A Deep Transfer Learning Framework for Robust IoT Attack Detection: A Review

Hanan Abbas Mohammed* and Idress Mohammed Husien
Department of Computer Science, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq
E-mail: stcm22005@uokirkuk.edu.iq,idress@uokirkuk.edu.iq
*Corresponding author

Keywords: Internet of things (IoT), deep transfer learning (DTL), attack detection, deep learning, data labeling

Received: March 26, 2024

Our lives have been significantly altered due to the digital revolution, and the Internet of Things (IoT) has played a significant part in this transformation. However, the fast expansion of the IoT into almost every aspect of life has resulted in various new cybersecurity dangers. As a result, detecting and preventing possible attacks on IoT networks have lately garnered significant attention from the academic and business worlds. Machine learning (ML)-based techniques, intense learning (DL), have shown considerable promise among the many different approaches to attack detection. This is because they can identify attacks at an early stage. However, for these DL algorithms to be effective, gathering substantial data from IoT devices, including label information, is necessary. On the other hand, the labeling process is often resource-intensive and time-consuming; hence, it may not be able to accommodate rapidly growing IoT threats in the real world. The introduction of DL methods to the IoT datasets is the main emphasis of this study, which also reviews the newest advancements in security measures for threat detection. This review aims to examine DL techniques and continuing breakthroughs in approaches that may be used to produce enhanced attack detection models for IoT frameworks. This is the objective of this review. When applying DL to IoT security, we address the benefits and research gaps associated with each strategy.

Povzetek: Podan je pregled uporabe metod globokega prenesenega učenja za robustno detekcijo napadov na IoT. Poudarjeni so napredki in vrzeli v razvoju učinkovitih modelov.

1 Introduction

Internet of Things (IoT) is the network of sensors, actuators, and devices in cars, appliances, buildings, and other structures. As sensors, data storage, and the Internet improve in speed, accessibility, and integration[1][2][3], IoT devices will have new uses. Smart buildings, cities, transportation, and healthcare are examples. IoT rapidly enters most parts of society, creating new cybersecurity challenges. IoT devices' minimal computing capability makes them vulnerable to third-party assaults. IoT devices need more security than Personal Computers (PCs) because they are more vulnerable [2, 3]. To prevent malicious activities, developing IoT applications need attack detection [1][3][4][5].

Machine Learning (ML)- or signature-based IoT threat detection techniques exist [5][6]. IoT Things attack signatures are checked in incoming traffic [6]. These solutions require IoT threat expertise to define signatures. During offline training, ML-based techniques learn typical and harmful data features. The models detect traffic attacks throughout pre-predicting and real stages. ML-based approaches detect IoT attacks early [1][5][6][7]. Their capacity to automatically and gradually gather crucial data and qualities allows this.
Classifying large data sets is time-consuming and expensive [8][9]. Thus, ML IoT threat detection is limited. IoT threat detection strategies can be signature-based or ML-based [5][6]. Incoming traffic is analyzed for IoT Things attack signatures [6]. These methods require extensive IoT threat expertise to define signatures. However, ML-based approaches try to learn normal and dangerous data features during the offline training phase. These models detect traffic attacks throughout the pre-predicting and live phases. ML-based methods can detect IoT attacks early [1][5][6][7]. Their ability to automatically and gradually gather vital data and attributes makes this possible.
Surveys [10], and [11] divide learning transfer approaches into three basic categories based on the source-destination domain relationship. These papers summarise the transfer learning (TL) literature, which has established many of the most influential methodologies. Additionally, many new and effective methods have been proposed recently. Educational researchers focus on domain adaptability and multi-source domain transfer. Deep Learning (DL) dominates several branches of study today. Finding effective deep neural network-based information transfer methods, or Data Transfer Language (DTL) is vital.
ML approaches, especially DL, are increasingly used in cyber threat detection because of their categorization ability. DL models are good at detecting cyberattacks.

DL models also identify new assault kinds [12]. DL models are good at detecting cyberattacks. DL models also identify new assault kinds This review assists DL and IoT scholars and programmers, especially security-focused ones. The paper's accomplishments are: We identified and underscored IoT security issues. We found 22 IoT security vulnerabilities that deep transfer learning may solve. Recent DTL-IoT security studies were thoroughly analyzed. Our goal was to evaluate this area's practicality and limits. Our objective categorization system uses important data from cutting-edge technology.

