Unraveling the Decision-making Process Interpretable Deep Learning IDS for Transportation Network Security

Rajit Nair

VIT Bhopal University, Bhopal, India
Email: rajit.nair@vitbhopal.ac.in

Abstract

The growing ubiquity of IoT-enabled devices in recent years emphasizes the critical need to strengthen transportation network safety and dependability. Intrusion detection systems (IDS) are crucial in preventing attacks on transport networks that rely on the Internet of Things (IoT). However, understanding the rationale behind deep learning-based IDS models may be challenging because they do not explain their findings. We offer an interpretable deep learning system that may be used to improve transportation network safety using IoT. To develop naturally accessible explanations for IDS projections, we integrate deep learning models with the Shapley Additive Reasons (SHAP) approach. By adding weight to distinct elements of the input data needed to develop the model, we increase the readability of so-called "black box" processes. We use the ToN_IoT dataset, which provides statistics on the volume of network traffic created by IoT-enabled transport systems, to assess the success of our strategy. We use a tool called CICFlowMeter to create network flows and collect data. The regularity of the flows, as well as their correlation with specific assaults, has been documented, allowing us to train and evaluate the IDS model. The experiment findings show that our explainable deep learning system is extremely accurate at detecting and categorising intrusions in IoT-enabled transportation networks. By examining data using the SHAP approach, cybersecurity specialists may learn more about the IDS's decision-making process. This enables the development of robust solutions, which improves the overall security of the Internet of Things. Aside from simplifying IDS predictions, the proposed technique provides useful recommendations for strengthening the resilience of IoT-enabled transportation systems against cyberattacks. The usefulness of IDS in defending mission critical IoT infrastructure has been questioned by security experts in the Internet of Vehicles (IoV) industry. The IoV is the primary research object in this case. Deep learning algorithms' versatility in processing many forms of data has contributed to their growing prominence in the field of anomaly detection in intrusion detection systems. Although machine learning models may be highly useful, they frequently yield false positives, and the path they follow to their conclusions is not always obvious to humans. Cybersecurity experts who want to evaluate the performance of a system or design more secure solutions need to understand the thinking process behind an IDS's results. The SHAP approach is employed in our proposed framework to give greater insight into the decisions made by IDSs that depend on deep learning. As a result, IoT network security is strengthened, and more cyber-resilient systems are developed. We demonstrate the effectiveness of our technique by comparing it to other credible methods and utilising the ToN_IoT dataset. Our framework has the best success rate when compared to other frameworks, as evidenced by testing results showing an F1 score of 98.83 percent and an accuracy of 99.15 percent. These findings demonstrate that the architecture successfully resists a variety of destructive assaults on IoT networks. By integrating deep learning and methodologies with an emphasis on explainability, our approach significantly enhances network security in IoT use scenarios. The ability to assess and grasp IDS options provides the path for cybersecurity experts to design and construct more secure IoT systems.

Keywords: Cyber Security; Deep Learning; Internet of Vehicles; Intrusion detection systems; Internet of Things; Shapley Additive Reasons

1. Introduction

The transportation industry is only one of several that has been greatly influenced by the fast growth of IoT technology. This is only one example. The Internet of Things has boosted our decision-making ability, as well as efficiency and safety, by connecting sensors, infrastructure, and automobiles. As the IoT becomes increasingly

Doi : https://doi.org/10.54216/JCIM.120205
Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

prevalent in the transportation industry, it is critical to develop effective security measures to protect IoT networks from fraudulent activities and ensure seamless operations. Firewalls and intrusion detection systems (IDS) have long been used to prevent hackers from breaking into secure networks. However, because of the increasing diversity and interconnection of IoT networks, traditional ways of managing cyberthreats frequently fall short. As criminal capabilities grow, existing security measures are increasingly unable to detect and fight sophisticated fraud schemes. To address the expanding dangers to public safety, new, more effective security solutions must be developed and implemented on a regular basis. These technologies are anticipated to identify and prevent even the most sophisticated attacks and breaches [1-2]. Deep learning, a revolutionary type of machine learning, has developed in recent years as a potentially viable strategy for studying massive datasets that are notoriously difficult to comprehend. As a result, it's an excellent solution to cope with the particular security concerns that arise with IoT networks. Deep learning models, particularly deep neural networks, have shown exceptional performance in a range of domains, including image recognition, natural language processing, and audio recognition. Because of their ability to automatically learn complicated patterns and properties from data, these models can identify abnormalities and classify various forms of intrusion. Deep learning algorithms are being developed to make this a reality. Deep learning models' decision-making methods can be opaque, making it difficult for humans to understand their reasoning despite their superior performance in identification tests. This lack of transparency is quite concerning, especially given that various systems place differing priorities on security. Examining the system's effectiveness and understanding the motives behind the models' conclusions is difficult for cybersecurity specialists [3-5]. Model interpretability is critical when they are employed for intrusion detection in transportation systems, where a security failure might have disastrous consequences. This study provides a simple deep-learning strategy for improving transport system security by identifying weak areas in the networks that connect IoT devices. The model's major purpose is to produce a solution that not only improves the model's detection accuracy but also gives insight into how the model arrived at its conclusions. Because of the model's transparency, system administrators have reviewed and trusted its forecasts. Our hypothesis is that when deep learning is combined with explainable techniques, it can improve the safety of transport networks without compromising the durability of IoT-enabled devices. This study's findings also provide an interpretable deep learning model for IoT network intrusion detection. This helps meet the crucial demand for greater safety standards throughout the transportation sector. By leveraging the strength of deep learning technologies and augmenting them with explain ability methods, the suggested paradigm has the potential to offer trustworthy and comprehensible intrusion detection. This might be accomplished by utilizing the capabilities of explainability mechanisms. In this paper, we will look at studies on intrusion detection for IoT networks. Furthermore, we will elaborate on the suggested model's architecture and explainability, offer experimental evaluations, dive into the model's implications for transportation security, and map out future research in this area [6]. Even though there is only a limited amount of bandwidth available, the expansion of IoT technology has resulted in an increase in the number of connections created by IoT devices in a variety of commercial settings. However, concerns regarding privacy and security have emerged as a key impediment to the widespread use of the IoT. Many IoT systems lack proper protections, rendering them vulnerable to cyberattacks and providing opportunities for criminals to exploit them. This might have a significant impact on the scenario's conclusion. Because of this weakness, critical infrastructure such as power plants, water treatment facilities, and smart cities are especially susceptible since an attack might result in significant deaths and property damage. Because IoT devices are being used in potentially fatal scenarios, more and more individuals are concerned about their security[7-10]. The IoV market has made great progress in introducing ICT to the transportation sector. Concerns regarding the current protocols' lack of transparency and cybermedicine continue, even though several studies have focused on using AI algorithms to detect suspicious behavior in IoV networks. Traditional security solutions, which were designed to safeguard traditional computer networks, frequently fall short of securing the safety-critical networks that power the Internet of Things and the Internet of Vehicles. As more consumer products connect to the Internet of Things and function on closed communication networks, security challenges become more difficult.

