sensor Networks, vol. 34, no. 1, pp. 56–69, 2020.

53. Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, "Nbaiot network-based detection of iot botnet attacks using deep autoencoders," IEEE Pervasive Computing, vol. 17, no. 3, pp. 12–22, 2018.
54. S. Raza, L. Wallgren, and T. Voigt, "Svelte: Real-time intrusion detection in the internet of things," Ad hoc networks, vol. 11, no. 8, pp. 2661–2674, 2013.
55. J. Krimmling and S. Peter, "Integration and evaluation of intrusion detection for coap in smart city applications," in 2014 IEEE Conference on Communications and Network Security, pp. 73–78, IEEE, 2014.
56. E. Hodo, X. Bellekens, A. Hamilton, P.-L. Dubouilh, E. Iorkyase, C. Tachtatzis, and R. Atkinson, "Threat analysis of iot networks using artificial neural network intrusion detection system," in 2016 International Symposium on Networks, Com- puters and Communications (ISNCC), pp. 1–6, IEEE, 2016.
57. Z. A. Khan and U. Abbasi, "Reputation management using honeypots for intrusion detection in the internet of things," Electronics, vol. 9, no. 3, p. 415, 2020.
58. S.-Y. Lee, S.-r. Wi, E. Seo, J.-K. Jung, and T.-M. Chung, "Profiot: Abnormal behavior profiling (abp) of iot devices based on a machine learning approach," in 2017 27th International Telecommunication Networks and Applications Confer- ence (ITNAC), pp. 1–6, IEEE, 2017.
59. A. Damodaran, F. D. Troia, C. A. Visaggio, T. H. Austin, and M. Stamp, "A comparison of static, dynamic, and hybrid analysis for malware detection," Journal of Computer Virology and Hacking Techniques, vol. 13, pp. 1–12, 2017.
60. B. Molina-Coronado, U. Mori, A. Mendiburu, and J. Miguel-Alonso, "Towards a fair comparison and realistic evaluation framework of android malware detectors based on static analysis and machine learning," Computers & Security, vol. 124, p. 102996, 2023.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.