2 Methodology

Deep Learning (DL) and Deep Transfer Learning (DTL) are critical components in bolstering cybersecurity protocols, specifically Intrusion Detection Systems (IDS) and the Internet of Things (IoT). Academics endeavor to construct resilient frameworks that effectively identify and alleviate cyber threats associated with the Internet of Things (IoT) by utilizing the resource optimization capabilities of DTL and the hierarchical data processing of DL. Nevertheless, extant research reveals notable deficiencies, such as concerns regarding privacy, scalability, and the necessity for interdisciplinary viewpoints. These findings underscore the intricate nature of protecting IoT environments amidst ever-changing cyber threats.

2.1 Data preprocessing

Data preprocessing is a crucial step in data analysis and machine learning processes. It involves cleaning, transforming, and preparing raw data for modeling and analysis. Common methods include feature selection or extraction, encoding categorical variables, handling missing data, and removing duplicates. Missing Value Handling techniques like as mean/median imputation, interpolation, and row/column removal are often used. Duplicate Removal is a process that removes duplicates to reduce bias and duplication. Techniques such as min-max scaling, z-score normalization, and log transformation are used for feature scaling and normalization. Categorical Variable Encoding is a technique used to convert categorical variables into numerical form to optimize processing efficiency. The process of feature selection and extraction identifies irrelevant characteristics and eliminates them. The process of "splitting" a dataset involves dividing it into several sets for testing, validation, and training. Data augmentation is beneficial in improving the quality of training data by increasing its diversity or addressing class imbalance. To mitigate the impact of anomalous values, the process of Outlier Detection and Handling eliminates data points that deviate significantly from the norm.

2.2 Deep learning (DL)

Deep learning (DL) refers to a subfield of machine learning that applies methods that mimic how the brain processes data[5][10]. For tasks like feature learning and pattern classification, DL architectures consist of a series of interconnected layers, where each layer takes data from the ones below it and rearranges it hierarchically. DL algorithms are often better suited than machine learning approaches in more complicated situations (i.e., with many characteristics and a large amount of data).
For neural network training, there are primarily two steps:

• The feed-forward phase, in which activation of network nodes is carried out from the input layer—which typically contains several nodes proportional to the number of features being considered—to the output layer—which typically contains several nodes proportional to the number of classes, in the case of classification problems. Every node in the intermediate levels represents a neuron that activates its output based on an appropriate and ad hoc activation function (like ReLu)[5][12][13], except the input layer nodes.

• The back-propagation phase, which uses convolution to boost the network's overall performance, is connected to the nodes themselves and, if required, updated weights and bias values to enhance the neural network's overall performance.

2.3 Deep transfer learning (DTL)

DTL involves using information acquired from a different task and dataset, even if they are not closely connected to the original task or dataset, to minimize the resources required for learning. For most DL models, having a substantial quantity of labeled data is essential, which may sometimes be challenging in many ML tasks. At the onset of the COVID-19 pandemic and even a year later, obtaining enough labeled chest X-ray data to train a DL model was difficult. However, by employing DTL, artificial intelligence (AI) was able to successfully detect the disease with a high level of accuracy using a small training set [13][14]. Another use case is implementing ML algorithms on edge devices, such as smartphones, for various activities. This is achieved by using DTL techniques to minimize the computational requirements. DTL and semi-supervised learning are distinct in that DTL allows for different distributions between the source and target datasets while maintaining a relationship. In contrast, semi-supervised learning involves using the same dataset for both the source and target data, with the target set lacking labels [15]. DTL and multiview learning are distinct from each other. Multiview learning involves using multiple datasets to enhance the performance of a single task, such as separating video datasets into image and audio datasets[15]. On the other hand, despite having some similarities, DTL and multitask learning are different. The primary distinction is in interconnections between activities in multitask learning, which facilitates mutual enhancement and enables simultaneous information transfer among related tasks. Unlike in DTL, where the emphasis is on the target domain, and the knowledge of target data is already acquired from source data, there is

no need for them to be connected or function simultaneously [15].

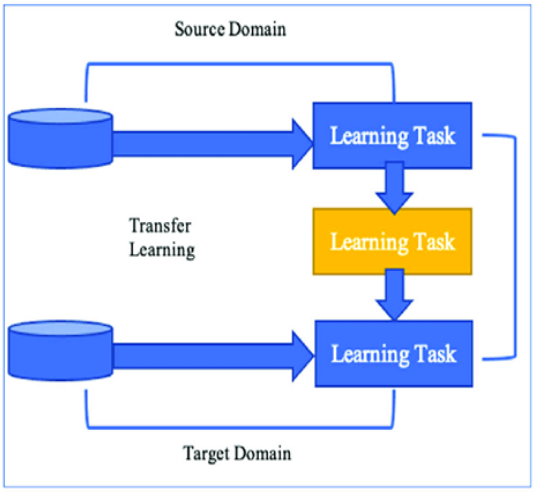

Figure 1: The transfer learning process [16]

