Investigation of Machine Learning Analytic-Based Data Classification Frameworks on Iot-Based Infrastructures

Dr. Badru Rahmon. A.[1], Adekoya Damola Felix[2], Peter-Ilesanmi Folasade Victoria[3].

[1,2,3]Department of Computer Science., Lead City University, Ibadan

DOI: https://doi.org/10.51244/IJRSI.2024.1102003

Received: 24 January 2024; Accepted: 29 January 2024; Published: 27 February 2024

ABSTRACT

The Internet of Things revolutionizes and enables convenient, automated lifestyles for people today. Over the past ten years, advances in computer, connectivity, and application design have led to its development. The Internet of Things has quickly spread around the world, impacting every person on the planet. Everyday IoT devices, such as smartphones, Google Home assistants, smart vehicles, and automated systems in buildings, including smart elevators and temperature control, as well as drones for environmental monitoring and leisure, play crucial roles in our daily lives. Therefore, the purpose of this research is to explore data classification frameworks based on machine learning analytics in Internet of Things-based infrastructures.

This study focuses on articles published between 2010 and 2023, utilizing a comprehensive literature review approach. The study entails reading through at least eighty peer-reviewed conference proceedings, industry white papers, and journal articles. An overview of the relationship between machine learning and the Internet of Things (IoT) is given in this study. The approach consists of four standard processes that are used in review papers: gathering articles from Thomson Reuters Web of Science, identifying categories, carrying out descriptive analysis, and assessing the information acquired.

A discussion on the categorization of diverse analytics methods for IoT is presented in the paper. Moreover, it introduces and discusses the architectural Framework for IoT and the challenges in IoT Data Classification. Furthermore, Machine Learning Strategies for Identifying and Classifying IoT were presented.

Researchers must investigate machine learning analytic-based data classification frameworks for Internet of Things infrastructures. The challenges arising from the unique characteristics of IoT data call for innovative solutions, and current frameworks demonstrate promising results.

Keywords – Machine learning, Internet of Things, Frameworks, Data classification, Computer security.

INTRODUCTION

The Internet of Things (IoT) revolutionizes and enables convenient, automated lifestyles for people today. Its development is the result of advancements in computing, communication, and application design over the past decade. The influence of IoT has rapidly expanded globally, affecting the entire human population. Everyday IoT devices, such as smartphones, Google Home assistants, smart vehicles, and automated systems in buildings, including smart elevators and temperature control, as well as drones for environmental monitoring and leisure, play crucial roles in our daily lives. The widespread adoption of IoT extends to storage centers, including geographically dispersed back-end cloud facilities, as highlighted by Adi et al.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

[1]. As a result, IoT devices and their platforms generate a large amount of data, which needs to be transferred and then stored and processed at the back-end cloud data centers. IoT devices generate raw data streams all the time, which don't provide meaningful insights unless they're exposed to processing techniques such as knowledge discovery or machine intelligence. The range of data output across different IoT implementations depends on the use case, including smart healthcare, smart social media, smart agriculture, smart e-health, intelligent electricity grids, smart vehicles, etc. Due to the resource constraints of IoT devices and to save power during operation, IoT devices have their protocols. Constrained Application Protocol, Message Queuing Telemetry Transfer, Advanced Message Queuing Protocol, and HyperText Transfer Protocol are some of the most widely used application-layer protocols for the Internet of Things [2].

In IoT devices with access to continuous or renewable power supplies, both Message Queuing Telemetry Transfer and Advanced Message Queuing Protocol can be used for the transfer of longer messages, but at the expense of consuming more energy Constrained Application Protocol, on the other hand, is lightweight and designed for IoT devices with limited computing resources and network bandwidth. Due to its higher resource demands, the HyperText Transfer Protocol is more appropriate for sophisticated IoT devices that possess advanced communication, storage capabilities, and computation. IoT devices generate substantial amounts of data, which undergoes localized processing to a limited extent before being transmitted to a centralized computing node or a cloud storage facility. In these locations, the data can undergo further processing or analysis to extract meaningful knowledge. Machine learning, a set of techniques automating the model-building process of training data with minimal human intervention, plays a crucial role. This results in the automated categorization of data into various classes. Data analytics plays a vital role in processing IoT data, and machine learning plays a significant role in identifying patterns in the data from IoT devices, facilitating the rapid processing of data [1].

Modern computing and storage models encompass cloud, fog, and edge computing. When integrated with the IoT, these paradigms form a robust framework for collecting, storing, processing, and analyzing data. This framework facilitates the implementation of machine learning techniques for intelligent data analytics on the Internet of Things and provides real-time insights into data trends. The cloud paradigm is a centralized data storage model that provides services for central processing and analyzing Internet of Things data, such as Software as a Service, Platform as a Service, and Infrastructure as a Service [3]. Edge computing, on the other hand, makes it easier to handle and analyze IoT data close to the IoT network, usually on localized computer nodes like base stations. By using this method, the expenses related to moving IoT data to centralized nodes are circumvented. Between cloud and edge computing, fog computing removes the need to process and analyze IoT data only at the edge of the network or in a centralized storage facility. Rather, the fog paradigm presents the idea of a virtual platform for IoT data processing and analysis that is not limited to the network's edge [1].

The Internet of Things connects a vast array of objects to the Internet within complex and diverse environments. This interconnected IoT environment generates massive amounts of data, creating a need for efficient storage, processing, and transmission capabilities. Beyond enhancing our daily lives, IoT applications, supported by technologies like Fog, Edge, and Cloud, play crucial roles in sectors such include environmental monitoring, applications for smart cities, remote patient monitoring, disaster mitigation, and precision agriculture. The expectation is that the prevalence of these applications will continue to grow exponentially. The primary limitation lies in the resources of IoT, posing challenges at the network, hardware, and software levels. At different levels of IoT systems, efficient resource management becomes crucial as the number of applications rises. This includes considering factors like battery life, size, processing power, storage, and bandwidth. To acquire, process, and store data from IoT applications, lightweight algorithms and protocols are implemented. IoT devices generate more data than ever before, increasing the demand for processing and storage. Therefore, edge devices, fog nodes, or smart gateways must be integrated to meet this demand [4]. The focus of this paper is to investigate machine learning

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

analytic-based data classification frameworks within IoT-based infrastructures. In this context, the paper aims to contribute specifically to the following research questions:

RQI: What is the categorization of diverse analytics methods for IoT?

