

Information Security

ISSN: 2073-607X, 2076-0930

Volume 14 Issue 02 Year 2022 Page 203:221

Security of Big Data over IoT Environment by Integration of Deep Learning and Optimization

Dr. N.Noor Alleema¹, Dr. Ramakrishnan Raman², Fidel Castro-Cayllahua³, Mr. Vinod Motiram Rathod⁴, Juan Carlos Cotrina-Aliaga⁵, Mrs. Supriya Sanjay Ajagekar⁶, Mrs. Reshma Ramakant Kanse⁷

 ¹Associate Professor, Department of Information Technology, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India <u>drnooralleeman@veltech.edu.in</u>
²Professor and Director, Symbiosis Institute of Business Management, Pune & Symbiosis International (Deemed University), Pune, Maharashtra, India <u>raman06@yahoo.com</u>
³Associate Professor, Universidad Peruana los Andes, Peru, South America <u>d.fcastro@upla.edu.pe</u>
⁴Assistant Professor, Department of Computer Science and Engineering, Bharati Vidyapeeth Deemed University Department of Engineering and Technology, Navi Mumbai, Maharashtra, India <u>vinod.rathod@bvucoep.edu.in</u>
⁵Associate Professor, Department of Post Grade, Universidad Cesar Vallejo, Peru, South America <u>jcotrinaal@ucvvirtual.edu.pe</u>
^{6,7}Assistant Professor of AI and ML, Bharati Vidyapeeth Deemed University Department of Engineering and Technology, Navi Mumbai, India

⁶ssajagekar@bvucoep.edu.in, ⁷rrkanse@bvucoep.edu.in

Article History	Abstract
Received: 11 April 2022 Revised: 26 July 2022 Accepted: 10 August 2022	This is especially true given the spread of IoT, which makes it possible for two-way communication between various electronic devices and is therefore essential to contemporary living. However, it has been shown that IoT may be readily exploited. There is a need to develop new technology or combine existing ones to address these security issues. DL, a kind of ML, has been used in earlier studies to discover security breaches with good results. IoT device data is abundant, diverse, and trustworthy. Thus, improved performance and data management are attainable with help of big data technology. The current state of IoT security, big data, and deep learning led to an all-encompassing study of the topic. This study examines the interrelationships of big data, IoT security, and DL technologies, and draws parallels between these three areas. Technical works in all three fields have been compared, allowing for the development of a thematic taxonomy. Finally, we have laid the groundwork for further investigation into IoT security concerns by identifying and assessing the obstacles inherent in using DL for security utilizing big data. The security of large data has been taken into consideration in this article by categorizing various dangers using a deep learning method. The purpose of optimization is to raise both
CC License CC-BY-NC-SA 4.0	accuracy and performance. Keywords: Big Data, IoT, Deep Learning, Optimization, Security

1. Introduction

1.1 Big Data

It is sometimes referred to as having a large variety of data, increasing volumes, and accelerating pace. Big data in particular is being gathered in enormous quantities, which increases complexity. These enormous data volumes cannot be handled by conventional technology. Large amounts of data are referred to as "big data". It is important to realize that value frequently only constitutes a small portion of the entire volume to deal with a particular business case. It is impossible to find this crucial element without prior research. When we talk about "big data," we mean a lot of data. Every second, the organizations that need to be handled generate vast amounts of data. Big Data will therefore gather, store, and organize data so that analysts may investigate it further. To put it another way, it refers to a huge quantity of data that can be mined for useful information. The amount and complexity of big data, particularly from new sources, are increasing. Traditional data processing approaches can't handle these massive datasets. Despite the sheer volume of data, it may be put to good use in solving business problems that you were previously unable to manage.

Since the amount of data generated is enormous and changes rapidly, innovative data processing methods are required to make better judgments, get a better knowledge of processes, and make them run more efficiently. If conventional or current technology has difficulty collecting, processing, storing, filtering, and visualizing the data, we may refer to it as "Big Data." In a nutshell, "Big Data" technology entails amassing, storing, and mining massive data sets for insights. Huge data refers to a sizeable amount of information that is described by a more intricate form of information and intricate connections between various data sets. Big data's main advantage is that it analyses massive data more effectively than conventional methods. Because of advancements in data collection, storage, and interpretation, big data has captured the attention of the current generation. The proliferation of digital media over the last three decades has resulted in an explosion of data in fields as diverse as finance, social media, and medicine, among others. Every day, data storage prices fall, making it possible to save all of the data rather than destroy it. Numerous ways for analysing data have also been developed, although few of them have done so properly and effectively. The gathering of large resources that can be used frequently is what big data is like.

1.2 IoT

It is no longer a mysterious idea. IoT has recently developed into a technology that has the potential to influence how we live in the future. As naturally curious creatures, humans seek to reduce manual labour and the chance of error by using Internet-connected gadgets and simplify and streamline our lives overall (IoT). We, therefore, added intelligence to our technology and focused on other issues that would increase our effectiveness. We've made it feasible for devices to collect and share data using ML and NN by connecting them and the internet. The results of this phase were excellent. IoT users can benefit from higher degrees of automation, analysis, and integration. They improve the depth and scope of coverage in these areas. The Internet of Things is made up of sensors, networks, and robotics. Modern viewpoints on how technology affects our lives, more affordable hardware, and cutting-edge software are all benefits of IoT.

The success of IoT depends on the usage of artificial intelligence, sensors, user engagement, and tiny devices. We can quickly see what can be done by putting each of these features into practice in fig 1. IoT transforms seemingly lifeless devices into "smart" ones by collecting and analysing data using artificial intelligence algorithms and networks. Installing sensors in your refrigerator and cabinets that alert you when you have running low on milk or cereal might be a simple solution. Because of new enabling technologies, IoT connections are no longer dependent on only one or two service providers, which is a significant advancement for the networking sector. The IoT effectively creates mini-networks to link its numerous nodes. The Internet of Things would be nothing without sensors. These defining tools enable IoT to develop into a useful, flexible system. It is becoming less common for people to use technology passively. A fresh method of connecting with data, goods, and services are offered by the IoT. Electronics have naturally moved in the direction of becoming more

compact, affordable, and potent. IoT depends on tiny, custom-made devices for their accuracy, scalability, and adaptability.

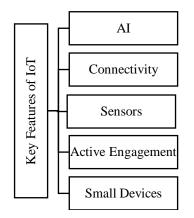


Figure 1. Key Features of IoT

1.3 Deep Learning

Deep learning architectures, prominent deep learning frameworks, and evaluation techniques for deep learning-based models were all explored in detail. The three main categories of learning models used in deep learning systems are supervised learning, unsupervised learning, and semi-supervised learning. With supervised learning, the data used to train the architecture is correctly labelled, but with unsupervised learning, the input is unlabelled and the architecture attempts to construct a structure by extracting important information. When it is challenging to extract key characteristics from the data, semi-supervised learning methods, which include labelled and unlabelled data in the training dataset, are useless. Furthermore, discriminative and generative deep learning systems can be distinguished. The generative model enables unsupervised learning techniques, whereas the discriminative model often supports supervised learning techniques.