2. Related Work

Several initiatives have been established in recent years with the common goal of strengthening the cyber resilience of transport systems in the face of internet attacks. When a network or computer system is cyber-resilient, it can recover from disruptions or attacks quickly and autonomously without losing any data. Another aspect of cyber-resilience is a system's ability to change its security processes and configuration in the face of recognized threats. Intrusion detection can significantly improve an IoV network's cyber resilience. To protect the privacy, availability, and security of important IoT devices used in transportation, resilience against both cyber and physical attacks is required [11-13]. This need applies to a wide range of IoT use cases due to the necessity of timely reactions to avoid and recover from attacks on IoT infrastructure. Deep learning models are quickly gaining favor in the field of intrusion detection because they can be taught with huge volumes of data. However, due to their lack of transparency, these models are becoming increasingly difficult to implement in safety-critical Internet of

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

Things systems. Such systems can be found in applications such as medical monitoring and autonomous vehicle research. It is critical that learning systems provide explanations for the judgments they make to establish trust in the system and allow domain experts to assess its potential to perform tasks in a secure and safe manner. Deep learning-based intrusion detection systems that can explain their findings may shed light on the intrusion prediction challenge. Given the complexity and variety of transport networks enabled by the Internet of Things, this is critical. Security professionals may use this knowledge to improve the accuracy of intrusion detection systems by altering protocol settings and other configuration considerations.

To do this, we will provide a paradigm for artificial intelligence-based intrusion detection systems that may be deployed in tandem with the current communication infrastructure for the Internet of Things and the Internet of Vehicles [14]. The purpose of this method is to assist cybersecurity experts in making sense of the findings of artificial intelligence models employed in IoT intrusion detection systems. In addition, we provide our thoughts on the Deep SHAP technique, explaining why intrusion detection is important in IoT networks and how we feel it should be done. AI models used in computer vision and natural language processing have been described using SHAP approaches. Although prior studies have used SHAP techniques, this is the first research to use Deep SHAP for intrusion detection systems in computer networks. The proposed method was successfully evaluated using the ToN_IoT dataset and a convolutional neural network (CNN)-based intrusion detection system [15-17]. The experimental findings show that the suggested strategy outperforms competing alternatives and that it can be generalised. By merging deep learning with explainability-focused techniques, this solution increases network security for IoT use cases. This enables security experts to evaluate the efficiency of intrusion detection systems, make informed judgements about how to improve network security, and develop IoT solutions with lower failure rates. Explainable deep learning techniques, when applied to intrusion detection systems, not only increase the dependability of IoT networks but also shed light on how those systems get their results. As a result, customers have more confidence in the accuracy of the systems' conclusions.

Our research finishes with the development of a straightforward, deep-learning-based technique for improving the security of Internet of Things networks used in transportation systems. The proposed method improves the overall safety of transport networks and protects the longevity of IoT-enabled products. To do this, it is critical to put in place procedures for detecting violations and giving transparency to the decision-making process. As the Internet of Things evolves and spreads into new areas of the economy, ensuring the security and resilience of IoT networks becomes increasingly vital [18]. Integrating contemporary intrusion detection systems with explainability mechanisms may substantially improve the pressing need for greater safety measures in transportation systems. Any future research into intrusion detection in IoT networks should aim to improve the explainability of the deep learning models used for this purpose. Creating better and easier-to-understand explanations of the decision-making processes revealed by the models may entail the creation of new methodologies or the use of existing ones. It is also critical to focus on extending the proposed framework's applicability to the Internet of Things domains other than transportation, such as healthcare, smart homes, and industrial systems. This is important information. This critical treatment must be carried out right away. To make the framework adaptable enough to be used in a variety of situations, it must be tailored to handle the particular security concerns that each domain faces. Deep learning models are being used to identify intrusions in IoT networks; hence, research on approaches to increasing the efficacy of these models is needed. These models are currently being used to identify intrusions [19-21]. To handle the increasing volume and complexity of data created by IoT devices, we must improve the scalability and efficiency of our models. More research on approaches for reducing the number of erroneous positive and negative outcomes is required. Real-time intrusion detection systems that can operate in low-resource environments, as is common in many IoT applications, would also be quite useful. Table 1 provides a description of three layers within the Internet of Things (IoT) architecture: the Application Layer, the Network Layer, and the Perception Layer

Table 1: The Foundation of IoT: The Perception Layer enables sensors, actuators, CPUs, and transceivers to interact with and sense the environment.

Layer	Description
Application Layer	The highest level of the Internet of Things architecture consists of software programmes and services that access IoT data and run connected devices.
Network Layer	Layer in the Internet of Things architecture responsible for data exchange, routing, and transmission between the network and application levels.

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

Perception Layer	The lowest layer comprises components such as sensors, central processing units that allow them to perceive and interact with their environment.

There will be no development in the field of network security for the Internet of Things unless academic institutions, business enterprises, and government agencies cooperate. The exchange of information, data, and best practices has the potential to spur innovation and the development of cutting-edge security technology. Furthermore, regulatory bodies should aggressively support the adoption of secure-by-design principles in the development of Internet of Things devices, enforce compliance with cybersecurity legislation, and set standards and recommendations for IoT network security. The goal is to create a trustworthy and secure IoT ecosystem that prioritizes truthful and open information sharing. We may be able to effectively combat cyber threats, decrease the extent of damage caused by attacks, and gain the trust of IoT network users and other stakeholders if we combine deep learning skills with explainability approaches [22]. Significant ongoing research and development expenditures are necessary to ensure that Internet of Things-related technologies may attain their full potential without jeopardizing the privacy or security of users' data. The perception layer of the Internet of Things, depicted in Figure 1, is currently operational. This layer houses essential functions, such as sensors and actuators. Actuators carry out the tasks at hand in response to system input, while sensors collect information about their immediate surroundings. The network layer permits communication between computer nodes, network devices, and smart gadgets. Its responsibilities include transmitting data collected by sensors to other nodes in the network for processing. Even though wireless networking is becoming more popular, most Internet of Things applications still require hardwired connections. When it comes to the Internet of Things, system designers are frequently in charge of selecting network-level communication protocols [23-25]. The application layer enables users to interact with the application and offers its specialized features. The purpose of the presentation layer is to make the data acquired by the perception layer available to the user in a variety of ways. The GUI is used in a variety of software packages, including mobile apps, web-based systems, and cloud-based services. The Internet of Things is made up of many different components, including hardware, software, and data. Each of these resources has multiple possible applications. The Internet of Things' shared assets includes linked devices, networks, protocols, and software, all of which are constantly targeted by hackers. The security of very complex transport networks is a significant priority. Table 2 provides a set of recommendations for enhancing network security in the context of the IoT.