RQ2: What is the Architectural Framework for IoT?

RQ3: What are the challenges in IoT Data Classification?

RQ4: What are the Machine Learning Strategies for Identifying and Classifying IoT?

LITERATURE REVIEW

Machine Learning frameworks

The realm of Machine Learning (ML) algorithms boasts a substantial and diverse array of both algorithms and their corresponding software implementations. Several tools for Data Mining have been developed over the past 25 years using Machine Learning techniques [5]. The tools offer integrated programming environments based on standard programming languages that simplify complex data analysis processes. Tools such as analytics platforms, predictive systems, recommender systems, and imaging processors are tailored to a variety of applications. While some focus on processing enormous amounts of data quickly and streaming them, others are experts at applying machine learning methods like Deep Learning and Neural Networks.

It's crucial to underscore that no single tool is universally suitable for all problems, often necessitating a combination of tools for success. Notably, many open-source tools have their code hosted on GitHub repositories [6]. GitHub serves as a repository not only for code but also for comprehensive monitoring information related to software development. This contains the information displayed with graphs and insights on contributors, commits (the team's past and present actions as well as the project's overall activity), watches, issues, and stars, forks. Furthermore, GitHub is integrated with various third-party applications that offer automated code reviews and code analytics. These apps have different perspectives, presentations, evaluation specifics, and preference settings, even if their analytics are based on GitHub data [7].

The intersection of the Internet of Things and machine learning

The goals of IoT encompass comprehending human preferences, understanding thought processes, predicting desired and undesired events, and effectively managing various situations. To achieve these objectives, IoT must interpret the data generated by countless objects, a task that can be accomplished through the utilization of machine-learning algorithms. Machine learning assumes a crucial role within the IoT paradigm, given its inherent ubiquity, making it accessible from anywhere [8]. Machine learning becomes instrumental in extracting insights from the data produced by myriad connected devices, enhancing the overall utility of IoT devices and contributing to the genuine ubiquity of IoT [9]. At the core of this transformative role lies embedded intelligence, which involves integrating product and intelligence to enhance automation, efficiency, productivity, and connectivity [10,11]. Whether operating in the physical or virtual realm, intelligence is acquired through the process of learning.

The inclination of Machine Learning to identify patterns forms the foundation for achieving human-like intelligence. The subsequent extrapolation of these patterns into more meaningful insights and trends enhances our comprehension of the surrounding world. In the context of IoT, the primary aim of ML is to achieve complete automation by fostering intelligent capabilities in objects, leading to smarter decision-

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

making [12]. ML empowers IoT-enabled systems to emulate human-like decision-making processes after being trained on data, thereby improving their understanding of the environment. Information visualization plays a pivotal role in enhancing the comprehension of data and insights by the human visual system [13]. This visualization offers several advantages, such as providing in-depth knowledge without extensive data analysis and leveraging cognitive skills to enhance human understanding of data. IoT has the potential to replace expensive and maintenance-intensive systems with cost-effective ML systems based on sensors. For instance, traditional radar-based weather-monitoring systems are costly and inaccessible in many regions. In contrast, an ML-enabled IoT system, incorporating an affordable sensor network to analyze lighting and cloud patterns for weather prediction, has been successfully deployed in poor countries like Haiti and Guinea, addressing the need for more accessible solutions [14].

The convergence of IoT and machine learning holds promise for improvements in productivity, accuracy, efficiency, and overall cost-effectiveness, especially for resource constrained IoT devices. The collaboration between machine learning algorithms and IoT yields enhanced performance in communication and computation, improved controllability, and more effective decision-making. With improved communication and sophisticated monitoring capabilities ranging from thousands to billions of widely distributed sensing devices, the Internet of Things (IoT) has enormous potential to improve human well-being and find uses for industrial expansion, especially in the framework of Industry 4.0 [1]. The synergy between machine learning and artificial intelligence has significantly enhanced the potential of IoT. The application of advanced machine intelligence techniques enables the extraction of valuable insights from the vast volume of sensory data generated by IoT, facilitating a better understanding of various real-world issues and the ability to make critical operational decisions [1]. Therefore, it is essential that IoT and machine learning work in tandem to solve complex real-world challenges and satisfy computational and communication needs. IoT data analytics has been more important and the subject of much discussion in recent years for several strong reasons.

Significant volume of data generated by dispersed Internet of Things devices.

As per Ericsson's mobility report, the projections indicate a global presence of 18 billion connected IoT devices by 2022 [15]. This figure is anticipated to rise continuously, driven by the widespread integration of IoT devices across various critical applications. The effective and intelligent mining of this vast data pool will be crucial for intelligent data analytics, playing a vital role in identifying and predicting future states of processes or systems.

Diverse data types with significant variations originating from heterogeneous data sources

Due to diverse applications and varying requirements, there is a wide array of IoT devices, ranging from mobile phones, PCs/laptops, and tablets to short-range and wide-area IoT devices. The data derived from these devices exhibit heterogeneity in terms of features, formats, and attributes, stemming from the distinct nature of their applications. For instance, IoT devices utilized in medical applications differ from those in smart home IoT setups. Managing the quality, processing, and storage of data has become challenging due to this heterogeneity. In their work [16], the authors address critical questions arising from the diversity of data sources. These include tasks like managing the high-frequency streaming data sampling process, noise cancellation and filtering, collecting, and combining data from multiple sources, guaranteeing data interpretation and interoperability, reasoning, situation awareness, deriving knowledge from the data, and gathering and storing data from multiple sources following application constraints [16].