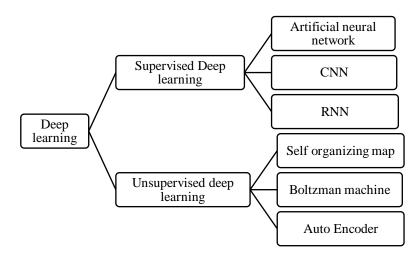


Figure 2. Deep Learning Techniques

1.4 Optimization

The goal of optimization is to provide the "optimal" design in light of a set of constraints or priorities. Some of them include increasing productivity, tenacity, reliability, endurance, effectiveness, and utilization. Optimization strategies are frequently used to find solutions that increase or decrease particular research criteria, such as decreasing costs involved with creating a good or service, maximizing revenues, minimizing the number of raw materials used to manufacture

a good or increasing output. Fig 1.3 is presenting different optimization techniques such as heuristic optimization, nature-inspired optimization, and Meta heuristic optimization.

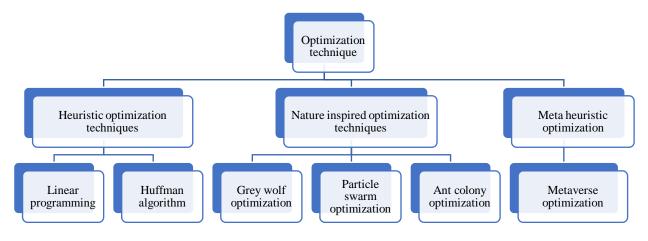


Figure 3. Optimization Technique

2. Literature Review

Recent studies in the fields of big data, security, and deep learning have taken datasets into account for both training and testing.

M. Mohammadi, et al. (2018) focused on big data and streaming analytics for IoT using DL. In this research, they explore how DL might be used to provide analytics and learning in the IoT domain. They also review and assess significant published research projects that used DL in the IoT area. Also addressed were the intelligent IoT gadgets that have DL built into their intelligence base. To support IoT applications, DL deployment strategies on fog and cloud centres are also studied. Finally, we discussed various difficulties and potential research avenues. [1]

D. P. Acharya, et al. (2019) reviewed analytics for big data: challenges, unanswered questions, and tools each day, cutting-edge IT infrastructures generate terabytes of data thanks to innovations like IoT & cloud computing. To generate information for decision-making from this kind of enormous data collection, a lot of work has to be put in at several levels of analysis. This has led to the present focus on research and development in the area of "big data analysis." The primary purpose of this research is to investigate the potential consequences of large data problems, unexplored research areas, and associated software. As a result, this article offers a framework for analysing big data in its many forms. Additionally, it provides researchers with a fresh opportunity to create a solution in light of the obstacles and unanswered questions that have plagued previous efforts. [2]

Dandugudum Mahesh, et al. (2019) introduced the future holds for big data analytics and AI in IoT. Today's most important developments are the IoT, Big Data, and AI. Several social structures, such as those of commerce, finance, industry, manufacturing, management, and the environment, stand to benefit from their implementation. IoT connects the real world to the virtual one and produces massive amounts of data. IoT, Big Data, and AI are all remarkable developments that, when put to use together, become much more crucial resources that may be utilized for good or evil. The utilization of IoT devices, large amounts of data, and artificial intelligence has posed several challenges to internet governance. [3]

Mr. K. Karthiban, et al. (2019) looked at stored & processed data as the major purpose of big data analytics in the creation of a safe and secure IoT. The phrase IoE describes the interconnection of smart devices, networks, systems, and people. In this analysis, we take a look at what's already in place to make the Internet of Things safe to use. Sensing devices embedded in everyday things contribute to the massive amounts of data known as "big data. This data can be applied to environmental analysis and development based on inference. This information is used by the Internet of Everything to automate the electronic devices in the immediate area. However, as automation levels rise, so does a system's susceptibility to attack. In this article, they take a close look at how big data analytics may be leveraged to build a secure IoE. [4]

M. Kantarcioglu, et al. (2019) provided research obstacles at the big data, security, and privacy intersection. Once data has been gathered, it may be shared between businesses and potentially linked or cleaned to enable 61 new applications and reveal potential value. For instance, municipal planners can use location information gathered from mobile devices 62 to better improve transportation systems. Sadly, security and privacy concerns could make such data exchange impossible. Even worse, in some circumstances, such data may be shared 64 with numerous parties that may have competing interests. For instance, some businesses might be reluctant to 65 shares their cybersecurity event data because of concern that a competitor might take advantage of it. Because of this, it's important to consider some issues, such as those of security, privacy, and 67 incentives for sharing big data. [5]

P. Angin, et al. (2019) presented analytics of big data for cybersecurity. With billions of connected devices, the Internet of Things era has increased the surface area that cybercriminals might exploit, necessitating the need for swift and precise threat detection. Cyber threat analysis systems may benefit greatly from the usage of big data analytics technology due to their ability to examine enormous volumes of data in real-time. [6]

Feng Zhang, et al. (2020) explained a security strategy for agriculture based on the IoT and big data mining. The swarm intelligence algorithm simulates the challenging issue of populations in nature by having individuals work together to solve it. The algorithm is independent of particular issues and offers excellent robustness and parallelism potential. Agricultural big data fusion problem concept, principle, and implementation approach are investigated. The large data fusion modelling algorithm's shortcomings are then examined. Finally, the ant colony large data fusion algorithm's source and fundamental phases are examined. The experimental findings demonstrate that the measured data validate the enhanced Big Data ACO algorithm. The enhanced algorithm suggested in this study significantly reduces the uncertainty of data fusion when compared to the D-S evidence theory, K-means, and Bayesian algorithm. [7]

S. Vats, et al. (2020) reviewed the evaluation of the efficiency of a time-optimized, symmetrical big data analytics framework. In the time-optimization paradigm presented in this study, three Namenodes (Master Nodes), three Data-nodes, & a single Client-node all shared HDFS. Within the DMZ, these vertices perform a mirroring role (DMZ). To use this approach, you'll need to hunt for machine learning jobs on a separate platform. To get started, the first node has to have the maven repository set up with Mahout and all of the other machine-learning libraries (Name-node 1). On the second node, R is being run through the shiny server and is linked to Hadoop (Name-node 2). The third node has Splunk set up so that the logs may be analysed (Name-node 3). For instance, processing Newsgroup data collection using Mahout since other tools are incompatible with it is an example of how tool choice may affect data independence. The empirical findings prove without a reasonable doubt that the suggested model outperforms the resource limitations of the baseline model. The proposed model is also capable of running any kind of algorithm on various data sets that are available in their native forms. [8]