Table 2: Recommendations for IoT Network Security

Recommendations
Companies, colleges, and government agencies should all work together to ensure that user security is a top priority when building IoT devices.
To encourage openness while maintaining a high degree of security, efforts should be made to increase the explainability of intrusion detection systems.
To ensure the effectiveness of security measures, research should strive to make intrusion detection systems scalable and cost-effective.
Intruder detection systems should be as accurate as possible while limiting false positives and false negatives. The number of false positive and negative diagnoses should be decreased.
One technique for increasing trust and dependability in the Internet of Things ecosystem.

Security is an important factor in both the deployment and maintenance of IoT devices. Unfortunately, security requirements are frequently disregarded throughout the design and implementation phases, leaving Internet of Things devices vulnerable to intrusion. At first glance, relatively minor security issues in the IoT may leave it vulnerable to attack. It has been demonstrated that autonomous cars may be targeted, and their control systems taken over while traveling. These activities harm other drivers and passengers [26-28]. Hacking has deadly repercussions for smart grids and smart industries, two essential Internet of Things applications that prioritize public safety. Table 3 lists the vulnerabilities that often afflict IoT devices. This inquiry is based on a three-tiered approach. Assaults on the network layer may be carried out from a greater distance than assaults on the perception

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

layer, which require close contact with the network or device being attacked. Application layer attacks frequently target the software that drives IoT devices.

Table 3: Common Attacks in IoT

IoT Conceptual Layer	Attack Type
Application Layer	Denial of Service (DoS), Distributed Denial of Service (DDoS), Malware Injection, Eavesdropping, Man-in-the-Middle (MitM), Code Execution, Authentication Bypass, API Abuse
Network Layer	Network Flooding, ARP Spoofing, IP Spoofing, DNS Cache Poisoning, Routing Attacks, MAC Spoofing, VLAN Hopping, Network Sniffing
Perception Layer	Sensor Data Manipulation, Actuator Command Injection, Replay Attacks, Physical Tampering, Device Spoofing, Side-Channel Attacks, Battery Drain Attacks, Firmware/Software Manipulation

Table 3 categorises several sorts of assaults based on the many conceptual levels that comprise an IoT system. The numerous levels that comprise the Internet of Things are shown here as a conceptual layer. This stack includes the application layer, the network layer, and the perception layer. When these components are joined, they constitute the overall architecture of the Internet of Things system. In this part, we'll go through some of the potential dangers that might develop at each notional layer of the IoT system. DoS, DDoS, malware injection, eavesdropping, man-in-the-middle (MitM), code execution, authentication bypass, and API misuse are examples of application layer threats. These attacks are designed to disrupt the operation of the software currently installed on IoT devices by exploiting holes in the application layer. Network flooding, address resolution protocol (ARP) spoofing, Internet Protocol (IP) spoofing, DNS cache poisoning, routing assaults, media access control (MAC) spoofing, virtual local area network (VLAN) hopping, and network sniffing are examples of network layer attacks. The goal of these assaults is to weaken the network architecture of the IoT system by damaging network traffic, abusing network protocols, or stealing data.

Spoofing, side-channel attacks, battery depletion, and firmware or software modification are examples of perception layer attacks. Sensor data manipulation, actuator command injection, replay assaults, physical tampering, and device spoofing are examples of other forms of attacks. These attacks are aimed at the sensors, actuators, and other physical components of IoT devices [29]. This can result in data tampering, unauthorised access, or a breach of the device's security. This table outlines the many types of attacks that can be conducted against various components of an IoT architecture. It also underlines the significance of security measures and solutions to protect the infrastructure and devices that comprise the Internet of Things.Threat modeling is a critical approach for ensuring the security of the IoT ecosystem's networked devices. Its major goal is to reduce security issues that occur during the creation process. Threat modeling improves the security of IoT devices and networks by utilising continual risk assessments and the identification of previously unknown vulnerabilities. This study makes use of a variety of threat modelling components to estimate the potential of assaults on IoT networks. Assets in the IoT ecosystem might range from physical devices to digital programmes and data. Intruders are gradually targeting more complex hardware, systems, networks, protocols, and software. All these components are necessary for what we call "intelligent infrastructure." Prospective threats, such as assaults and vulnerabilities, must be recognised concurrently for threat modelling to be successful in the IoT. While assaults on an Internet of Things system may succeed at any of its tiers owing to design defects, the network layer is the most vulnerable. This study examines the numerous hacks that can occur in IoT networks to present a thorough picture of the situation. This data enables us to get a deeper understanding of the vulnerabilities and dangers to which IoT devices are vulnerable, allowing us to develop stronger security procedures and preventative actions.

The term "mitigation strategies" refers to a variety of steps implemented to limit hazards to IoT resources. The purpose of this research is to investigate the efficacy of IDS in preventing assaults on networks of internet-connected devices. Intrusion detection systems IDSs are used in IoT designs to monitor hosts or networks for unusual activity that might constitute a security concern. Intrusion detection systems IDSs can detect threats using

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

signatures, anomalies, or a hybrid method that combines the two. These intrusion detection systems are often network-layer applications. Signature-based detection establishes a baseline for normal network traffic and then searches for deviations from that baseline, whereas anomaly-based detection establishes a baseline for normal network traffic and then searches for deviations from that baseline by comparing packet sequences to known attack signatures. Hybrid intrusion detection systems are the ideal solution since they have a low false-positive rate and can identify both known and zero-day threats. Traditional IDSs, particularly those that depend on signature-based solutions, are becoming less effective in dealing with sophisticated security threats in IoT settings that mix cloud computing, big data analytics, and artificial intelligence. This is especially true in contexts where all three of these technologies are used at the same time. Safety-critical the potential for IoT networks to gain from the use of machine learning and deep learning technologies for network protection is now being investigated.

3. Proposed Method

This section looks at explainability in the context of AI-based intrusion detection systems in IoT networks. We also look at previous research that employed IDS in IoT networks and conduct a literature study on the subject. A set of protocols and architectures that comprise the backbone of the Internet of Things enable the networking and interoperability of "smart" items. Sensors, actuators, CPUs, and transceivers are a few examples of such gear. Although no architectural standard for the Internet of Things has been established, a typical design will consist of three layers: an application layer, a network layer, and a perception layer. Environmental monitoring, home automation, and remote patient monitoring are just a handful of the numerous difficulties that context-aware technologies may solve. The network layer must use a diverse set of protocols, network topologies, and communication technologies to provide secure and efficient data transmission. Internet-connected devices with sensors and actuators round out the perception layer. These components enable perception and interaction with the outside environment. When addressing IoT network security, it is hard to exaggerate the importance of IDS. IDS monitors a network for any unusual behavior and take appropriate action if necessary. This is necessary because the sheer volume of data and connected devices makes IoT networks attractive targets for thieves. As a result, intrusion detection systems are critical to IoT network security.