Variability and unpredictability in the data streams of IoT

Practical data analysis often encounters uncertainty, as noted by Anwar, Mahmood, and Pickering [17].

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

Uncertainty in IoT data streams can come from several places, including IoT device or communication channel failures during data transfer. IoT data streams commonly exhibit gross errors and missing data, necessitating advanced analytics for effective data preprocessing. Additionally, cyber intrusions can contribute to uncertainty in the data, as highlighted by Stellios et al. [18]. Evaluating, propagating, and representing uncertainties appropriately is essential to improving the accuracy of decision-making processes. The development of models and solutions capable of addressing these factors is essential, as emphasized by Sun et al. [16].

Striking a balance between scalability and efficiency

The majority of IoT data analytics takes place in the cloud. However, the process of transferring data from IoT devices to the cloud incurs a significant cost in terms of delay, posing challenges for time-sensitive applications, particularly when dealing with many IoT devices. For example, in a connected vehicle situation, a sizable fleet of vehicles could have to make judgments in real-time or very close to it. Therefore, as the number of cars rises, it becomes increasingly important to balance the speed and accuracy of the analysis.

Categorization of diverse analytics methods for IoT

Descriptive analytics

IoT systems can collect data from a small number to thousands of intelligent devices, transmitting this data to a cloud environment. Utilizing advanced machine-learning techniques on historical data allows for the generation of detailed insights into past events. Descriptive analytics is a field that consists of collections of machine learning-based algorithms that process and summarize unprocessed data to provide meaningful insights. Examples of descriptive analytics include data aggregation, data summarization, mathematical logical operations, and data mining using clustering algorithms. This kind of analytics works especially well with large amounts of data. Recent developments in technology have shown that cloud storage can manage enormous volumes of Internet of Things (IoT) data. IoT cloud

IoT predictive analytics

Advanced statistical or machine learning techniques are used in predictive analytics to model behavior or patterns based on historical data, which allows for the prediction of potential future trends or patterns in the data. In essence, it predicts future events by analyzing data correlations and past trends. Predictive analytics finds widespread applications, including predictive maintenance, forecasting price trends, predicting supply-demand dynamics, or determining the likelihood of specific outcomes. Two categories of predictive models exist, according to SAS, a prominent analytics company: (i) regression-based models, which forecast a numerical value based on likelihood and past data, and (ii) classification-based models, which carry out prediction studies based on class membership [19]. The latest methods for predictive modeling include decision trees, neural networks, and deep neural networks, combined with statistical regression. Gradient boosting, ensemble model-based analysis, and Bayesian analysis are other popular techniques. For well-informed decision-making, these predictive analytics methods mostly rely on data. The Internet of Things paradigm is essential for optimizing the data collection process from intelligent IoT devices and for offering an analytical framework that can be used with the cloud or the network edge.

Prescriptive Analytics for IoT

Prescriptive analytics is the practice of making recommendations about future courses of action based on

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

data analysis. Unlike predictive analytics, which forecasts future states, prescriptive analytics not only predicts outcomes but also offers recommendations for adopting those outcomes. It functions as a method for future scenario analysis that combines the advantages of predictive and descriptive analytics.

Prescriptive analytics provides insights into future predictions coupled with impact studies, whereas predictive analytics only reveals what and when an event might occur based on future projections. This approach is extensively employed for optimizing business outcomes. Prescriptive analytics works especially well in Industrial IoT environments where machine learning, cloud/edge computing, and business intelligence are used to make decisions based on facts rather than hypotheses. Services within an IoT cloud platform can help with decision-making by utilizing analytics and business intelligence technologies.

IoT-based Adaptive Analytics.

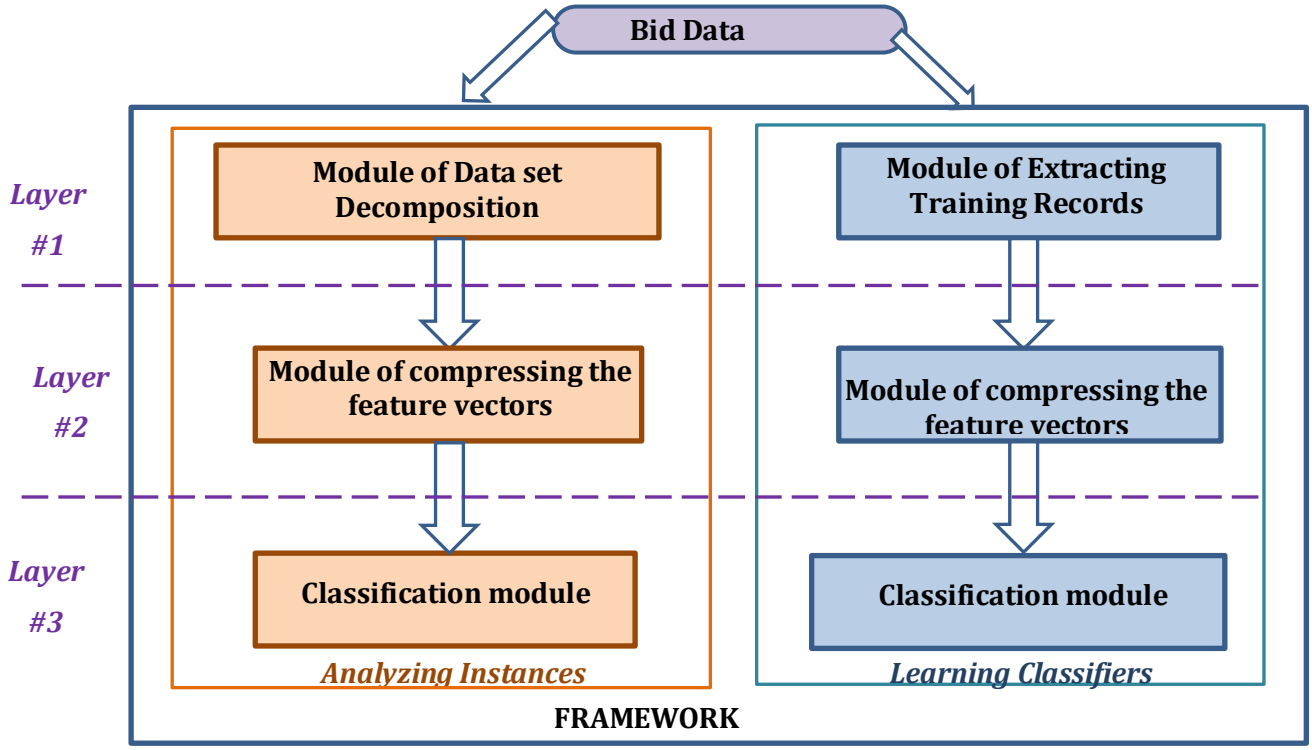
In the practical implementation phase, the results derived from predictive analytics need to be adjusted based on real-time data. By considering the recent history of the data and looking at their correlations, adaptive analytics can be used to change or optimize the process outcome. This type of analysis helps to reduce errors and improve model performance. The capacity of adaptive analytics to modify the solution's result when new sets of input data are received is one of its main advantages. Adaptive analytics are a good fit for real-time stream data processing, especially in an Internet of Things setting. The use of adaptive analytics for efficient data analysis can also be advantageous for the real-time assessment of dynamic data streams, such as those found in malware scenarios [20].