S. Riaz, et al. (2020) focused on privacy and security concerns around big data in the cloud at the current time and where the field is headed in terms of research. In this paper, they analyse and critique existing frameworks and architectures for data security that is continually established against threats to improve how to keep and store data in the cloud setting, and we provide an overview of the characteristics and current state of big data and data security and privacy top threats, open issues, current challenges, and their impact on business for future research perspective. [9]

S. A. Janab, et al. (2020) presented IDA and IoTs are used in a clever method to provide the ideal learning environment in higher education. Five components make up the created system given here that uses IDA and the Internet of Things to solve an issue in higher education (IoTs). The biggest advantage of this approach is that it frees up a lot of time for teachers by eliminating the need to manually record attendance each day. Fewer people are required for the administration now that it is more exact (This will (iii) improve teachers' productivity by automatically starting and stopping class sessions when students arrive and leave the classroom. This consolidated system is seen as fundamental to the success of Iraq's planned e-government platform. If the violator's name has already been accepted to the institution, the judgment against his or her pupils should be applied automatically, barring them from attending classes while they are under punishment. The system is

affordable and efficient. Last but not least, the system is built on a 254-node private network and is protected by a pre-prepared encryption method. [10]

J. Koo, et al. (2020) introduced big data privacy and security is examined throughout the data's life cycle, and new issues are uncovered. This research verifies the existing standards developed by global standardization bodies and analyses relevant studies to identify risks and security concerns that arise throughout the big data life cycle. [11]

A.singh, et al. (2020) provided the processing of massive volumes of data for multimedia networks, using software that takes advantage of IoT. These wireless connections are made possible by technologies like Wi-Fi, Bluetooth, Infrared, and Hotspots. Devices like this must be connected to servers so that user requests may be fulfilled. Big data, also known as structured, semi-structured, or unstructured data, is produced in vast quantities by these sensing devices. Big data approaches are used to store, alter, and analyse the data to make well-defined judgments. As a result, corporate leaders at the top may manage their companies in real-time. The Personal, Group, Community, and Industrial application categories of the Internet of Things are all seeing a tremendous surge in the use of smart devices. Smart devices are a crucial part of IoT due to their easy connectivity to the internet, independence from a human power source, and ability to sense without it. This chapter begins with a quick overview of the IoT and its structure. The topic of IoT-related technology is then discussed. Also mentioned are the many IoT application fields. [12]

M. I. Tariq, et al. (2020) presented deep learning methods for improving large data in medicine Muhammad. Similar to this, there are several limitations on a significant amount of data processing with current conventional methods of data analysis and processing. Big data analytics and DL are both currently very active fields of medical science research. Due to the vast volumes of data relevant to its field that have been gathered by numerous medical organizations, big data is becoming more and more important. Big data management skills allow for the execution of novel research projects and the adoption of fresh delivery strategies for medical science. Therefore, unique contexts like medical and big data machine learning landscapes heavily rely on the existing DL algorithms utilized in general scenarios. Another contributor to poor performance is the heavy reliance on human input throughout the DL-based algorithm development process for medical big data. [13]

X. Nie, et al. (2020) looked at big data & IoT analytics-based safety management for subsea operations. This study explores how SCADA may be used in conjunction with IoT and Big Data Analytics to promote sustainable water management in smart cities. Sensors connected to the Internet of Things may be set up to capture data on how often a gadget is used, how well it performs, and how high quality it is made to be. This data can then be analyzed using big data techniques. To put this notion of big data analysis into practice, it may be possible to expand IoT software to include the whole water supply system and device product usage. The results of the study suggest that the implementation will actively regulate water usage by enterprises and consumers to improve the reliability of the water supply. [14]

V. Seethalakshmi1, et al. (2020) utilized the HGDSMO algorithm, big data processing in heterogeneous environments can schedule resources effectively. Addressing the problems and difficulties in the HGDSMO technique is suggested for effective resource allocation. The proposed HGDSMO algorithm takes inspiration from the foraging and social behaviour of spider monkeys to attain the same aim of effective resource allocation. [15]

Daissaoui, et al. (2020) researched a survey on IoT & big data analytics for smart buildings. The goal of the digital transformation procedures has been to increase productivity, safety, and execution quality, as well as to promote sustainable development, teamwork, and solutions for the sustainable smart city mainstream digital advancements are revealing new tendencies in the integration of information technology into the construction sector, which is undergoing radical change as a result. Systems for managing smart buildings today use a range of sensors, actuators, and dedicated networks. Their goals were to assess the state of particular places and implement the necessary regulations to maintain or enhance comfort while conserving energy. In this article, they suggest reviewing works on IoT and big data analytics for smart buildings. [16]

Khalid Haseeb, et al. (2020) provided EBDS: A Trustworthy IoT-based Big Data-based Safe Framework for the Environment. The Internet is used by the IoT-based network to interact with and store massive amounts of data on the cloud. Therefore, security and data integrity are additional study concerns along with the reduction in energy use. To this end, they provide a safe, big data-

based, energy-saving infrastructure for IoT. First, IoT-based sensors are linked together to collect data, and the Dijkstra-based optimum algorithm is used for data routing. This proposed method determines the most reliable and shortest transmission channels while using the least amount of energy possible. Protecting the produced big data from network attackers and encouraging green practices are additional benefits. According to the results of the experiments, the EBDS framework improves the green environment's performance by 16% in terms of energy consumption, 33% in terms of packet drop ratio, 13% in terms of network throughput, 15.5% in terms of end-to-end delay, and 16% in terms of route stability, compared to the baseline performance established by the previous studies. [17]

R Anitha, et al. (2020) Used deep learning techniques for data security in an IoT environment. Recently, IoT frameworks have struggled with signal validation, and it might be difficult to spot a digital attack when it connects to the cloud. The method utilized to identify the digital intrusion in the IoT entrance is deep learning. A new calculation is used to overcome this problem and ensure that the information is transferred securely. An autonomous vehicle is equipped with a variety of sensors. Information is not transferred because there is a possibility of an attack. The main concept is to use computations to ensure that information from IoT devices is sent to the cloud. [18]

Abhay Narayan Tripathi, et al. (2020) introduced deep learning at depth for big data and its applications. They have practical patterns that were previously impractical. Self-driving cars, visual recognition, healthcare, transportation, and other industries use deep learning. Companies nowadays are beginning to understand how critical it is to have access to big volumes of data to make the best decisions and advance their plans. Big data, so-called because of its massive quantity, varied nature, and rapid processing speed, is used to study patterns and trends. With its history, evolution, and introduction to some of the most complex neural networks, including DBN, CNN, and RNN, this paper serves as an introductory lesson to the field of deep learning for Big Data (RNN). [19]