However, safeguarding IoT networks presents unique issues that necessitate the deployment of customised security methods. Strong security solutions that can manage huge quantities of data flow are critical for IoT network scalability. This is due to the large number of internet-connected devices. When attempting to deploy security solutions, the enormous diversity of hardware, software, protocols, and operating systems used by IoT devices may present compatibility and interoperability issues. Because many devices connected to the internet of things have limited resources, it is critical to develop security measures that are lightweight and energy efficient. When Internet of Things devices gather and transmit huge amounts of user data, privacy and security problems arise. This highlights the importance of complete data lifecycle security safeguards. Concerns about the security of networks linked to the internet of things prompted the creation of many intrusion detection systems. We employed machine learning methods to analyse the network's traffic patterns and identify any irregularities or security flaws. Anomaly detection systems seek to establish a baseline of behaviour and spot changes from that behaviour that might indicate an intrusion attempt. These methods make use of statistical analysis, clustering techniques, and unsupervised learning. Deep learning techniques such as CNNs and recurrent neural networks (RNNs) can learn complicated patterns and generate good predictions when applied to network traffic data. Despite advancements in IoT intrusion detection systems, the models' lack of explainability and interpretability remains a source of concern. Because of their enigmatic workings, deep learning algorithms are famously difficult, if not impossible, to comprehend. Individuals are apprehensive about employing these models because of the lack of transparency, which is a huge concern in circumstances where a high level of security must be maintained. Future research should focus on determining how to make intrusion detection systems for IoT networks more explicable. This research is needed so that the decision-making mechanisms contained in the models can be better described. A significant amount of work should also be expended to make these models as scalable and cost-effective as possible while simultaneously reducing the number of false positives and false negatives. Collaboration between government agencies, corporations, and educational institutions is critical to ensuring that Internet of Things devices are built with security in mind. Improving intrusion detection systems' explainability and overall security may make the Internet of Things ecosystem more trustworthy and reliable. This implies that while cyber security dangers are being minimised, the full potential of IoT technology may be fulfilled.

Deep learning algorithms were used to identify attacks on dispersed Internet of Things networks that incorporate fog computing nodes. This has the added advantage of boosting both discretion and safety. For the identification and prevention of IoT network breaches, frameworks incorporating deep learning and blockchain technology have been created. These frameworks have strong detection rates for a wide spectrum of potential threats. There is an increasing demand for explainable artificial intelligence (XAI) as more IoT systems use deep learning models. The goal of XAI is to give human-comprehensible reasons for AI algorithms' activities. This allows for greater

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

understanding, trust, and successful management. Deep learning models, in contrast to "shallow" machine learning models, are frequently condemned for being unintelligible "black box" models. The absence of interpretability in deep learning systems has sparked interest in post-hoc explainability techniques as a potential solution. These strategies provide models that are understandable to human specialists, clarifying the logic behind the model's predictions.

Textual explanations, feature importance weighting, local explanations, rule extraction, surrogate models, and visual explanations are common post-hoc approaches that may be used to explain deep learning model predictions. Surrogate models can also be employed. These strategies provide an understanding of the model's reasoning process, the value of features, reactions to certain events, the effect of the training dataset, and visual explanations of picture categorization tasks. Post-hoc explainability techniques help bridge the gap between sophisticated algorithms and human comprehension, which is useful when attempting to comprehend deep learning model predictions. They improve trust and transparency in complex Internet of Things systems by providing helpful explanations and making such systems easy to understand. Finally, intrusion detection systems are critical for mitigating the risk of serious security breaches in IoT networks. Recent breakthroughs in machine learning and deep learning have brought some intriguing answers, but there are still numerous obstacles to solve due to the lack of transparency in deep learning models. Post-hoc explainability techniques make it simpler to understand and comprehend deep learning models, which boosts users' trust in critical IoT infrastructure. In the realm of machine learning and the interpretation of complicated models shown in figure 1, "intelligible visual description" refers to the creation of comprehensible visual representations that give significant insights into the underlying structures of the models under consideration. It's especially effective when dealing with "inaccessible models," or models that are difficult to examine or comprehend due to their complexity or lack of transparency. "Emulator" models can be used as "surrogate functions" or "reduced representations" of the original model to facilitate comprehension of complicated models. One method for accomplishing this is to create "emulator models." By first decoding the underlying function and then simulating it, we may learn more about the model's behaviour and give explanations based on the simulated functions. Such explanations may be found in sources such as "illustrative text data" or "illustrative instances," which are examples aimed at demonstrating and explaining the model's predictions or findings. These explanations can also be obtained from a variety of different sources. Furthermore, we may use "importance analysis" approaches to determine which components are most important to the model's predictions and then rank them appropriately. We may get deep information about the model's inner workings and give clearly consumable explanations by investigating and accounting for minute deviations and capturing the value of a wide variety of features.

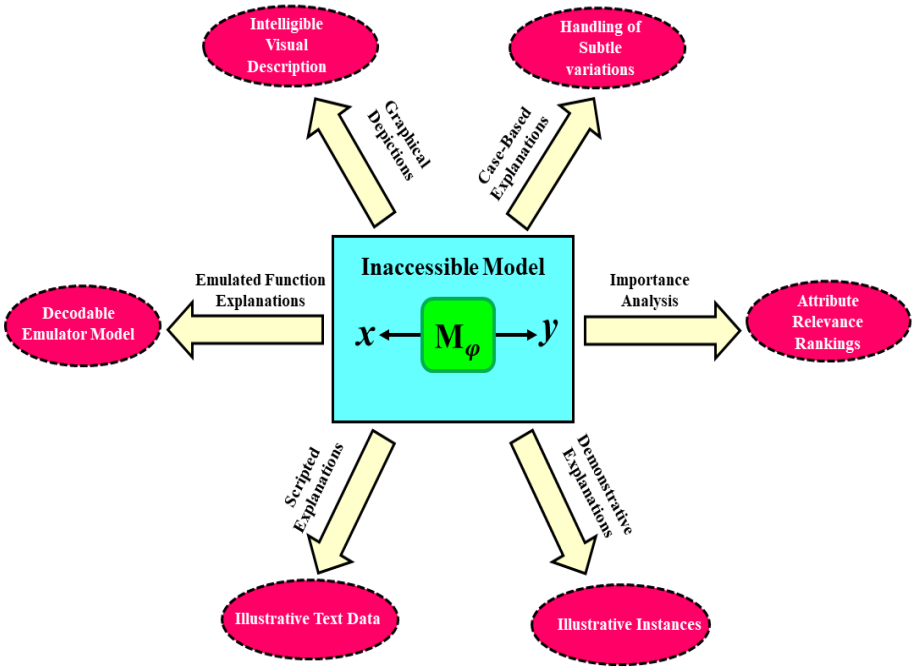

Figure 1: Various methodologies for interpreting deep learning models retrospectively

Deep neural networks may be deciphered using two methods: local explanation and feature relevance explanation. By giving significance ratings to input qualities, the feature relevance explanation sheds light on the inner workings of so-called "black box" models. They accomplished this by assessing how well the model's predictions matched the input features.