2.4 Internet of things

The Internet of Things (IoT) establishes a bridge between the real and virtual worlds by combining various applications built on the merging of smart things with the Internet. Anything from a basic smart home device to complex machinery in an industrial facility might be one of these uses. While the goals of many Internet of Things applications are very varied, they do have some commonalities. Data gathering, transmission, processing, management, and use are the three main stages of an Internet of Things (IoT) operation[17]. Gathering physical environment data is the main focus of the collection phase. This objective is accomplished by integrating sensing devices and technology for short-range communication. Typically, devices used in the collecting phase are compact and have limited resources. At this stage, protocols and technologies for communication are developed with a focus on small distances, restricted data speeds, memory capacity, and energy usage. These features are why Low Latency Networking (LLN) (Low-Power and Lossy Networks) is a common name for collecting phase networks. Compared to the traditional Internet, LLN solutions for error control, medium access control, routing, and addressing could be different.

Transferring the information acquired in the collecting phase to the applications and, ultimately, the users is the goal of the transmission phase. At this stage, the network that connects people and objects across larger distances is built using Transmission Control Protocol/Internet Protocol (TCP/IP) protocols in conjunction with technologies like Ethernet, WiFi, Hybrid Fiber Coaxial (HFC), and Digital Subscriber Line (DSL). Gateways must connect the traditional Internet protocols used during transmission with the LLN protocols used during collection.

2.5 Intrusion detection

Intrusion detection describes finding malicious attempts to access computer networks. An incursion is any attempt to gain unauthorized access to a computer system using these means. Both internal and external intruders are possible. Intruders inside the network who have some lawful access but want more rights so they may abuse their position without authorization are known as internal intruders. People outside the target network attempting to access system data without authorization are known as external intruders[12][10]. Sensors, an analytical engine, and a reporting system comprise a standard Intrusion Detection System (IDS). Sensors are set up at various nodes or hosts in the network. They collect information such as host or network statistics, packet headers, service requests, Operating System (OS) calls, and file system modifications. The analysis engine receives data from the sensors and uses it to study the data to identify persistent intrusions. The reporting system notifies the network administrator whenever the analysis engine finds evidence of an intrusion. There are two main types of intrusion detection systems (figures 2, 3): host-based and network-based. Network-based intrusion detection systems (NIDS) link to several network regions to detect malicious activity in network traffic. Attached to a computer, host-based intrusion detection systems (HIDS) watch for any harmful activity on the device. Compared to NIDS, HIDS is more comprehensive in analyzing system calls, operating processes, file-system modifications, interprocess communication, application logs, and network traffic. Signature-based, anomaly-based, and specification-based intrusion detection systems are further possible categories.

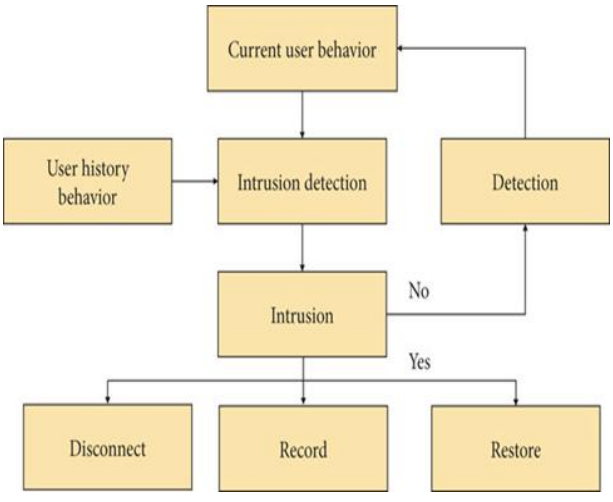

Figure 2: Principle of intrusion detection[18]

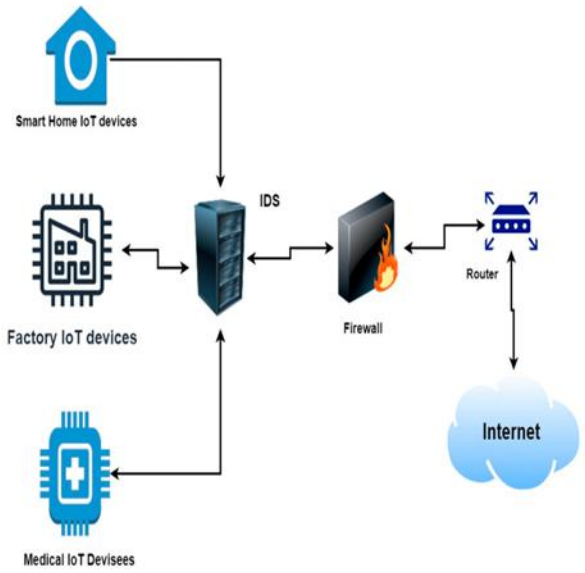

Figure 3. ID-based IoT[19]