M. A. Amanullah, et al. (2020) looked into past, studies that have demonstrated that the ML technique known as deep learning is effective in identifying vulnerabilities in large data-based IoT systems. The information gathered from IoT devices is extensive, varied, and precise. Therefore, using big data technology allows for improved speed and data processing. [20]

P. B. Dash, et al. (2020) used multi-class adaptive boosting classified, anomaly detection in IoT networks. Anomaly detection and attack identification are two of the most pressing issues in the IoT space today. Threats and abnormalities are being amplified to a sufficient degree as IoT-based infrastructure is being used exponentially across all industries. Attacks such as eavesdropping, service denial, malicious behaviours, etc. are the primary cause of an IoT system failure. By integrating many models, ensemble learning boosts the efficiency of ML strategies. In comparison to using a single technique, ensemble learning-based models provide a higher level of predictability when working with high-dimensional data this research proposes an adaptive boosting-based approach to detect anomalies in IoT environments. The suggested method outperforms numerous existing machine learning-based competitors on all the measures assessed in the performance comparison. [21]

M. A. Amanullah, et al. (2020) reviewed systems for protecting IoT devices using big data and machine learning algorithms. IoT has made it feasible to connect and interact with a myriad of objects, therefore integrating technology into almost every aspect of human existence. IoT has had security flaws uncovered, though. Therefore, it is crucial to construct solid answers by creating new technologies or merging current ones to handle security concerns. With this in mind, the identification of security breaches is one area of machine learning known as "DL" that has shown potential. Countless bits of information, of varying quality, are produced by IoT gadgets. Therefore, the use of big data technology may improve productivity and precision. That's why they probed state-of-the-art choices in areas like deep learning and massive data protection for the Internet of Things. Data security for IoT devices and big data analytics both have been studied in connection to deep learning. They were able to establish shared ground and create a hierarchical classification of related issues by comparing technical studies in these three disciplines. At last, we've taken a close look at the challenges posed by combining deep learning and big data technologies for IoT security. They have also paved the way for future academics by outlining a methodology for addressing IoT security issues. [22]

Sagu, et al. (2020) explained techniques for machine learning to secure IoT environments. IoT (Internet of Things) is spreading throughout society and making otherwise dumb objects smarter by allowing them to share data through networks. IoT is present in a variety of industries and is not just found in homes or utilities. The number of linked devices is expected to surpass 20 billion by the year 2024. It's not without its advantages and disadvantages. The safety of the networks upon which a plethora of devices depend is a major cause for worry. The focus of the current article was on all the concerns related to protecting IoT environments and how machine learning approaches could be able to aid with these security concerns. The research also examines the parameters, strategies, and potential approaches, as well as which strategy would be more efficient. [23]

Amit Sagu, et al. (2020) used Security for IoT using ML methods. One of the most exciting and invigorating breakthroughs in IT is the advent of IoT. IoT refers to a system in which electronic devices and other computational assets are networked together to exchange and transmit data. Things security is not as high as promised despite the rapid global expansion of IoT devices. Due to the pervasiveness of the IoT environment, the majority of users lack the knowledge or motivation to secure devices independently. Using machine intelligence to fix IoT security flaws might be a game changer. The researcher has employed machine learning techniques, methodologies, or methods in recent relevant studies to secure objects in an IoT setting. This article aims to examine the relevant literature on machine-learning strategies for safe IoT devices. [24]

Taran Singh Bharati, et al. (2020) provided big data challenges, issues, security, and privacy big data can be used to describe extraordinarily large amounts of information regarding things like jeans, cancer, pharmaceuticals, HIV, social networking sites, etc. Humans are attempting to decipher biological data to solve puzzles involving biological systems. People from other fields are drawn to big data. Big Data has more uses and applications than before, and it is gaining popularity in the biological streams of data scientists. Big data comes in massive quantities and is generated rapidly from many sources. Thousands of posts are created on social media every second. Some important concerns that are taken into account in this study are its nature, sharing, storage management, security, and privacy. The problems are carefully examined and debated. [25]

C. Li, et al. (2020) introduced the concept of "smart agriculture" which is shaped by the use of big data and the IoT in the planning process. Because of the proliferation of IoT devices and the advent of the age of big data in agriculture, a smart agricultural design based on IoT technology can effectively realize the function of real-time data transfer and information processing, therefore fostering the growth of smart agriculture. The analysis and processing of a vast number of planting and environmental data present a significant difficulty in the form of figuring out how to extract relevant information from these massive volumes of data. This piece uses the maximum distancebased k-means algorithm for data mining as a technique for investigating and optimizing the vast volumes of data generated during the agricultural production process. Agriculturalists' requirements drive this practice. An experimental study compares the performance of the modified K-means algorithm to that of the original K-means algorithm by simulating the crop growth curve. The experimental data reveal that the improved K-means clustering technique has an average improvement in the F metric value of 7.67% and a decrease in the overall time of 0.23 seconds. The algorithm presented in this article is vital in boosting agricultural modernization and launching the sector by better realizing the functions of real-time data interchange and information processing. [26] Yi. Chen, et al. (2020) reviewed that Improvements in cold chain logistics may be made via the use of cloud computing and the analysis of massive data sets. This research endeavours to examine cloud-based big data analysis for its potential in optimizing distribution in smart cold chain logistics. The outcomes of the experiment demonstrate the viability of the vehicle routing strategy for cold chain logistics optimization. Execution times are 19.89, 14.52, 8.12, and 6.41 correspondingly for 1, 2, 4, and 8 CPUs. The calculation time decreases with the number of processors. [27]

I.M. E. Hasnony, et al. (2021) used fog and mist may help analyse large amounts of data in the IoT. This article provides a structured approach for analysing the IoT ecosystem using big data analytics, along with its limitations and challenges. In addition, both centralized and decentralized data mining approaches are detailed for how cloud, fog, and mist might be used together to manage data from the Internet of Things. However, naive Bayes topped them for the time measure, using the smallest amount of time to generate the mode. [28]

W. Ding, et al. (2021) explained in a large-data industrial setting, they have enhanced the effectiveness of a chemical process that employs fuzzy and real-coded logic for intrusion detection. To monitor data flow and stop malicious actions in the Big Data setting, each communication network must analyse and model the intrusion detection system. Due to the nature of the machine learning model used by IDSs, it is necessary to examine vast amounts of network data, some of which may include redundant information that increases the processing and analytical effort needed to analyse the data. [29]

J. Chun-Wei Lin, et al. (2021) focused on producing a Multi-Objective Sanitization Model with Privacy Protection for the 6G IoT. To protect personal information, this article explains how to use the ACO method of ACO by using transaction deletion and adopting some different goals. To protect 6G IoT networks, this strategy is necessary. Delete transactions, one for each ant in the population, may be used to obfuscate up to eighteen pieces of information. They use a pre-large concept to decrease the amount of time spent scanning databases throughout the assessment process. To keep the 21 Pareto solutions we've already identified going, while also welcoming suggestions from the outside. This boosts efficiency and aids in identifying 22 optimal outcomes. PSO and GA, two additional cutting-edge bio-inspired algorithms, are compared to our technique in a series of experiments. In addition to producing fewer unintended consequences, our 25 findings show that the planned technique also has a relatively low computing cost. [30]