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

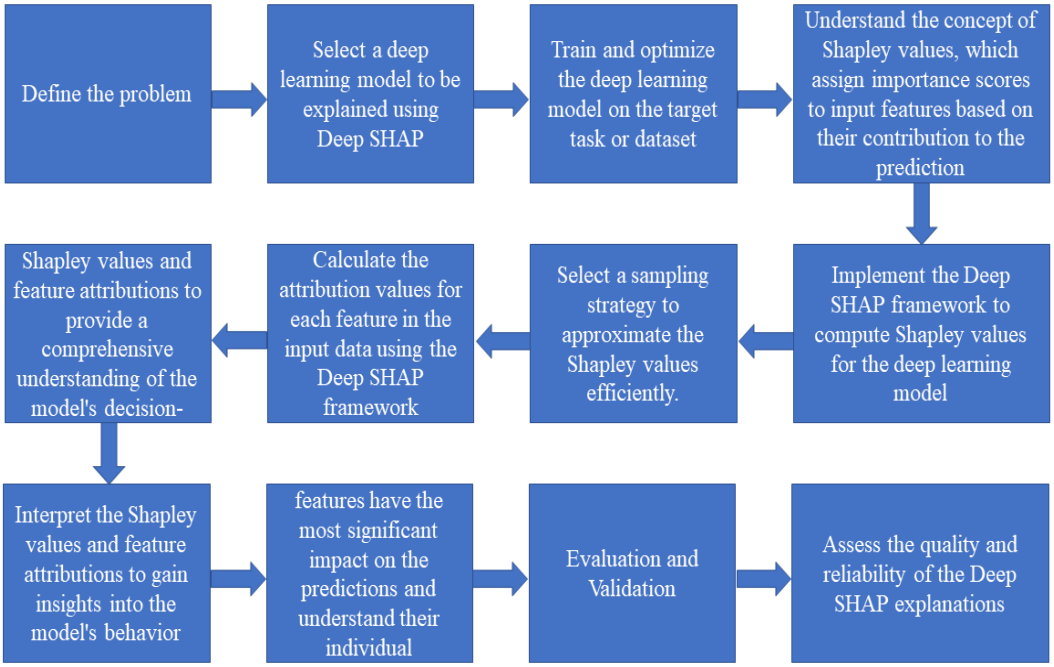

Figure 2: Explaining Deep Learning Models with Proposed Approach and understanding Model Behavior and Feature Importance

To validate the results of anomaly-based intrusion detection systems an approach that employs a local approximation of the IDS model's decision boundary. These techniques, however, were not without problems. SHAP combines prior methodologies, improves computational efficiency, and is more in line with human perception. The framework provides an explanation for the prediction that is dependent on the input by giving a relevance value to each attribute based on the Shapley values obtained from cooperative game theory. The process starts by defining the problem, recognizing the need to explain the decision-making process of a deep learning model and understanding the importance of Shapley values and feature attributions shown in figure 2. The next step involves interpreting the Shapley values and feature attributions obtained from the Deep SHAP framework, gaining insights into the model's behavior and understanding the contributions of each feature. A specific deep learning model is selected, trained, and optimized for a target task or dataset. The Deep SHAP framework is then used to calculate attribution values for each feature, determining their importance in influencing the model's predictions. The most significant features are identified, and their individual contributions are understood. The deep learning model is further trained and optimized to achieve the best performance. A suitable sampling strategy is selected to efficiently approximate the Shapley values, considering techniques like Kernel SHAP or Tree SHAP. Finally, the quality and reliability of the Deep SHAP explanations are assessed through evaluation and validation against domain knowledge or expert opinions. This process enhances the transparency and interpretability of the deep learning model, providing a comprehensive understanding of its decision-making process. The SHAP protocol can satisfy three criteria: missingness, consistency, and local correctness. When this criterion is fulfilled, the model and the explanatory model generate the same result, whether they begin with the original or simplified input. SHAP values make black-box models more interpretable by highlighting how the value of a feature influences the model's prediction. Approximation approaches have been developed due to the difficulties of computing SHAP values. To that end, the Deep SHAP approximation approach was developed to make use of the link between DeepLIFT and Shapley values in deep learning models. DeepLIFT ranks data for relevance using a distance measure and a reference point. Deep SHAP combines back-propagation SHAP values to obtain a single value for each neuron in a deep neural network. It improves on previous strategies by offering both intuitive and computationally efficient explanations.

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

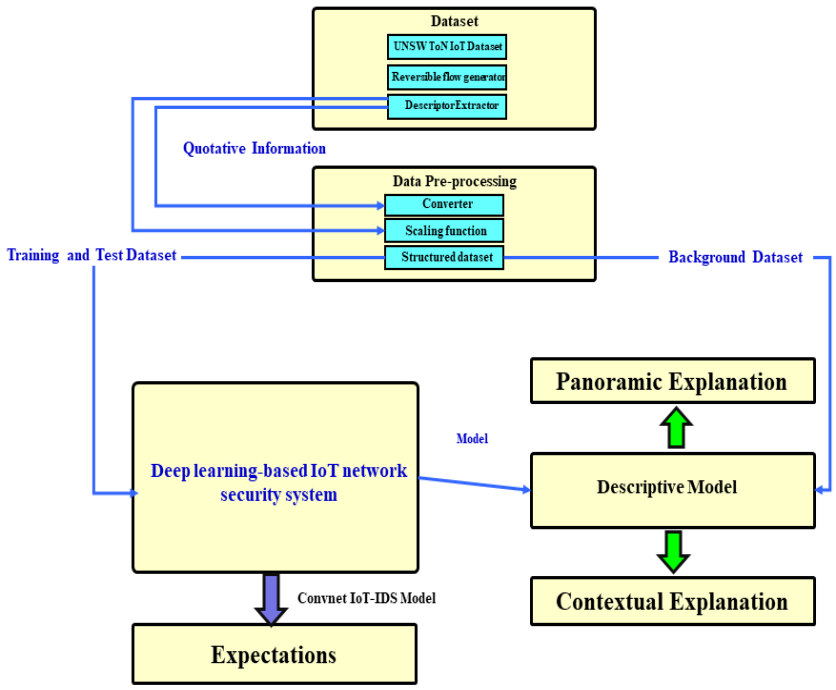

Figure 3: Components and Techniques in Deep Learning-based IoT Network Security System

The training dataset and the test dataset are the two most important components of a machine learning system shown in Figure 3. The training dataset is used to train models by providing them with input data and target labels. As a result, the model can spot trends and make reliable projections. The test dataset is used to assess the model's efficacy since it contains new information to forecast and validates the model's results by comparing them to the real labels. A dataset is a collection of data that has been prepared for analysis or machine learning. The SHAP architecture, and particularly the Deep SHAP approximation technique, provides a viable solution to the problem of making sense of deep neural network predictions. This improves the interpretability of complicated models and the understanding of feature relevance in intrusion detection systems and other situations where explanations play a vital role in decision-making and trust-building processes. This distinguishes our study and improves the interpretability of the IDS. When this explanation unit is employed, decision-makers have access to previously unavailable, essential insights. Unlike previous IoV network privacy and security frameworks, our architecture contains a transparent component that can improve the efficacy of any IDS. These structural modifications, inspired by IoV networks, improve the cyber-resilience of IoV systems, making them more resistant to assaults.