2.6 Studies gaps

We identified the most significant gaps in the studies covered by examining a large number of those studies. As a result, we can advise and offer future researchers an approach that will allow them to overcome these challenges in their endeavors. In this section, we discuss such shortcomings. In addition, Table 1 provides a concise summary of the most significant comparisons and a presentation of the difficulties encountered in the examined research.

Much of the research focuses on real-world application, which suggests creative ideas or frameworks but does not provide empirical validation or real-world implementation. For example, while Shukla, Amogh, et al. [1] address creating energy-efficient artificial intelligence models for Internet of Things threat detection, they have not yet explored the actual deployment issues or the system's performance in the real world.

Concerns Regarding Privacy and Ethics That Are Not Given Sufficient Attention: Despite technological progress, there is sometimes a lack of conversation about its consequences for privacy and the ethical issues involved with the offered solutions. Some studies, such as Khoa, Tran Viet, et al. [12] discuss collaborative learning models for detecting cyberattacks; however, these studies do not dive into the possible hazards to privacy or ethical implications of sharing sensitive network data or information.

Numerous studies, including Jaiswal, Aayush, et al.[14], Banaamah[20], and Vu, Ly, et al. [21], concentrate on assessing suggested models by using standard datasets. However, there is a lack of diversity in the exploration of datasets. Despite this, there is a need for more varied and representative datasets to guarantee the robustness and generalizability of the solutions that have been established, particularly in the context of IoT security and medical diagnostics.

Scalability and generalization issues: Although many studies have shown encouraging results in controlled environments or particular use cases, they have not addressed the issues of scalability and generalization. It is not obvious if the collaborative learning framework proposed by Khoa, Tran Viet, et al.[12] is scalable or applicable to various network topologies. For example, the framework presents a collaborative learning framework for intrusion detection.

Integrating interdisciplinary viewpoints Because of the interdisciplinary character of new technologies such as the IoT and deep learning, it is necessary to integrate these technologies with various viewpoints, such as those from the social sciences, ethics, and policymaking. On the other hand, most research concentrates only on technical concerns, ignoring the wider social ramifications and prospects for cooperation across disciplines.

Limited Consideration of Environmental Effects: As the IoT and deep learning technologies become more widely used, there is a rising concern about the effect these technologies have on the environment, especially in terms of the amount of energy they use and the amount of electronic trash they produce. Nevertheless, the majority of the studies that were examined here fail to take these environmental factors into account.

3 Literature review

Only the most recent five years' published research on deep transfer learning for various tasks and data sources were considered for inclusion in our list. A list of chosen works from dozens of examined works is shown in Table 1, arranged according to the DTL techniques used in each study. The following is a list of the inclusion criteria that we used throughout the selection process: (1) it has been published within the last five years; (2) it is repeatable (in terms of detailed implementation and models); (3) it has been applied to actual ML situations; and (4) it is generalizable.

To successfully identify and mitigate various information security risks in a Software-Defined Networking for Internet of Things (SDN-IoT) environment, Lahlou, Sara, et al. [22] suggested a lightweight, safe Threat Detection (TD) and Rule Automation (RA) architecture called "TD-RA" in 2022. To detect threats to the Internet of Things (IoT), the suggested solution includes a Policy-Enforcement Module (PEM) and BCM/MCM. Many ML algorithms have been used and evaluated to address the categorization issues.

In 2022, Mazaed Alotaibi et al. [23]. The multifaceted Deep Generative Adversarial Networks Model (MDGAN) was created as part of this effort to identify malicious software on mobile phones and tablets. To accurately depict Android Package Kit (APK) archives, the grayscale and API sequence of the hybrid Google Net and Long Short-Term Memory (LSTM) attributes have been processed pixel-by-pixel using conditioned Generative Adversarial Network (GAN). The generator of words for distinction in the

discrimination network produces syntactic harmful traits. According to the results of validation tests performed on the amalgamated Andro Zoo and Drebin databases, the accuracy was 0.962, and the F-score was 0.947. These results continue to be superior to the systems that were previously disclosed.

Benazzouza, Salma, et al.[24] presented a novel method that relies on two ML results in 2022. As a first step, they offer a stacked model-based fake user recognition system that applies two novel approaches: an ensemble ML system for user categorization and a method of authentication that utilizes disorganized compressive sensing for gathering features with few measures. As a second option, they provide a unique DL method for main user spectrum categorization that takes scalogram pictures as inputs.