Citation	IoT Environment	Deep Learning	Big Data	Security	Optimization
[1]	Yes	Yes	No	No	No
[2]	No	No	Yes	No	No
[3]	Yes	No	Yes	No	No
[4]	Yes	No	Yes	Yes	No
[5]	No	No	Yes	Yes	No
[6]	No	No	Yes	Yes	No
[7]	No	No	Yes	No	No
[8]	No	No	Yes	No	Yes
[9]	No	No	Yes	Yes	No
[10]	Yes	No	No	No	Yes
[11]	No	No	Yes	Yes	No
[12]	Yes	No	Yes	No	No
[13]	No	Yes	No	No	Yes
[14]	Yes	No	Yes	No	No
[15]	No	No	Yes	No	No
[16]	Yes	No	Yes	No	No
[17]	Yes	No	Yes	Yes	No
[18]	Yes	Yes	No	Yes	No
[19]	No	Yes	Yes	No	No
[20]	No	Yes	Yes	Yes	No
[21]	Yes	No	No	No	No

Table 1: Comparison of feature chart

International Journal of Communication Networks and Information Security

[22]	No	Yes	Yes	No	No
[23]	Yes	No	No	Yes	No
[24]	No	No	Yes	Yes	No
[25]	No	No	Yes	Yes	No
[26]	Yes	No	Yes	No	No
[27]	No	No	Yes	No	Yes
[28]	Yes	No	Yes	No	No
[29]	No	No	Yes	No	Yes
[30]	Yes	No	No	No	No

3. Problem Statement

There have been several types of research in the area of big data security in IoT environments. But it has been observed that the security of big data is a challenging operation. Deep learning mechanisms that are used for classification need to be used for the categorization of attack. Moreover, an optimization mechanism is also required to get the best solution so that training and testing operations made by the deep learning approach should provide better accuracy along with better performance. Filtering a dataset on the bases of optimized value would result in high accuracy and performance.

4. Proposed Work

The topic of how to best secure vast amounts of data in an IoT setting has been the subject of several academic investigations. Nonetheless, it is a known challenge that protects massive data sets from unauthorized access. The classification of attacks has to be done with the use of deep learning mechanisms, which are often employed for other purposes. Additionally, an optimization mechanism is needed to get the optimal answer. This is necessary for the training and testing operations carried out by the deep learning technique to deliver improved accuracy in addition to enhanced performance. A high level of accuracy and performance might be achieved by filtering the dataset based on the grounds of optimum value.

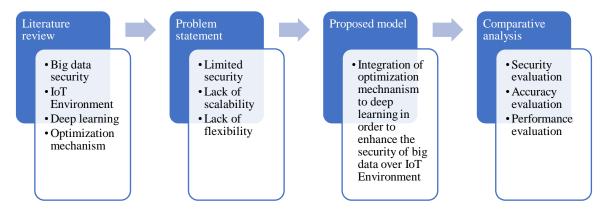


Figure 4. Proposed research methodology

The proposed model has considered records of big data transactions over an IoT environment. An optimization mechanism has been applied to get the best solution. Considering the best solution the dataset is filtered.

Security of Big Data over IoT Environment by Integration of Deep Learning and Optimization

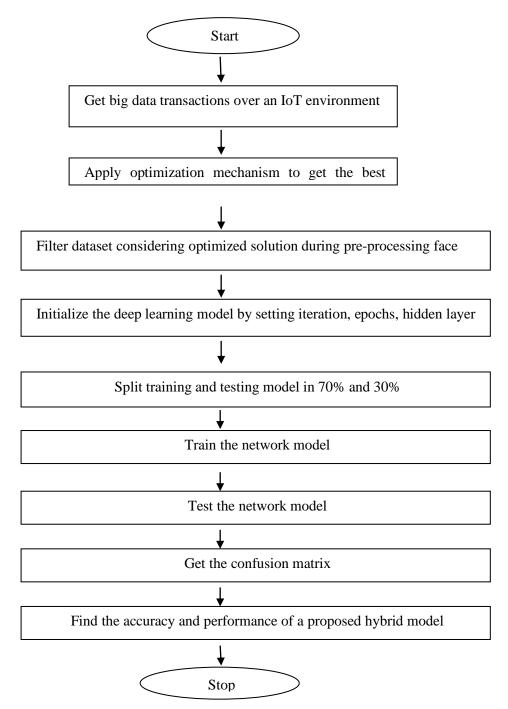


Figure 5. Flow chart of proposed work

Read datasets from big data
Initialize parameter and objective function of optimization
Set lower bond, upper bond, iteration
Get the best optimum solution
Filter dataset considering guest
Set hidden layer, batch size, epochs
Perform partition for training and testing for 70% and 30%
Train network
Test Network
Find true positive (TP), true negative (TN), false positive (FP), false negative (FN)

^{4.2} Algorithm for Proposed Model

Find accuracy
Accuracy = (TN + TP) / (TN + TP + FN + FP)
Precision = TP / (TP + FP)
Recall= TP / (TP + FN)
F1 score= 2* (Precision * Recall)/ (Precision + Recall)

5. Results and Discussion

A deep model is used for a dataset containing a variety of big data entries over an IoT environment for the proposed research. Because of this, many forms of cyberattacks may be recognized and labelled. Two situations have been simulated: one in which the dataset is filtered, and the other in which it is not. The use of an optimizer to the dataset to filter it ought to result in enhanced detection and classification outcomes. The present work has considered 1000 epochs, 32 batch sizes, the PSO optimization technique, and ANN as a classification technique.

5.1 Confusion Matrix in the case of Conventional Model

This section is presenting accuracy in the case of the conventional model. Table 2 is presenting the confusion matrix obtained after testing the conventional model while table 3 is presenting the accuracy table produced on the bases of table 2.

	Brute force attack	Man in middle	Denial of services	SQL injection
Brute force attack	879	34	12	75
Man in middle	65	822	45	68
Denial of services	182	12	783	23
SQL injection	45	22	22	911

Table 2. Confusion matrix of a conventional classification model

Results TP: 3395

Overall Accuracy: 84.88%

After applying recall, precision, accuracy, and F1-score in table 2, accuracy parameters are extracted and presented in Table 3

Table 3. Accuracy of	Confusion matrix	of the convention	onal model

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
1	1171	1000	89.68%	0.88	0.75	0.81
2	890	1000	93.85%	0.82	0.92	0.87
3	862	1000	92.6%	0.78	0.91	0.84
4	1077	1000	93.63%	0.91	0.85	0.88

5.2 Confusion Matrix of Filtered Dataset

This section is presenting accuracy in the case of the conventional model. Table 4 is presenting the confusion matrix obtained after testing the conventional model while table 5 is presenting the accuracy table produced on the bases of table 4. Table 4 is considering the Confusion matrix of the filtered dataset.