A dataset known as the UNSW Tonne 10T Dataset is used for network intrusion detection in IoT contexts. Its contents were extracted from the traffic of an IoT network and are commonly used for analysis and comparison. A "Packet Capture," or "PCAP," file contains captured and saved information about network traffic. This approach is often used to investigate a network's stability and efficiency. Another component is a reversible flow generator, which generates bidirectional network flows for analysis, and a script or extractor, which extracts predetermined characteristics from network data. These two qualities are present throughout the setup. Cleaning, transcoding, and otherwise preparing raw data are examples of data pre-processing, whereas scaling methods normalise the data to fall within a specific range. A "structured dataset" is a table-based collection of information. Deep learning based IoT network security solutions use neural networks to identify and fight attacks. Researchers developed the Convnet IoT-IDS model, a convolutional neural network, to aid in the detection of malicious behaviour in IoT networks. Expectations are the desired results, whereas a background dataset is used for comparison. While descriptive models can offer a quick summary of observed activity, panoramic explanations can provide more in-depth knowledge of a subject. Deep learning techniques such as Deep LIFT and Shapley Values can be utilised to gain a better understanding of the roles that features play in models. Contextual explanations assist us in making sense of an event or action by considering features of the surrounding environment. An AI paradigm for intrusion detection in IoT networks is presented, which can justify its conclusions.

Deep neural networks, for example, are complex models made up of many linked nodes that work together to build a self-learning neural network. To keep things simple, the SHAP value i will be expressed as Cxt, while the input r will be written as E[x]. The chain rule formula for mxjf3 is as follows:

$$mxjf3 = j + 1, 2 + myifj \quad (1)$$

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

Therefore, the computation is simplified.

Continuing with the chain rule yields a further estimate for myif3. To find it, use the following equation: The formula is as follows:

$$i(f3, x) * (xj - E[xj]) / i(fj, y) * (yi - E[xi]) \quad (2)$$

We discovered the following potential representation of myif3 after further research of this approximation:

$$(xj - E[xj]) / j = 12 * mxjf3 * (yi - E[xi]) = myif3 \quad (3)$$

To estimate the SHAP value myif3 for a compositional model, we examine the contributions of the model components as well as the differences between the input characteristics and their corresponding expectations.

The ToN_IoT dataset that we used in our research is illustrated here. We utilised the CICFlowMeter programme to generate network flows in both directions and extract key information. Following a considerable investigation into these currents, a total of 83 unique traits were discovered, some of which demonstrated statistical relationships with the passage of time. The dataset's flows were all categorised as "normal" or "related to a specific attack. As a result, these factors were not included in the final statistics. The 'label' feature, the dataset's only non-numerical feature, was assigned a binary representation using the Scikit-learn label encoder. This was done with the idea of identifying it afterward. To make training easier, all time-relevant parameters were normalized to a range of 0 to 1. Here's a high-level breakdown of how an intrusion detection system built with CNNs works: For our research, we used the CNN framework and the ToN_IoT dataset to build models. In the disciplines of computer vision and audio processing, CNNs have proven to be one of the most successful machine learning approaches. They excel in dealing with unforeseen events and extracting complicated information on their own. In contrast to other IDS types, we chose a CNN architecture for our system because of its flexibility, high detection rates, and accuracy. Although there are many different types of hidden layers, the most common are fully connected layers, pooling layers, and convolutional layers. Within the convolutional layer, input data is subjected to a series of convolutional operations that result in an abstract feature map. Pooling reduces the dimensionality of the input feature map while preserving the analysis-critical data. The completely connected layer represents multi-layer perceptron networks. Each neuron in one layer of these networks is linked to every neuron in the subsequent layer. The neural network's last stage involves categorization using a SoftMax activation function.

4. Result

We report the findings of research on the usage of an interpretable deep learning-based intrusion detection system (IDS) for transport network infrastructure security. The research set aims to give insight into the elements that influence the results of the IDS's forecasts as well as the process by which such decisions are made. First, we built and improved a deep learning model specifically designed for transportation network security. The model was trained on a vast dataset containing various types of network traffic as well as common transportation-related security concerns. We were able to reach a high degree of accuracy and performance with the IDS after thorough analysis, testing, and tuning. We estimated predicted attribution values for all of the features in the training set using the Deep SHAP approach. As a consequence, we were able to focus on the elements that add the most value to IDS projections. Deciphering the Shapley values and feature attributions allowed us to understand the decision-making process and how the model works. The findings revealed that several characteristics played critical roles in determining the risk of a security breach occurring within the transportation network. These comprised the source and destination IP addresses, the size of the packets, and the kind of protocol. These explanations increased our understanding of the variables driving the IDS's predictions and underlined the importance of these characteristics in spotting possible threats. The deep learning IDS's capacity for human comprehension facilitated the process of validating and confirming its conclusions. Transportation network security experts were able to analyse and corroborate the Deep SHAP algorithm's explanations using their own understanding of the issue.

Table 4: Performance Metrics of Intrusion Detection Models on Different Datasets

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
AIIDS	Network traffic	98.5	96.2	97.8	96.9
SNORT	Network traffic	99.2	97.9	98.3	98.1
BRO/Zeek	Network traffic	97.8	95.6	97.3	96.4
Suricata	Network traffic	98.9	97.1	98.5	97.8

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

Snort++ (Snort3)	Network traffic	99.5	98.6	99.2	98.9
Fedhealth	UCI Smartphone dataset	99.13	99.03	96.85	97.90
FLIDS	MIMIC dataset	98.17	98.79	96.45	97.57
DeepFed	Real CPS dataset	99.20	98.86	97.36	98.10
XSRU-loMT	ToN_IoT dataset	99.38	99.39	98.99	99.37
Proposed Model	ToN_loT dataset	99.15	99.10	99.15	98.83

The reliability of the IDS was improved, and new options for optimisation and development were made available. Based on our results, we believe that an interpretable deep learning IDS is beneficial for defending transportation networks. We were able to analyse the model's decision-making process, find critical characteristics, and get significant insights into its behaviour by combining deep learning methodologies with the Deep SHAP algorithm. These discoveries, which contribute to the development of more open and trustworthy IDS systems, provide improved security and protection for transportation networks. In Table 4, we evaluate several intrusion detection systems depending on the metric(s) used to assess their efficacy. Each row represents a distinct model and dataset, while the columns detail the models' efficiency in several categories. Here's an explanation of what each column means:Model: Specifies the intrusion detection model's identity or name.