In 2022, by Shafiq, Unsub et al. [25], Internet of Things (IoT) devices infested with Bash lite and Mirai were compared. The researchers also included benign recordings that showed uninfected behavior in each dataset. On average, they found that the model's detection accuracy for Bash lite was 44.59% higher, while for Mirai, it was 9.52%. The greatest significant performance gains of 26.68% and 73.000% were seen when the Eco Bee thermometer anomalous model was evaluated on various devices before and after TL for the Bash lite and Mirai, etc. Further, compared to Bash lite, 47.31% and 58.27% of the time were saved by TL in Mirai. Using the CIC-IDS2017 dataset, they also tested a dependent anomalous model using flow-based network traffic records.

In 2023, David A. Bierbrauer et al.[26] used neural networks (NN) to show that TL can identify intrusions in contexts with low computing resources using raw network data. Their findings demonstrate that they can achieve over 96% accuracy on edge machines with only 5,000 training samples and a final training period of around 67 seconds by combining a re-trained random forest framework with a transfer 1-D CNN model.

In 2023, Debicha, Islam, et al.[27] evaluated the efficacy of employing several strategically placed antagonistic detectors instead of a solitary antagonistic sensor for systems to detect intrusions. The antagonistic detector was built utilizing efficient TL techniques. Existing state-of-the-art intrusion detection algorithms were put into action. After that, they use a predetermined set of evasive attacks to target such models. To identify these hostile assaults, they develop and deploy several hostile sensors based on TL, with each detector getting a portion of the data that passes through the intrusion detection system. Their combined conclusions show that using many detectors instead of just one in a parallel IDS architecture might make hostile traffic even more detectable.

In 2023, Lubeyd et al.[28] presented a new Multi-modal Deep Transfer Learning(MMDTL) framework to effectively identify attacks in SDN settings. This framework allows them to study a wide range of attack types. The MMDTL framework extensively uses several data modalities, such as analytics of user activity, system logs, and network traffic patterns. This framework's TL technique is its most important feature since it allows the integration of pre-trained models' insights, improving the accuracy of identifying attacks.

In 2023, Singh, Amardeep, et al.[29] introduced a modern hybrid approach called RANSOMNET+. It effectively tackles the difficult problem of malware categorization by combining CNNs with trained transformers. By fusing the best parts of both designs, RANSOMNET+ can capture hierarchy characteristics and local patterns, making it superior to previous models. They found that RANSOMNET+ has great capability. With a remarkable 0.995 precision, 98.5% recall, and 0.977 F1 score, the model achieved 0.99799 training accuracy and 99.1% testing accuracy. The loss metrics for RANSOMNET+ were very low throughout the testing and training phases. They compared our model to two state-of-the-art options: ResNet 50 and VGG 16. RANSOMNET+ performed better in recall, accuracy, precision, and F1 score than the other two models. The visual illustrations and comprehension analysis provided by RANSOMNET+ further shed light on the method's decision-making method. Integrating feature payments, outlier identification, and feature significance analysis increased the model's accessibility and applicability.

In 2023, Okey, Og Obuchi Daniel, et al.[30] introduced a TL intrusion detection system (IDS) built on the CNN framework. This IDS has shown outstanding performance in image categorization. For training on two particular datasets, namely CIC-IDS2017 and CSE-CICIDS2018, they used five CNN models that had already been pre-trained. These models were VGG16, VGG19, Inception, MobileNet, and EfficientNets. Before the training, they performed processing, unbalance correction, reducing dimensionality, and the transformation of the vector of features into pictures that were acceptable for the CNN design using the Quantile Converter. The model average strategy is used to construct an ELETL-IDS. The three algorithms that have shown the highest performance levels are InceptionV3, MobileNetV3Small, and EfficientNetV2B0. The assessment results indicate that the ELETL-IDS surpassed the current state-of-the-art approaches in each assessment parameter, achieving a perfect score in all metric categories, including accuracy, precision, recall, and F-score.

In 2023, Çavuşoğlu et al.[31] developed a novel DL model that relied on TL. This model was designed to identify and safeguard cloud systems from harmful assaults. A DTL-IDS has been created, and it creates 2D-prepared feature maps from network data. In the subsequent step, the feature maps are processed using the transmitted and adjusted convolutional regions of the DL model. This is done before the dense layer, recognizing and categorizing traffic data.