		Brute attack	force	Man middle	in	Denial services	of	SQL injection
Brute attack	force	956		18		23		3
Man in m	niddle	55		912		23		10
Denial of	services	109		12		856		23
SQL inje	ction	33		22		22		923

Table 4. Confusion matrix of the proposed model

Results

TP: 3647

Overall Accuracy: 91.18%

After applying precision, accuracy, recall, and F1-score in table 4, accuracy parameters are extracted and presented in Table 5

Table 5. Accuracy	of C	Confusion	matrix o	of the	proposed model
	~ , ~			.,	p · · · p · · · · · · · · · · · · · · ·

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
1	1153	1000	93.98%	0.96	0.83	0.89
2	964	1000	96.5%	0.91	0.95	0.93
3	924	1000	94.7%	0.86	0.93	0.89
4	959	1000	97.18%	0.92	0.96	0.94

5.3 Comparative Analysis

5.3.1 Accuracy

Table 6 displays the results of checking the accuracy of prior work and planned work for each of classes 1, 2, 3, and 4. It has been found that the suggested work is accurate in comparison to the standard model.

Table 6. Comparison Analysis of Accuracy

Conventional model	Proposed model
89.68%	93.98%
93.85%	96.5%
92.6%	94.7%
93.63%	97.18%
	89.68% 93.85% 92.6%

Considering table 6 fig 6 is drawn to visualize the accuracy of the proposed model concerning the conventional model.

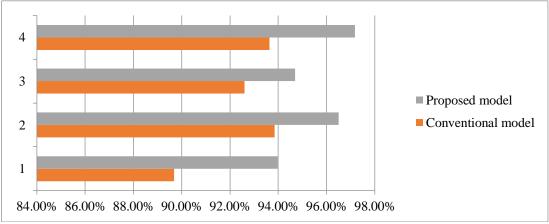


Figure 6. Comparison Analysis of Accuracy

5.3.2 Precision

Table 7 displays the results of comparing the accuracy of past and projected work for each of the three classes. It has been noted that the suggested model is more precise than the standard model.

Class	Conventional model	Proposed model
1	0.88	0.96
2	0.82	0.91
3	0.78	0.86
4	0.91	0.92

Table 7. Comparison Analysis of Precision

Considering table 7 fig. 7 is drawn to visualize the precision of the proposed model for the conventional model.

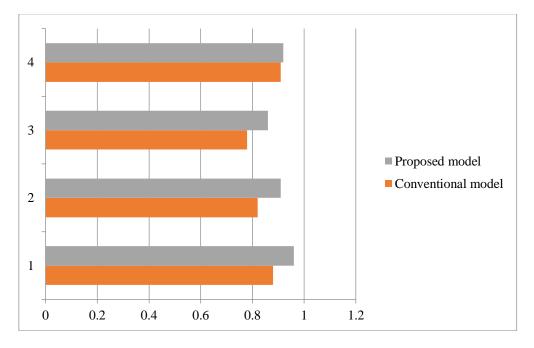


Figure 7 Comparison Analysis of Precision

5.3.3 Recall Value

Table 8 displays the recall values from prior work and planned work for classes 1, 2, and 3. Comparing the suggested model to the standard model, it is shown that the Recall value is higher.

Table 8. Comparison Analysis of Recall Value

Class	Conventional model	Proposed model
1	0.75	0.83
2	0.92	0.95
3	0.91	0.93
4	0.85	0.96

Considering table 8, figure 8 is drawn to visualize the recall value of the proposed model concerning the conventional model.

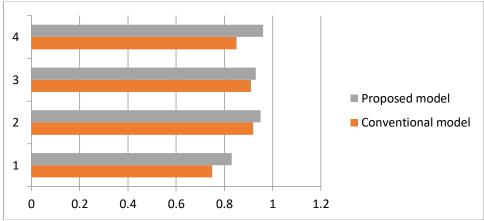
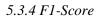


Figure 8. Comparison Analysis of Recall Value



F1-Score of previous work & proposed work are taken for class 1, class2, and class 3 and shown in table 9. It is observed that the F1-Score proposed concerning conventional.

Class	Conventional model	Proposed model
1	0.81	0.89
2	0.87	0.93
3	0.84	0.89
4	0.88	0.94

Table 9. Comparison Analysis of F1-Score

Considering table 9, figure 9 is drawn to visualize the F1-Score of the proposed model concerning the conventional model.

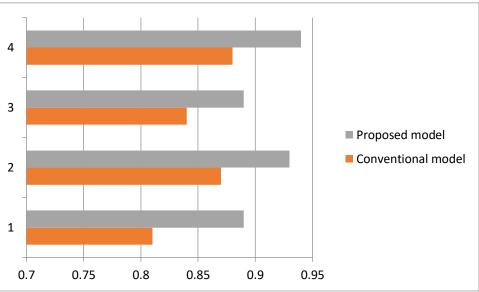


Figure 9. Comparison Analysis of F1-Score

6. Conclusion

It is possible to conclude, on basis of the results of simulations, that the work that has been provided has delivered a higher level of accuracy when compared to past techniques that made use of a deep learning methodology. It is feasible that the work that has been recommended will increase accuracy while simultaneously decreasing the amount of time that is spent thanks to the incorporation of an optimization mechanism. In addition, the classification strategy that was used while classifying attacks has led to improvements in terms of recall value, accuracy, precision, and F1 Score as a consequence of the work that was done.

7. Future Scope

Present research would play a significant role in securing big data in an IoT environment. Such a mechanism would allow the classification of different types of security threats in the IoT environment. The optimization mechanism might be enhanced to get the best solution. Moreover, hidden layers used in deep learning are capable to improve further accuracy.