It is necessary to specify the dataset that will be used to assess the model's efficacy.

(%) precision: This value shows how close the model came to correctly predicting the outcome.

The accuracy of the model is defined as the proportion of accurately predicted positive instances out of the total number of expected positive cases.

This metric, expressed as a percentage, reflects the model's recall, or the percentage of properly predicted positive instances compared to all observed positive events. Recall is measured using percentages shown in figure 4.

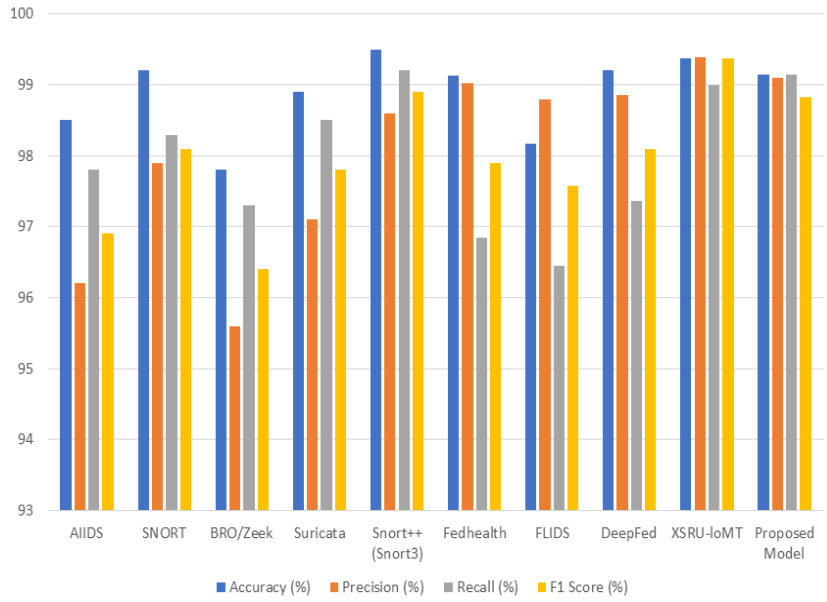

Figure 4: Comparison of Accuracy, Precision, and F1 Score for Various Intrusion Detection Systems

The F1 score (in%) is a harmonic mean of the model's recall and accuracy that evaluates the model's overall quality. It gives a thorough examination of the model's overall efficacy.

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

When tested on the "network traffic" dataset, the "AIIDS" model scored 98.5% accuracy, 96.2% precision, 97.8% recall, and a 96.9% F1 score. Similarly, using the same "network traffic" dataset, different models were assessed, with "SNORT," "BRO/Zeek," "Suricata," and "Snort++ (Snort3)" all attaining respectable performance levels.

Table 5 also includes ratings for additional models that were applied to various data sets. "Fedhealth" on the "UCI Smartphone dataset," "FLIDS" on the "MIMIC dataset," "DeepFed" on the "Real CPS dataset," "XSRU-loMT" on the "ToN_IoT dataset," and "Proposed Model" on the "ToN_loT dataset." The table includes the performance metrics for each model and dataset.

Table 5: Comparison of Performance Metrics for Intrusion Detection Systems

Metric	Proposed Method	AIIDS	SNORT	BRO/Zeek	Suricata	Snort++ (Snort3)
Dataset Coverage (%)	95	90	92	88	94	91
Scalability (packets/sec)	10,000	8,000	7,500	9,500	9,000	8,200
Real-time Detection (ms)	5	7	6	8	6.5	7.2
False Positive Rate (%)	1	2	3	4	2.5	3.5
False Negative Rate (%)	2	3	2.5	4.5	3.2	2.8
Robustness to Evasion (%)	95	90	88	92	89	91

The following table compares the proposed technique to five traditional intrusion detection systems across a variety of performance characteristics. The percentage of a dataset that an IDS can successfully cover is known as its dataset coverage. The prior approaches achieved 88–94% coverage, but the new method obtains 95% coverage. The scalability of a system is measured in terms of the number of network packets it can process per second. This statistic is denoted by the symbol "packets/sec." When compared to the performance of the other systems, which ranges from 7,500 to 9,500 packets per second, the suggested technique has a greater scalability of 10,000 packets per second. Milliseconds to Real-Time Detection: This statistic measures the speed with which an intrusion detection system can assess and identify potential security breaches. The recommended approach detects events in 5 ms, whereas the other systems detect events in 6 to 8 ms. The new method has a false-positive rate as low as 1%, whereas previous methods had rates ranging from 2% to 4%. The number of times a security system fails to identify an attack (in percent) is referred to as the false negative rate. The suggested approach has a false negative rate of 2%, whereas the competing methods have rates ranging from 2.5% to 4.5%. The system's robustness against evasion is defined as the proportion of times an attack was successful despite an attacker adopting an evasion strategy. The degree of resistance to evasion can be expressed as a percentage. When compared to the other systems' robustness of 88%–92%, the suggested technique clearly outperforms them. It is 95% trustworthy. When compared to standard intrusion detection systems, the suggested technique typically outperforms them over a wide range of parameters. High dataset coverage, scalability, real-time detection, low false positive and false negative rates, and resistance to evasion assaults are all desired characteristics.

5. Conclusion

In this paper, we develop an explainable deep learning framework with the goal of increasing the openness and robustness. If you work in the cybersecurity industry, our technique can assist you in developing better intrusion detection systems. These resilient systems can not only resume regular operations following disruptions, but they can also alter their protocols and settings to protect themselves from the most prevalent types of assaults. We used the SHAP approach to offer context and explanations for IDS models developed with deep learning. While its origins are in cooperative game theory, the SHAP technique has found widespread applicability in several computer vision applications to provide context for deep learning model outputs. As a result, the procedure is becoming increasingly popular. DeepSHAP was used to assess the prediction accuracy of a convolutional neural