Architectural Framework for IoT

The architecture framework created for the mobile Internet of Things security monitoring is shown in Figure 1. The explicit distinction between the two operational modes—the training mode and the analysis mode—is the central idea of this architectural design. The analyzed dataset is first divided up among several classifiers. For every binary classifier, a specific training sample is created during the training phase. The dataset is first split into segments in the analysis mode, and separate classifier copies are allocated to handle each segment. A unique training sample is created for every binary classifier throughout the training phase. First, the dataset is split into segments, and separate classifier copies are assigned to handle each segment in the analysis mode. Then, the compositions and parameters of the basic classifiers are adjusted in the training mode, and the classification indicators of the trained classifiers are calculated in the analysis mode. Consequently, the suggested design implements the MapReduce concept: in the first two stages, incoming data are processed beforehand and distributed among several processes (map), and in the third stage, outputs are aggregated (reduce).

The two operational modes included in the designed framework for mobile IoT security monitoring are the instance analysis mode and the classifier learning mode. Three layers comprise each mode: (1) data set extraction and deconstruction; (2) feature vector compression; and (3) learning and classification. The examined data set is decomposed at the first stage, and the training sample is generated. Typical representative training sample construction techniques, like CADEX and DUPLEX [21], have quadratic complexity when it comes to the cardinality of the source data set. This feature may harm how well algorithms are intended to handle large volumes of data function. But instead of using the full data set, Kotenko, Saenko, and Branitskiy [21] used a heuristic approach to extract training instances, eliminating strongly correlated records from randomly chosen groups.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024



Challenges in IoT Data Classification

A highly networked society that bridges the gap between the physical and digital realms is possible with the IoT paradigm. It must, however, overcome major technological and non-technological obstacles to accomplish this objective. These issues can be divided into four groups based on information gleaned from several IoT literature evaluations [23, 24, 25]: (1) technology, (2) individual, (3) business, and (4) social. A brief explanation of each is given in the subsections that follow.

Technological Difficulties

Although it is still a way off, the global deployment of IoT systems is beginning to take shape in many regions of the world. Four essential technologies—sensing, networking, actuators, and security—must be integrated for IoT to function. But there are still problems with connectivity, especially when it comes to mobile and internet availability, especially in low-income nations. Furthermore, the lack of cross-platform compatibility in the existing IoT platform has contributed to its sluggish adoption [26]. The client-server architecture that currently dominates the Internet of Things (IoT) landscape might not be viable in the future because of increased latency, increased energy consumption, single-point failure, and security risks. Though they are still in their infancy and suffer difficulties with network capacity, latency, accessibility, control, and management, developing edge and fog computing platforms have emerged to address these worries [27]. The security of these networked devices is a key challenge that seriously jeopardizes the acceptability of IoT systems on a large scale [26]. Because the Internet of Things is decentralized and allows devices to connect with other companies, there is a greater vulnerability to cyberattacks. Security flaws are made worse by the lack of global standards for authorization and authentication of IoT devices. For the IoT paradigm to be successfully implemented and widely accepted, security problems must be addressed.

Challenges at the Individual Level

The goal of the Internet of Things is to improve people's lives by providing enhanced automation and an intelligent environment, catering to a variety of demands. Because of the "anytime and anywhere" aspect of IoT, privacy becomes a major problem that varies from person to person and presents more difficult issues

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

for servers and devices to handle. Abi et al.'s [28] analysis of privacy preservation tactics and new IoT trends is crucial. The public's acceptance of IoT presents another challenge to its success. How effectively our cognitive needs match these changes is a challenge raised by the revolutionary impact of IoT on daily life [29]. The importance of public acceptability has been examined by Bestepe and Yildirim [30], particularly in the context of smart cities that use IoT infrastructure for sustainability. People need to be educated and trained about the advantages of these cutting-edge services and applications and how they can transform everyday life towards higher wealth to increase the adoption of IoT.