References

- [1] Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. IEEE Communications Surveys and Tutorials, 20(4), 2923–2960. https://doi.org/10.1109/COMST.2018.2844341
- [2] Muthulakshmi, P., & Udhayapriya, S. (2018). a Survey on Big Data Issues and Challenges. International Journal of Computer Sciences and Engineering, 6(6), 1238–1244. https://doi.org/10.26438/ijcse/v6i6.12381244
- [3] Mahesh, D., Yamani, N., Harshavardhan, A., Srikanth, Y., Engineering, C., & Urban, W. (2019). New Challenges and Future Trends: Big Data Analytics and Artificial Intelligence in IoT. Studia Rosenthaliana (Journal for the Study of Research), XI(165), 165–172.
- [4] K., K., & S. Raj, J. (2019). Big Data Analytics for Developing Secure Internet of Everything. Journal of ISMAC, 01(02), 49–56. https://doi.org/10.36548/jismac.2019.2.006
- [5] Kantarcioglu, M., & Ferrari, E. (2019). Research Challenges at the Intersection of Big Data, Security, and Privacy. Frontiers in Big Data, 2. https://doi.org/10.3389/fdata.2019.00001
- [6] Angin, P., Bhargava, B., & Ranchal, R. (2019). Big Data Analytics for Cyber Security. Security and Communication Networks, 2019. https://doi.org/10.1155/2019/4109836
- [7] Zhang, F., & Zhang, Y. (2020). A big data mining and blockchain-enabled security approach for agriculture based on the internet of things. Wireless Communications and Mobile Computing, 2020. https://doi.org/10.1155/2020/6612972
- [8] Vats, S., Sagar, B. B., Singh, K., Ahmadian, A., & Pansera, B. A. (2020). Performance evaluation of an independent time-optimized infrastructure for big data analytics that maintains symmetry. Symmetry, 12(8), 1–15. https://doi.org/10.3390/SYM12081274
- [9] Riaz, S., Khan, A. H., Haroon, M., Latif, S., & Bhatti, S. (2020). Big data security and privacy: Current challenges and future research perspective in the cloud environment. Proceedings of 2020 International Conference on Information Management and Technology, ICIMTech 2020, August, 977–982. https://doi.org/10.1109/ICIMTech50083.2020.9211239
- [10] Al_Janabi, S. (2020). Smart system to create an optimal higher education environment using IDA and IOTs. International Journal of Computers and Applications, 42(3), 244–259. https://doi.org/10.1080/1206212X.2018.1512460
- [11] Koo, J., Kang, G., & Kim, Y. G. (2020). Security and privacy in the big data life cycle: A survey and open challenges. Sustainability (Switzerland), 12(24), 1–32. https://doi.org/10.3390/su122410571
- [12] Singh, A., & Mahapatra, S. (2020). Network-based applications of multimedia big data computing in IoT environment. In Intelligent Systems Reference Library (Vol. 163). Springer Singapore. https://doi.org/10.1007/978-981-13-8759-3_17
- [13] Tariq, M. I., Tayyaba, S., Ashraf, M. W., & Balas, V. E. (2020). Deep learning techniques for optimizing medical big data. In Deep Learning Techniques for Biomedical and Health Informatics. Elsevier Inc. https://doi.org/10.1016/B978-0-12-819061-6.00008-2
- [14] Nie, X., Fan, T., Wang, B., Li, Z., Shankar, A., & Manickam, A. (2020). Big Data analytics and IoT in Operation safety management in Under Water Management. Computer Communications, 154, 188–196. https://doi.org/10.1016/j.comcom.2020.02.052
- [15] Seethalakshmi, V., Govindasamy, V., & Akila, V. (2020). Hybrid gradient descent spider monkey optimization (HGDSMO) algorithm for efficient resource scheduling for big data

processing in a heterogeneous environment. Journal of Big Data, 7(1). https://doi.org/10.1186/s40537-020-00321-w

- [16] Daissaoui, A., Boulmakoul, A., Karim, L., & Lbath, A. (2020). IoT and Big Data Analytics for Smart Buildings: A Survey. Procedia Computer Science, 170, 161–168. https://doi.org/10.1016/j.procs.2020.03.021
- [17] Haseeb, K., Lee, S., & Jeon, G. (2020). EBDS: An energy-efficient big data-based secure framework using the Internet of Things for a green environment. Environmental Technology and Innovation, 20, 101129. https://doi.org/10.1016/j.eti.2020.101129
- [18] Anitha, R., & Raja, A. B. (2020). Data Security in IoT Environment using Deep Learning Technique. International Journal of Creative Research Thoughts, 8(6), 2320–2882. www.ijcrt.org
- [19] Narayan Tripathi, A., & Sharma Assistant Professor, B. (2020). A Depth of Deep Learning for Big Data and its Applications. 8(10), 20–23. www.ijert.org
- [20] Amanullah, M. A., Habeeb, R. A. A., Nasaruddin, F. H., Gani, A., Ahmed, E., Nainar, A. S. M., Akim, N. M., & Imran, M. (2020). Deep learning and big data technologies for IoT security. Computer Communications, 151, 495–517. https://doi.org/10.1016/j.comcom.2020.01.016
- [21] Dash, P. B., & Rao, K. S. (2020). Anomaly Detection in IoT Network by using Multi-class Adaptive Boosting Classifier. 9(3), 164–171.
- [22] Amanullah, M. A., Habeeb, R. A. A., Nasaruddin, F. H., Gani, A., Ahmed, E., Nainar, A. S. M., Akim, N. M., & Imran, M. (2020). Deep learning and big data technologies for IoT security. Computer Communications, 151(January), 495–517. https://doi.org/10.1016/j.comcom.2020.01.016
- [23] Sagu*, A., & Gill, N. S. (2020). Machine Learning Techniques for Securing IoT Environment. International Journal of Innovative Technology and Exploring Engineering, 9(4), 977–982. https://doi.org/10.35940/ijitee.d1209.029420
- [24] Sagu, A., & Gill, N. S. (2020). Securing IoT Environment using Machine Learning Techniques. International Journal of Engineering and Advanced Technology, 9(3), 870–873. https://doi.org/10.35940/ijeat.c5339.029320
- [25] Bharati, T. S. (2020). Challenges, issues, security, and privacy of big data. International Journal of Scientific and Technology Research, 9(2), 1482–1486.
- [26] Li, C., & Niu, B. (2020). Design of smart agriculture based on big data and the Internet of things. International Journal of Distributed Sensor Networks, 16(5). https://doi.org/10.1177/1550147720917065
- [27] Chen, Y. Hua. (2020). Intelligent algorithms for cold chain logistics distribution optimization based on big data cloud computing analysis. Journal of Cloud Computing, 9(1). https://doi.org/10.1186/s13677-020-00174-x
- [28] El-Hasnony, I. M., Mostafa, R. R., Elhoseny, M., & Barakat, S. I. (2021). Leveraging mist and fog for big data analytics in IoT environment. Transactions on Emerging Telecommunications Technologies, 32(7), 1–16. https://doi.org/10.1002/ett.4057
- [29] Ding, W., Nayak, J., Naik, B., Pelusi, D., & Mishra, M. (2021). Fuzzy and Real-Coded Chemical Reaction Optimization for Intrusion Detection in Industrial Big Data Environment. IEEE Transactions on Industrial Informatics, 17(6), 4298–4307. https://doi.org/10.1109/TII.2020.3007419
- [30] Lin, J. C. W., Srivastava, G., Zhang, Y., Djenouri, Y., & Aloqaily, M. (2021). Privacypreserving multiobjective sanitization model in 6G IoT environments. IEEE Internet of Things Journal, 8(7), 5340–5349. https://doi.org/10.1109/JIOT.2020.3032896
- [31] Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2021). IoT-based big data security management in the fog over a 6G wireless network. IEEE Internet of Things Journal, 8(7), 5164–5171. https://doi.org/10.1109/JIOT.2020.3033131
- [32] Duraisamy, A., Subramaniam, M., & Robin, C. R. R. (2021). An Optimized Deep Learning Based Security Enhancement and Attack Detection on IoT Using IDS and KH-AES for Smart Cities. Studies in Informatics and Control, 30(2), 121–131. https://doi.org/10.24846/v30i2y202111