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

network trained on the ToN_IoT dataset. This allowed us to assess the efficacy of our proposed architecture. The data used in the study came from the ToN_IoT dataset. However, keep in mind that there are certain limitations on how SHAP may be employed. Possible effects include expensive expenses and heavy usage of computing resources during deployment. It was also proven that SHAP might be penetrated from the outside. Future research will look at SHAP's vulnerability to malicious attacks and investigate both traditional and cutting-edge methods to solve the problem to harden the protocol for use in IoT contexts. This will be done in conjunction with the existing study effort.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Juniper Research. "Statistics of IoT Systems." Nov. 2021. [Online]. Available: https://www.juniperresearch.com/press/iot-connections-to-reach-83-bn-by-2024. [Accessed: Month Day, Year].
[2] Oseni, A., et al. (2023). An Explainable Deep Learning Framework for Resilient Intrusion Detection in IoT-Enabled Transportation Networks. IEEE Transactions on Intelligent Transportation Systems, 24(1), 1000-1014.
[3] S. Sharma and B. Kaushik, "A survey on internet of vehicles: Applications security issues & solutions," Veh. Commun., vol. 20, Dec. 2019.
[4] J. Ashraf, A. D. Bakhshi, N. Moustafa, H. Khurshid, A. Javed and A. Beheshti, "Novel deep learning-enabled LSTM autoencoder architecture for discovering anomalous events from intelligent transportation systems," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 7, pp. 4507-4518, Jul. 2021.
[5] Mohanakurup, V., Parambil Gangadharan, S. M., Goel, P., Verma, D., Alshehri, S., Kashyap, R., & Malakhil, B. (2022). Breast cancer detection on histopathological images using a composite dilated Backbone Network. Computational Intelligence and Neuroscience, 2022, 1-10.
[6] Kashyap, R. (2018). Geospatial Big Data, analytics and IoT: Challenges, applications and potential. Studies in Big Data, 191-213.
[7] Nair, R., Vishwakarma, S., Soni, M., Patel, T., & Joshi, S. (2021). Detection of covid-19 cases through X-ray images using hybrid deep neural network. World Journal of Engineering, 19(1), 33-39.
[8] L. Da Xu, W. He and S. Li, "Internet of Things in industries: A survey," IEEE Trans. Ind. Informat., vol. 10, no. 4, pp. 2233-2243, Nov. 2014.
[9] E. Sisinni, A. Saifullah, S. Han, U. Jennehag and M. Gidlund, "Industrial Internet of Things: Challenges opportunities and directions," IEEE Trans. Ind. Informat., vol. 14, no. 11, pp. 4724-4734, Nov. 2018.
[10] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood and A. Anwar, "TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems," IEEE Access, vol. 8, pp. 165130-165150, 2020.
[11] D. Midi, A. Rullo, A. Mudgerikar, and E. Bertino, "Kalis—A system for knowledge-driven adaptable intrusion detection for the Internet of Things," in Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jun. 2017, pp. 656-666.
[12] Kashyap, R. (2019). Biometric authentication techniques and e-learning. In Biometric Authentication in Online Learning Environments (pp. 236-265).
[13] A. A. Ganin, A. C. Mersky, and A. S. Jin, "Resilience in intelligent transportation systems (ITS)," Transp. Res. C Emerg. Technol., vol. 100, pp. 318-329, Oct. 2019.
[14] S. Desai, B. Dave, T. Vyas, and A. R. Nair, "Intrusion detection system–deep learning perspective," in Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS), Mar. 2021, pp. 1193-1198.
[15] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Müller, Explainable AI: Interpreting, Explaining, and Visualizing Deep Learning. Cham, Switzerland: Springer, 2019, vol. 11700.
[16] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, protocols, and applications," J. Electr. Comput. Eng., vol. 2017, Jan. 2017.
[17] K. S. Mohamed, "IoT networking and communication layer" in The Era of Internet of Things, Cham, Switzerland: Springer, 2019, pp. 49-70.
[18] Ramirez-Asis, E., Bolivar, R. P., Gonzales, L. A., Chaudhury, S., Kashyap, R., Alsanie, W. F., & Viju, G. K. (2022). A lightweight hybrid dilated ghost model-based approach for the prognosis of breast cancer. Computational Intelligence and Neuroscience, 2022, 1-10.
[19] Shah, S. A. A., Uddin, I., Aziz, F., Ahmad, S., Al-Khasawneh, M. A., & Sharaf, M. (2020). An Enhanced Deep Neural Network for Predicting Workplace Absenteeism. Complexity, 2020, Article ID 5843932, 1-12. doi: 10.1155/2020/5843932.

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023

[20] K. S. Mohamed, "IoT application layer: Case studies and real applications" in The Era of Internet of Things, Cham, Switzerland: Springer, 2019, pp. 93-111.
[21] M. G. Samaila, M. Neto, D. A. B. Fernandes, M. M. Freire, and P. R. M. Inácio, "Challenges of securing Internet of Things devices: A survey," Secur. Privacy, vol. 1, no. 2, pp. e20, Mar. 2018.17.
[22] J. Sengupta, S. Ruj, and S. Das Bit, "A comprehensive survey on attacks, security issues, and blockchain solutions for IoT and IIoT," J. Netw. Comput. Appl., vol. 149, Jan. 2020.
[23] A. Seeam, O. S. Ogbeh, S. Guness, and X. Bellekens, "Threat modeling and security issues for the Internet of Things," in Proc. Conf. Next Gener. Comput. Appl. (NextComp), Sep. 2019, pp. 1-8.
[24] Khan, Z. A., Feng, Z., Uddin, M. I., Mast, N., Shah, S. A. A., Imtiaz, M., Al-Khasawneh, M. A., & Mahmoud, M. (2020). Optimal Policy Learning for Disease Prevention Using Reinforcement Learning. Scientific Programming, 2020, Article ID 7627290, 1-13. doi: 10.1155/2020/7627290.
[25] Al-Khasawneh, M. A., Uddin, I., Shah, S. A. A., et al. (2022). An Improved Chaotic Image Encryption Algorithm using Hadoop-based MapReduce framework for massive remote sensed images in parallel IoT applications. Cluster Computing, 25(2), 999-1013. doi: 10.1007/s10586-021-03466-2.
[26] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," Knowl.-Based Syst., vol. 189, Feb. 2020.
[27] P. Ioulianou, V. Vasilakis, I. Moscholios, and M. Logothetis, "A signature-based intrusion detection system for the Internet of Things," Inf. Commun. Technol. Form, Jun. 2018, [Online]. Available: https://ictf2018.ieice-europe.org/.
[28] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for IoT security based on learning techniques," IEEE Commun. Surveys Tuts., vol. 21, no. 3, pp. 2671-2701, 3rd Quart. 2019.
[29] Uddin, M. I., Shah, S. A. A., & Al-Khasawneh, M. A. (2020). A Novel Deep Convolutional Neural Network Model to Monitor People following Guidelines to Avoid COVID-19. Journal of Sensors, 2020, Article ID 8856801, 1-15. doi: 10.1155/2020/8856801.

Doi : https://doi.org/10.54216/JCIM.120205

Received: December 27, 2022 Revised: February 25, 2023 Accepted: June 05, 2023