Challenges in the Business Realm

IoT has not lived up to the hype that has been created, even though it offers substantial commercial prospects in manufacturing, applications, and services. Companies are facing a few difficulties, such as the lack of industry standards, universal platforms, compatibility problems, connectivity problems, difficulties gathering data, and security problems [31]. A further major problem is the lack of qualified IoT specialists. IoT developments have also slowed down and lost interest from businesses and governments as priorities have changed due to global market disruptions like the COVID-19 pandemic [32], a global computer chip shortage [33], and geopolitical events like the Russia-Ukraine war [34].

Challenges in the Societal Context

As a society, the aspiration is to foster a thriving and sustainable living environment. IoT holds the potential to significantly contribute by providing support for actionable decision-making [35]. However, the pivotal questions that arise are whether our society is currently equipped to effectively leverage IoT and if we are prepared to embrace the cognitive changes it will introduce into our daily lives. The answers to these fundamental questions will play a crucial role in shaping the public acceptance of IoT applications and services [36]. Addressing these concerns will not only boost the demand side but also prompt industries to invest substantial efforts in the widespread deployment of IoT, its applications, and services. Furthermore, one of the major challenges arising from the imbalance in global economic growth is the current digital divide.

FINDING AND RESULT

Creating frameworks for enhancing IoT security through deep learning

The great capacity of deep learning algorithms to generalize data makes them popular for use in high-dimensional object analysis. The tuning mechanism for these structures is designed to imitate the process of building a set that includes items that are encountered during training as well as ones that are not. A deep learning-based approach was presented by Zhou et al. [37] with the express purpose of identifying internet intrusions on the Internet of Things. The authors emphasize how useful the framework is in the context of a "smart city." Importantly, this method balances the trade-off between attack detection precision and processing time, allowing for an improvement in classification accuracy.

Nguyen et al. [38] used three datasets (NSL-KDD, UNSW-NB15, and KDDCup 1999) to investigate the potential of deep learning for cyberattack detection. Their developed framework consists of four steps that are applied sequentially: (1) using principal component analysis to reduce dimension and redundancy in input data; (2) using the Gaussian binary restricted Boltzmann machine to pre-train the neural network; (3) deep learning involving iterative weight adjustments; and (4) using softmax regression to construct the output signal. In a different study, Diro and Chilamkurti [39] concentrated on building a deep learning model-based distributed attack detection system that was especially utilized to identify attacks in the social IoT. Because of the way the system is designed, it is possible to update the learning parameters on a single master node and then distribute those adjustments to worker nodes. Deep neural network distributed

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

learning was discussed by Alsheikh et al. [40] in the context of mobile big data analytics. Their method, which is applied in the Spark environment, consists of repeating the processes of parameter dispersion among worker nodes, parameter averaging, and partial model learning until convergence requirements are satisfied.

Creation of Frameworks for Big Data

Numerous works delve into the development of frameworks for Big Data, typically involving the execution of MapReduce operations and relying on specialized systems such as Hadoop, Spark, Flink, and others [41, 42]. These frameworks, with extensive scopes, find applications in various domains, including mobile networks and mobile IoT. For example, Kim et al. [44] talked about a framework for medical data analysis, Zygouras et al. [45] investigated a framework for monitoring data on bus traffic control, and Scherbakov et al. [43] presented a framework for processing online applications. Furthermore, there are Big Data frameworks specifically designed for computer security [46, 47]. Nevertheless, there is still work to be done in terms of effectively integrating machine learning algorithms into these frameworks. Branitskiy and Kotenko [48, 49] offered a mixed classifier strategy; nevertheless, the particulars of Big Data processing have not been fully explored.

Utilizing Machine Learning for Computer Security

The first group's studies demonstrate the increasing importance of using machine learning techniques in computer security. Chan and Lippmann [50] claim that increased connection brought about by the widespread usage of computers in mobile networks, particularly mobile IoT, has made these networks more vulnerable to a wide range of pervasive and varied attacks. The effectiveness of machine learning algorithms can be advantageous for traditional security software, which relies heavily on human labor to identify threats. A distributed Support Vector Machine technique was successfully implemented, as shown by Shamili et al. [51], to identify dangerous software (malware) on mobile device networks. Sahs and Khan [52] introduced a machine learning-based system for detecting malware in Android devices, utilizing Support Vector Machine algorithms. Combining several machine learning methods, however, can result in even higher efficacy. According to Joseph et al. [53], mobile device protection combined with aided malware analysis is a promising use for machine learning. In an evaluation of machine learning-based intrusion detection systems, Ford and Siraj [54] emphasized the systems' flexibility in responding to novel and unidentified threats. The evaluation considers several attacks on mobile IoT networks, including monitoring energy use from smart meters and robbing security cameras.

An overview of previous studies concentrating on the use of machine learning techniques for mobile threat identification was carried out by Arslan et al. [55]. According to their research, Support Vector Machine algorithms are especially well-liked, and detection systems that use machine learning techniques have success rates between 80% and 99.6%. IoT security solutions that make use of machine learning techniques like supervised learning, unsupervised learning, and reinforcement learning were investigated in a study by Xiao et al. [56]. These techniques were used for malware detection, safe offloading, access control, and authentication, among other IoT security concerns. This analysis has shown that the accompanying computer and communication overheads are a major barrier to deploying machine learning-based security techniques in real-world IoT systems.

Machine Learning Strategies for Identifying and Classifying IoT

In industrial automation control systems, Falk and Fries [57] suggested some authentication techniques for device identification and whitelisting (a list of approved devices). The study, which focused on industrial automation control systems, demonstrated success in settings where devices had established communication relationships. It did, however, highlight the possibility of failure in dynamic, expansive organizational

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

contexts where new device types are frequently introduced. In a different study, Meidan et al. [58] extracted features from network traffic data for unauthorized device identification in intelligent environments using random forest, a machine learning technique. After manually labeling the traffic data and training a multiclass classifier for each type of IoT device, the researchers gathered data from 27 different IoT devices of nine different sorts. The classifier properly classified the remaining eight classes while effectively identifying an unlawful device. After manually labeling the traffic data and training a multiclass classifier for each type of IoT device, the researchers gathered data from 27 different IoT devices of nine different sorts. The classifier properly classified the remaining eight classes while effectively identifying an unlawful device. Nevertheless, restrictions included the lack of consideration for complex mixed real-time traffic and the classification of particular device types.