In 2023, Hazman, Chaimae, et al.[32] introduced the IDS-SIoEL malware detection system; it is designed for use in smart settings that rely on the Internet of Things and use ensemble learning. In most cases, the structure suggests a perfect detection of anomalies model that incorporates AdaBoost, various feature selection methods, Boruta, shared data, and

association. Applying GPU to the IoT-23, BoT-IoT, and Edge-IIoT datasets allowed us to assess the suggested model. With an ACC, recall, and accuracy of about 0.999, the method offers significant rating gains to current IDS.

In 2022, Shukla et al. [33] proposed a dynamic multi-population teaching–learning optimization algorithm called DMPTLBO to protect against malicious intruders in network systems.

In 2022, Abdulmajeedand Inam [34] said that dataset choice is exceptionally critical to guarantee that it matches the IDS requirements. The dataset structure can greatly influence the selection of the ML algorithm. Hence, metrics provide a numerical relation between the ML algorithm and specific datasets.

In 2022, Abdulmajeed, I.A., and Husien, I.M.[35] The use of state-of-the-art IDS datasets, such as CIC-IDS2017 and CSE-CIC-IDS2018, in developing and assessing IDS systems based on machine learning with a hybrid CNN-LSTM architecture is a topic of extensive research. The novel strategy involves combining the two datasets to produce a new dataset

According to the experimental testing, training using the mixed dataset produced better metric values than individual datasets. This was particularly true when doing the inter-datasets assessment, eliminating the generalization issue.

In 2022, Aljanabi, Yaser Issam, et al.[36] suggested that a blockchain architecture integrates smart contracts and Machine Learning (ML) technologies, offering fresh, optimal chances for effective DDoS mitigation solutions across various collaborative sectors. The deployment of still-existing distributed and public infrastructure to block IP addresses or even advertise white is another major benefit of this structure.

It can be used with additional defense mechanisms against DDoS attacks, eliminating the need for distribution mechanisms or specialized registries and facilitating the implementation of procedures across various domains.

In 2024, El Ghazi, Mariam, and Noura Aknin [37] presented a deep model based on LSTM that was enhanced with batch normalization. Then, Bayesian Optimization was used to tune the model's hyperparameters, and the model was assessed using the PAMAP2 public dataset. The model achieves performance parameters of 96.76%, 96.55%, 96.85%, and F1 score, respectively, for accuracy, precision, and recall, with an accuracy of 97.71%.

In 2022, Qader and colleagues[38] showcased that creating and training the neural network populations is needed to play the Dama board game effectively. The NEAT algorithm was put into practice. Different network sizes and input/output combinations are tested for the game to surpass the human level. This article aimed to create a neural network that can play Dama like humans or is near enough to teach many neural networks over several generations.

In 2023, Khaleefah et al.[39] developed a strong machine learning algorithm-based model to identify and mitigate botnet-based assaults in Internet of Things networks. The suggested model addresses the common security problem caused by malevolent bot activity. Using various machine learning techniques, such as logistic regression, K-Nearest Neighbor (KNN), support vector machine (SVM), and linear regression, the model was trained using the BoT-IoT dataset to maximize its performance.

Table 1: A concise comparison of how relevant research is compared

Ref.	Method	Model	Dataset	Contribution	Results
[25]	ThreatDetection and Rule Automation Framework	Random Forest and Decision Tree	SDN-IoT	Utilize the BCM/MCM system to identify threats in IoT devices and employ a PEM system to mitigate attacks.	Accuracy for: RF=0.911 D.T.=0.987
[28]	Autoencoder-based Anomaly Detection	Autoencoder.	Mirai, Bash lite, CIC-IDS2017	Transferability of a trained automatic encoder system across comparable and dissimilar devices	Accuracy=0.999
[26]	TL	1-D-CNN	Raw network traffic5〖x10〗^3 samples	TL can identify intrusions using raw network traffic in cognitively restricted circumstances.	Accuracy = 0.96
[27]	TL-based Adversarial Detector	Multiple transfer learning (MTL)-based detectors	ID data	Demonstration showing a parallel IDS system with several sensors may identify hostile traffic better than a single sensor.	Detected rates of 0.717 and 0.741