- [33] Islam, N., Altamimi, M., Haseeb, K., & Siraj, M. (2021). Secure and sustainable predictive framework for IoT-based multimedia services using machine learning. Sustainability (Switzerland), 13(23), 1–15. https://doi.org/10.3390/su132313128
- [34] Et. al., L. J. M. (2021). Enabling Intelligence through Deep Learning using IoT in a Classroom Environment based on a multimodal approach. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(2), 381–393. https://doi.org/10.17762/turcomat.v12i2.818
- [35] Sanjay, V., & Suganthi, N. (2021). Deep Learning Techniques In Internet Of Things (IoT) Security. 12(03), 1453–1459.
- [36] Han, T., Muhammad, K., Hussain, T., Lloret, J., & Baik, S. W. (2021). An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks. IEEE Internet of Things Journal, 8(5), 3170–3179. https://doi.org/10.1109/JIOT.2020.3013306
- [37] Saleem, T. J., & Chishti, M. A. (2021). Deep learning for the internet of things: Potential benefits and use-cases. Digital Communications and Networks, 7(4), 526–542. https://doi.org/10.1016/j.dcan.2020.12.002
- [38] Aversano, L., Bernardi, M. L., Cimitile, M., & Pecori, R. (2021). A systematic review on Deep Learning approaches for IoT security. Computer Science Review, 40, 100389. https://doi.org/10.1016/j.cosrev.2021.100389
- [39] Lee, N. (2021). Consumer Privacy in the Age of Big Data. Facebook Nation, 22(2), 157– 171. https://doi.org/10.1007/978-1-0716-1867-7_8
- [40] Goel, P., Patel, R., Garg, D., & Ganatra, A. (2021). A review on big data: Privacy and security challenges. 2021 3rd International Conference on Signal Processing and Communication, ICPSC 2021, 8(5), 705–709. https://doi.org/10.1109/ICSPC51351.2021.9451749
- [41] Sun, L., Zhang, H., & Fang, C. (2021). Data security governance in the era of big data: status, challenges, and prospects. Data Science and Management, 2(June), 41–44. https://doi.org/10.1016/j.dsm.2021.06.001
- [42] Lakshmanna, K., Kaluri, R., Gundluru, N., Alzamil, Z. S., Rajput, D. S., Khan, A. A., Haq, M. A., & Alhussen, A. (2022). A Review on Deep Learning Techniques for IoT Data. Electronics (Switzerland), 11(10), 1–23. https://doi.org/10.3390/electronics11101604
- [43] Bian, J., Arafat, A. Al, Xiong, H., Li, J., Li, L., Chen, H., Wang, J., Dou, D., & Guo, Z. (2022). Machine Learning in Real-Time Internet of Things (IoT) Systems: A Survey. IEEE Internet of Things Journal, 1–1. https://doi.org/10.1109/jiot.2022.3161050
- [44] Prof. Romi Morzelona. (2019). Histogram Based Data Cryptographic Technique with High Level Security. International Journal of New Practices in Management and Engineering, 8(04), 08 14.
- [45] A. Gupta, A.Verma, S. Pramanik, "Advanced Security System in Video Surveillance for COVID-19", in An Interdisciplinary Approach to Modern Network Security, S. Pramanik, A. Sharma, S. Bhatia and D. N. Le, CRC Press, 2022.
- [46] A. Gupta, A. Verma and S. Pramanik, Security Aspects in Advanced Image Processing Techniques for COVID-19, in An Interdisciplinary Approach to Modern Network Security, S. Pramanik, A. Sharma, S. Bhatia and D. N. Le, Eds, CRC Press, 2022.
- [47] K. Dushyant, G. Muskan, A. Gupta and S. Pramanik, "Utilizing Machine Learning and Deep Learning in Cyber security: An Innovative Approach", in Cyber security and Digital Forensics, M. M. Ghonge, S. Pramanik, R. Mangrulkar, D. N. Le, Eds, Wiley, 2022, https://doi.org/10.1002/9781119795667.ch12
- [48] A. Mandal, S. Dutta, S. Pramanik, "Machine Intelligence of Pi from Geometrical Figures with Variable Parameters using SCILab", in Methodologies and Applications of Computational Statistics for Machine Learning, D. Samanta, R. R. Althar, S. Pramanik and S. Dutta, Eds, IGI Global, 2021, pp. 38-63, DOI: 10.4018/978-1-7998-7701-1.ch003
- [49] Venu, S., Kotti, J., Pankajam, A., Dhabliya, D., Rao, G. N., Bansal, R., . . . Sammy, F. (2022). Secure big data processing in multihoming networks with AI-enabled IoT. Wireless Communications and Mobile Computing, 2022 doi:10.1155/2022/3893875
- [50] Garg, M & Gupta, A & Kaushik, D & Verma, A. (2020). Applying machine learning in IoT to build intelligent system for packet routing system, Materials Today: Proceedings. 10.1016/j.matpr.2020.09.539.

- [51] Aggarwal, B. & Gupta, A. & Goyal, D. & Gupta, P. & Bansal, B. & Barak, D. (2021), A review on investigating the role of block-chain in cyber security. Materials Today: Proceedings. 10.1016/j.matpr.2021.10.124.
- [52] Dhabliya, D. (2021). Feature Selection Intrusion Detection System for The Attack Classification with Data Summarization. Machine Learning Applications in Engineering Education and Management, 1(1), 20–25.
- [53] Kshirsagar, P. R., Yadav, R. K., Patil, N. N., & Makarand L, M. (2022). Intrusion Detection System Attack Detection and Classification Model with Feed-Forward LSTM Gate in Conventional Dataset. Machine Learning Applications in Engineering Education and Management, 2(1), 20–29.
- [54] Gupta, A. & Garg, M. & Verma, A. & Kaushik, D. (2020). Implementing lossless compression during image processing by integrated approach. Materials Today: Proceedings. 10.1016/j.matpr.2020.10.052.
- [55] Verma, A. & Gupta, A. & Kaushik, D. & Garg, M. (2021). Performance enhancement of IOT based accident detection system by integration of edge detection. Materials Today: Proceedings. 10.1016/j.matpr.2021.01.468.