A technique for categorizing IoT and non-IoT devices was presented by Sivanathan et al. [59] using three weeks' worth of network traffic data collection. The authors were able to classify devices accurately by using a random forest multiclass classifier in conjunction with 12 parameters that were taken from network traffic, such as protocols, packet length, and port number. This method's drawback is that it requires the classifier to be trained for every device, which renders it unfeasible for many IoT devices available in the commercial market. Conversely, Pêgo and Nunes [60] created an application to automatically identify a new device's class based on its attributes. To identify devices engaging inside a network based on data shared by IoT devices, this program creates an interface and integration drivers for new devices. The study addressed typical integration issues for platforms containing heterogeneous IoT devices and investigated the accuracy of several machine-learning approaches for device discovery in the IoT smart environment. These findings contribute to the automation of IoT devices in intelligent settings. Using an iPhone OS application, the researchers used communication data gathered from smart environment traffic to create a database containing device information. TF-IDF tables, synonyms match, multi-property matching, the Levenshtein distance method, and other machine learning techniques were used to precisely identify devices that were interacting within the IoT network.

Ferrando and Stacey [61] discussed the difficulties in protecting IoT devices and suggested a unique security detection method for data streams. This approach attempts to categorize threats at an early stage, with a special emphasis on identifying a wide range of network irregularities and classifying sensor-generated information. The method was assessed by the researchers as an anomaly detection methodology, utilizing information from a network device to identify and categorize different types of anomalies. It was hypothesized that a variety of anomalies may be effectively detected and classified by looking at the feature distribution in network data. Similarly, Shen et al. [62] showed how several supervised machine-learning methods may be used to examine information gathered from traffic in intelligent environments. Accurately identifying illegal IoT devices was their main objective to safeguard the confidential data of a company. Utilizing a manually labeled dataset derived from the network traffic data of twenty-seven Internet of Things devices across nine different categories, the researchers trained and assessed a multiclass classifier. The findings accurately classified the remaining devices as allowed, while the ninth category was appropriately identified as unknown.

Suárez and Salcedo [63] collected data from twelve different devices, including lights, refrigerators, cameras, and sensors, and used a variety of categorization approaches, such as K-means and ID3. They made use of twelve characteristics that were taken out of the network communication data of Internet of Things devices, including Bluetooth capabilities, memory size, needed internet bandwidth, battery life, and gateway utilization. They divided the gadgets into four groups using these shared characteristics together with machine learning algorithms. After testing with three, four, and five clusters, K-means was able to classify the devices into four groups: fixed followers, mobile orchestrators, mobile orchestrators, and dummy followers.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

METHODOLOGY

This study focuses on articles published between 2010 and 2023, utilizing a comprehensive literature review approach. The study entails reading through at least eighty peer-reviewed conference proceedings, journal articles, and industry white papers. An overview of the relationship between machine learning and the Internet of Things (IoT) is given in this study. The approach consists of four standard processes that are used in review papers: gathering articles from Thomson Reuters Web of Science, identifying categories, carrying out descriptive analysis, and assessing the information acquired.

CONCLUSION

As the Internet of Things (IoT) has grown, a sizable number of IoT devices have been installed in a variety of settings, including homes, workplaces, warehouses, and roads. Ensuring the security of IoT devices is paramount due to the distinct properties of various IoT devices. In summary, exploring ML analytic-based data classification frameworks for IoT infrastructures is a critical area of research. The challenges arising from the unique characteristics of IoT data call for innovative solutions, and current frameworks demonstrate promising results. As technology advances, opportunities for improvement, such as Explainable AI and federated learning, should be investigated to ensure the ongoing efficacy and ethical use of machine learning in IoT contexts.

REFERENCES

1. Adi E, Anwar A, Baig Z, Zeadally S. Machine learning and data analytics for the IoT. Neural computing and applications. 2020 Oct; 32:16205-33.
2. Naik N. Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In2017 IEEE international systems engineering symposium (ISSE) 2017 Oct 11 (pp. 1-7). IEEE.
3. Cao H, Wachowicz M, Renso C, Carlini E. Analytics everywhere: generating insights from the internet of things. Ieee Access. 2019 May 28; 7:71749-69.
4. Farooq O, Singh P, Hedabou M, Boulila W, Benjdira B. Machine Learning Analytic-Based Two-Staged Data Management Framework for Internet of Things. Sensors. 2023 Feb 22;23(5):2427.
5. Jovic A, Brkic K, Bogunovic N. An overview of free software tools for general data mining. In2014 37th International convention on information and communication technology, electronics and microelectronics (MIPRO) 2014 May 26 (pp. 1112-1117). IEEE.
6. GitHub (2018) GitHub the world's leading software development platform. https://github.com/. Accessed 15. Sept 2018.
7. Nguyen G, Dlugolinsky S, Bobák M, Tran V, López García Á, Heredia I, Malík P, Hluchý L. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review. 2019 Jun 1; 52:77-124.
8. Mattern F, Floerkemeier C. From the Internet of Computers to the Internet of Things. Springer Berlin Heidelberg; 2010.
9. Tanwar S, Garg J, Gupta M, Rana A. Opportunities and challenges in machine learning with IoT. Machine Learning Paradigm for Internet of Things Applications. 2022 Mar 4:209-27.
10. Guo B, Zhang D, Wang Z. Living with internet of things: The emergence of embedded intelligence. In2011 International conference on internet of things and 4th international conference on cyber, physical and social computing 2011 Oct 19 (pp. 297-304). Ieee.
11. Guo B, Zhang D, Yu Z, Liang Y, Wang Z, Zhou X. From the internet of things to embedded intelligence. World Wide Web. 2013 Jul; 16:399-420.
12. Jain, R. Recent Machine Learning Applications to Internet of Things (IoT). Abstract: Table of Contents: 1–19. Available online: https://www.cs.wustl.edu/~ {}jain/cse570-15/ftp/iot_ml/#sec7