[28]	Multi-modal Deep Transfer Learning (MMDTL)	DTL model	CIC-IDS2017	Create an MMDTL framework for the identification of attacks by SDN.	Accuracy=0.9997
[29]	RANSOMNET	Hybrid model combining CNNs with pre-trained transformers	Cloud-encrypted data	Using RANSOMNET+ and pre-trained converters to classify attacks by ransomware effectively.	Precision = 0.995 Recall = 0.98 F1 score = 0.977
[30]	TL-IDS	Pre-trained CNN models (VGG16, VGG19, Inception, Mobile Net, Efficient Nets)	CIC IDS2017, CSE-CICIDS2018	CAN Transfer detects CAN bus TL using a Convolutional LSTM algorithm.	Precision =0.88 Recall =0.89 F1-score =0.953
[31]	DTL-IDS	TL with CNN	NSL-KDD	NSL-KDD	Accuracy=0.9985
[32]	Ensemble Learning-based IDS	AdaBoost	IoT-23, BoTIoT, and Edge-IIoT	The framework presents an ideal discovery of anomalies model employing AdaBoost, choosing features, Boruta, shared data, and correlation.	Accuracy=0.999 Precisio =0.999 Recall=1.0 F1 score=1.0 AUC=1.0
[33]	Optimization based on dynamic multi-population teaching and	DMPTLBO	BoT-IoT and UNSW-NB15	Dynamic sub-population learning and intentional detection increase network security by improving accuracy, detection rate, false alarm rate, and computing economy.	N/A
[35]	Hybrid CNN-LSTM	Machine Learning-based IDS Systems	CIC-IDS2017, CSE-CIC-IDS2018	They are designing and evaluating machine learning-based IDS systems using hybrid CNN-LSTM architecture.	Accuracy=0.9889 Precision=0.989 Recall=0.9889 F1-score=0.9889 AUC=0.999
[36]	Blockchain Design	Smart Contracts and Machine Learning	N/A	Presents a blockchain concept using smart contracts and machine learning to mitigate DDoS in cooperative domains.	N/A
[37]	LSTM-based Deep Model with Batch Normalization	N/A	PAMAP2 Public Dataset	A new deep learning-based HAR method leveraging smart home wearable sensors	Accuracy=0.9771 Precision=0.9685 Recall=0.9655 F1-score=0.9666
[38]	NeuroEvolution of Augmenting Topologies (NEAT)	ANNs	Dama Board Game	Develops and trains neural networks for efficient Dama gameplay using a NEAT algorithm.	AI5 achieved the highest winning rate of 0.8125
[39]	ML Algorithm-Based Model	Liner R, Logistic R, KNN and SVM	BoT-IoT Dataset	Creates a machine learning algorithm-based model to detect and mitigate IoT botnet assaults	Accuracy for: Liner R= 0.978 Logistic R=0.977 KNN= 0.983 SVM= 0.978

4 Constraints and challenges

In the context of IoT networks, edge computing, power systems, and malware detection, the extracts show the significance of transfer learning approaches in tackling a variety of limits and issues in the field of cybersecurity. These strategies are designed to enhance detection accuracy, resource use efficiency, flexibility in dynamic situations, and durability against assaults from adversaries. Engaging in multidisciplinary initiatives that include specialists in cybersecurity, ML, and network engineering and developments in algorithmic approaches and infrastructural capabilities is necessary to address these limitations and difficulties.

In this review, several different strategies and frameworks that are targeted at resolving cybersecurity concerns in the context of IoT and other network settings are discussed. Based on the extracts that were supplied, the following explanation is presented, which includes restrictions and challenges:

Data availability: The difficulty of acquiring sufficient labeled data for training effective intrusion detection systems (IDS) or malware detection models is discussed in several articles. For example, Lahlou, Sara, et al. [25] seek to solve the problem of restricted data availability to train deep neural network-based intrusion detection techniques for Controller Area Network (CAN) bus systems.

Data Imbalance: The training of correct models is made much more difficult by imbalanced datasets.

Several articles, including Mazaed Alotaibi [26] and Bierbrauer, David A. et al. [29], discuss preprocessing approaches and feature selection methods as potential solutions to the class imbalance problem. Because the number of normal instances often exceeds the number of intrusion instances, this imbalance problem is frequently

seen in datasets used for intrusion detection, Abdulmajeed. et al. [34].

Labelling Constraints: A great number of DL-based intrusion detection algorithms are dependent on labeled data, which may be difficult to gather and can be both costly and time-consuming. This limitation may make it more difficult for typical deep learning approaches to be successful in constantly expanding Internet of Things attack situations.

Resource Constraints in Edge Networks: Conventional DL methods for intrusion detection often call for a substantial amount of computing resources, which may not be accessible in situations involving edge networks. This constraint underscores the need for distributed and resource-efficient models to provide successful intrusion detection in contexts that need edge computing and the Internet of Things.

Model Generalization: Obtaining generalizations of a model over a wide variety of situations continues to be difficult. Transfer learning is a technique that has been presented in several studies (Lahlou, Sara et al. [25], Mazaed Alotaibi [26], Benazzouza, Salma, et al. [37], Shafiq, Unsub, et al. [28], Bierbrauer, David A., et al. [29], Debicha et al. [30], and Elubeyd [31]) to adapt models that have been trained in one domain to carry out well in another domain with less labeled data. Nevertheless, achieving a successful transfer learning across various datasets and contexts is not s-simple[40][41].