(accessed on 2 June 2022).
13. Keim DA, Munzner T, Rossi F, Verleysen M. Bridging information visualization with machine learning (dagstuhl seminar 15101). InDagstuhl reports 2015 (Vol. 5, No. 3). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
14. Davarzani, B.L.; Purdy, M. The Internet of Things Is Now a Thing 2015. Available online: https://ssir.org/articles/entry/the_internet_of_things_is_now_a_thing (accessed on 2 June 2022).
15. Internet of things forecast mobility report (2019). https://www.ericsson.com/en/mobility-report/internet-of-things-forecast. Accessed 18 Oct 2019
16. Sun Y, Song H, Jara AJ, Bie R. Internet of things and big data analytics for smart and connected communities. IEEE access. 2016 Feb 12; 4:766-73.
17. Anwar A, Mahmood AN, Pickering M. Modeling and performance evaluation of stealthy false data injection attacks on smart grid in the presence of corrupted measurements. Journal of Computer and System Sciences. 2017 Feb 1;83(1):58-72.
18. Stellios I, Kotzanikolaou P, Psarakis M, Alcaraz C, Lopez J (2018) A survey of iot-enabled cyberattacks: assessing attack paths to critical infrastructures and services. IEEE Commun Surv Tutor 20(4):3453–3495. https://doi.org/10.1109/COMST. 2018.2855563
19. Predictive analytics history & current advances. SAS. https://www.sas.com/en_au/insights/analytics/predictive-analytics. html. Accessed 15 Jan 2020.
20. Vu DL, Nguyen TK, Nguyen TV, Nguyen TN, Massacci F, Phung PH. HIT4Mal: Hybrid image transformation for malware classification. Transactions on Emerging Telecommunications Technologies. 2020 Nov;31(11):e3789.
21. Reitermanova, Z. "Data splitting," in Proc. WDS's Contributed Papers Matfyzpress, Prague, Czech Republic, vol. 10, 2010, pp. 3136.
22. Kotenko I, Saenko I, Branitskiy A. Framework for mobile Internet of Things security monitoring based on big data processing and machine learning. IEEE Access. 2018 Nov 18; 6:72714-23.
23. Nižetić S, Šolić P, Gonzalez-De DL, Patrono L. Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. Journal of cleaner production. 2020 Nov 20; 274:122877.
24. Malekshahi Rad M, Rahmani AM, Sahafi A, Nasih Qader N. Social Internet of Things: vision, challenges, and trends. Human-centric Computing and Information Sciences. 2020 Dec;10(1):1-40.
25. Sadique KM, Rahmani R, Johannesson P. Towards security on internet of things: applications and challenges in technology. Procedia Computer Science. 2018 Jan 1; 141:199-206.
26. O'Halloran, J. Connectivity Issues Disrupting Most Businesses' IoT Roll-Outs. Available online: https://www.computerweekly. com/news/252509991/Connectivity-issues-disrupting-most-businesses-IoT-roll-outs (accessed on 2 June 2022).
27. Rahul, Y.; Weizhe, Z.; Neeraj, K.; Omprakash, K. Recent Trends and Challenges in Fog/Edge Computing for Internet of Things. Available online: https://www.hindawi.com/journals/wcmc/si/975923/ (accessed on 2 June 2022).
28. Abi Sen AA, Eassa FA, Jambi K, Yamin M. Preserving privacy in internet of things: a survey. International Journal of Information Technology. 2018 Jun; 10:189-200.
29. Falcone R, Sapienza A. On the users' acceptance of IoT systems: A theoretical approach. Information. 2018 Mar 1;9(3):53.
30. Be¸stepe, F.; Yildirim, S.Ö. Acceptance of IoT-Based and Sustainability-Oriented Smart City Services: A Mixed Methods Study. Sustain. Cities Soc. 2019, 526, 121009. [CrossRef]
31. Patil, Y. Key Challenges to Consider for Successful IoT Implementation. Available online: https://www.saviantconsulting.com/blog/iot-implementation-challenges-enterprises.aspx (accessed on 2 June 2022).
32. Business. News IoT News—IoT Market Growth Will Slow down due to the COVID-19 Pandemic and the Limited Take-up of LPWA Solutions. Available online: https://iotbusinessnews.com/2021/02/25/20654-iot-market-growth-will-slow-down-due- to-the-covid-19-pandemic-and-the-limited-take-up-of-lpwa-solutions/ (accessed on 2 June 2022).

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

33. Developer Burnout and a Global Chip Shortage: The IoT Is Facing a Perfect Storm|ZDNet. Available online: https://www.zdnet. com/article/developer-burn-out-and-a-global-chip-shortage-the-iot-is-facing-a-perfect-storm/ (accessed on 2 June 2022).
34. News European IoT Spending Continues Its Double-Digit Growth, Despite Global Uncertainty and Slow Demand, Says IDC. Available online: https://www.idc.com/getdoc.jsp?containerId=prEUR149276822 (accessed on 2 June 2022).
35. Bernard Marr The 5 Biggest Internet of Things (IoT) Trends In 2021 Everyone Must Get Ready for Now. Available on-line: https://www.forbes.com/sites/bernardmarr/2020/10/26/the-5-biggest-internet-of-things-iot-trends-in-2021-everyone-must-get-ready-for-now/?sh=5c8dd2c641fd (accessed on 2 June 2022).
36. Chohan SR, Hu G. Success factors influencing citizens' adoption of IoT service orchestration for public value creation in smart government. IEEE Access. 2020 Nov 5; 8:208427-48.
37. Zhou Y, Han M, Liu L, He JS, Wang Y. Deep learning approach for cyberattack detection. InIEEE INFOCOM 2018-IEEE conference on computer communications workshops (INFOCOM WKSHPS) 2018 Apr 15 (pp. 262-267). IEEE.
38. Nguyen KK, Hoang DT, Niyato D, Wang P, Nguyen D, Dutkiewicz E. Cyberattack detection in mobile cloud computing: A deep learning approach. In 2018 IEEE wireless communications and networking conference (WCNC) 2018 Apr 15 (pp. 1-6). IEEE.
39. Diro AA, Chilamkurti N. Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems. 2018 May 1;82:761-8.
40. A. Alsheikh, D. Niyato, S. Lin, H.-P. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," IEEE Netw., vol. 30, no. 3, pp. 22 29, May/Jun. 2016.
41. Holmes, A. Hadoop in Practice. Greenwich, CT, USA: Manning Publications, 2012.
42. Shoro AG, Soomro TR. Big data analysis: Ap spark perspective. Global Journal of Computer Science and Technology: C Software & Data Engineering. 2015;15(1):7-14.
43. Shcherbakov M, Kachalov D, Kamaev V, Shcherbakova N, Tyukov A, Sergey S. A design of Web application for complex event processing based on hadoop and java servlets. International Journal of Soft Computing. 2015;10(3):218-9.
44. Kim MJ, Yu YS. Development of real-time big data analysis system and a case study on the application of information in a medical institution. International Journal of Software Engineering and Its Applications. 2015 Jul;9(7):93-102.
45. Zygouras N, Zacheilas N, Kalogeraki V, Kinane D, Gunopulos D. Insights on a scalable and dynamic traffic management system. InEDBT 2015 Mar (pp. 653-664).
46. Garware CP, Tidke BA. A security framework for big data computing through distributed cloud data centres in G-hadoop. International Journal of Computer Science and Mobile Computing. 2016 Jun;5(6):355-60.
47. Kotenko I, Kuleshov A, Ushakov I. Aggregation of elastic stack instruments for collecting, storing and processing of security information and events. In2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) 2017 Aug 4 (pp. 1-8). IEEE.
48. Branitskiy A, Kotenko I. Hybridization of computational intelligence methods for attack detection in computer networks. Journal of Computational Science. 2017 Nov 1; 23:145-56.
49. Branitskiy A, Kotenko I. Network anomaly detection based on an ensemble of adaptive binary classifiers. InComputer Network Security: 7th International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security, MMM-ACNS 2017, Warsaw, Poland, August 28-30, 2017, Proceedings 7 2017a (pp. 143-157). Springer International Publishing.
50. Chan, P. K. and Lippmann, R. P. "Machine learning for computer security,' J. Mach. Learn. Res., vol. 7, pp. 2669_2672, Dec. 2006.
51. Sharifi Shamili A, Bauckhage C, Alpcan T. Malware detection on mobile devices using distributed machine learning.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XI Issue II February 2024

52. Sahs J, Khan L. A machine learning approach to android malware detection. In2012 European intelligence and security informatics conference 2012 Aug 22 (pp. 141-147). IEEE.
53. Joseph AD, Laskov P, Roli F, Tygar JD, Nelson B. Machine learning methods for computer security (Dagstuhl Perspectives Workshop 12371). InDagstuhl Manifestos 2013 (Vol. 3, No. 1). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
54. Ford V, Siraj A. Applications of machine learning in cyber security. In Proceedings of the 27th international conference on computer applications in industry and engineering 2014 Oct 13 (Vol. 118). Kota Kinabalu, Malaysia: IEEE Xplore.
55. Arslan B, Gunduz S, Sagiroglu S. A review on mobile threats and machine learning based detection approaches. In 2016 4th International Symposium on Digital Forensic and Security (ISDFS) 2016 Apr 25 (pp. 7-13). IEEE.
56. Xiao L, Wan X, Lu X, Zhang Y, Wu D. IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? IEEE Signal Processing Magazine. 2018 Sep 3;35(5):41-9
57. Falk R, Fries S. Using managed certificate whitelisting as a basis for internet of things security in industrial automation applications. International Journal on Advances in Security Volume 8, Number 1 & 2, 2015. 2015.
58. Meidan Y, Bohadana M, Shabtai A, Ochoa M, Tippenhauer NO, Guarnizo JD, Elovici Y. Detection of unauthorized IoT devices using machine learning techniques. arXiv preprint arXiv:1709.04647. 2017 Sep 14.
59. Sivanathan A, Sherratt D, Gharakheili HH, Radford A, Wijenayake C, Vishwanath A, Sivaraman V. Characterizing and classifying IoT traffic in smart cities and campuses. In2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) 2017 May 1 (pp. 559-564). IEEE.
60. Pêgo PR, Nunes L. Automatic discovery and classifications of IoT devices. In2017 12th Iberian Conference on Information Systems and Technologies (CISTI) 2017 Jun 21 (pp. 1-10). IEEE.
61. Ferrando R, Stacey P. Classification of device behaviour in internet of things infrastructures: towards distinguishing the abnormal from security threats. In Proceedings of the 1st International Conference on Internet of Things and Machine Learning 2017 Oct 17 (pp. 1-7).
62. Shen J, Li Y, Li B, Chen H, Li J. IoT eye an efficient system for dynamic IoT devices auto-discovery on organization level. In2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud) 2017 Jun 26 (pp. 294-299). IEEE.
63. Suárez JN, Salcedo A. ID3 and k-means based methodology for Internet of Things device classification. In2017 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE) 2017 Nov 21 (pp. 129-133). IEEE.