5 Discussion

Driven by the digital revolution that the IoT is spurring, the rapid integration of Internet of Things devices into numerous facets of everyday life underscores the essential need to guarantee the security of these devices and networks. While recent advancements have improved efficiency and convenience, they have also sparked concerns about cybersecurity due to vulnerabilities in Internet of Things networks. Various cybersecurity threats, including network attacks utilizing machine learning and learning techniques and emerging risks like ransomware and darknet activities, have emerged with the rapid progress and adoption of IoT technologies. Traditional cybersecurity measures are no longer adequate to address the changing threat landscape. Therefore, incorporating learning methods is crucial to enhancing detection and prevention capabilities. Despite challenges such as a lack of labeled data resources on edge devices and entities, leveraging trained models and datasets for transfer learning can help overcome these hurdles. This strategy enables better resource utilization and performance by adapting detection systems to attack scenarios. Research findings indicate that transfer learning-based approaches significantly improve accuracy and effectiveness in network intrusion detection systems.

This underscores the importance of researching transfer learning techniques, developing defense strategies against attacks, and addressing privacy concerns related to data exchange in evolving IoT frameworks like 5G networks and smart cities. The discussion section provides an overview of the status of research in security matters. It also underscores the significance of transfer learning in addressing cybersecurity issues in environments. This is achieved by summarizing five themes.

6 Conclusion

In this study, we've introduced a deep transfer learning (DTL) framework to detect attacks tackling the crucial cybersecurity risks in IoT. The need for strong security measures is growing in importance as the IoT is infiltrating more and more parts of our lives. Our study explores the difficulties of conventional DL methods, namely how they rely on massive labeled datasets, which may be difficult, if not impossible, to acquire in today's dynamic threat environments. This review shows how DTL may use knowledge transfer from various activities and datasets to tackle these problems. To improve attack detection models for IoT frameworks, we classified DTL techniques into four categories: example-based, mapping-based, network-based, and adversarial-based. Each of these approaches has its own set of benefits. Lack of real-world validation, privacy and ethical implications considered, scalability, variety of datasets, multidisciplinary cooperation, and environmental effects are just a few of the holes in current research that we have uncovered. If we want to build complete and effective solutions for IoT security, we must fill these gaps. Researchers and industry professionals interested in how deep learning and Internet of Things security interact may use our work as a roadmap moving forward. We hope that by illuminating important research difficulties and suggesting future research pathways, we may help improve security measures in IoT networks and create a digital ecosystem that is safer and more robust for everyone.

References

[1] K. A. Shukla, S. Ahamad, G. N. Rao, A. J. Al-Asadi, A. Gupta, and M. Kumbhkar, “Artificial Intelligence Assisted IoT Data Intrusion Detection,” in *2021 4th International Conference on Computing and Communications Technologies (ICCCT)*, IEEE, Dec. 2021, pp. 330–335. doi: 10.1109/ICCCT53315.2021.9711795.

[2] A. D. Boursianis *et al.*, “Internet of Things (IoT) and Agricultural Unmanned Aerial Vehicles (UAVs) in smart farming: A comprehensive review,” *Internet of Things*, vol. 18, p. 100187, May 2022, doi: 10.1016/j.iot.2020.100187.

[3] J. Wang, M. K. Lim, C. Wang, and M.-L. Tseng, “The evolution of the Internet of Things (IoT) over the past 20 years,” *Comput. Ind. Eng.*, vol. 155, p. 107174, May 2021, doi: 10.1016/j.cie.2021.107174.

[4] Y. Hajjaji, W. Boulila, I. R. Farah, I. Romdhani, and A. Hussain, “Big data and IoT-based

applications in smart environments: A systematic review," *Comput. Sci. Rev.*, vol. 39, p. 100318, Feb. 2021, doi: 10.1016/j.cosrev.2020.100318.

[5] A. Nagaraj, *Introduction to sensors in IoT and cloud computing applications*. Bentham Science Publishers, 2021.

[6] T. I. Gritskevich, M. G. Leukhova, P. V. Gagin, K. K. Kalichkin, and N. S. Yakimova, "Introduction of IoT Solutions to Business Processes at Locomotive Enterprises: Efficiency and Transformation of Social Communications," 2022, pp. 875–883. doi: 10.1007/978-981-16-8829-4_86.

[7] M. Abdulraheem, J. B. Awotunde, R. G. Jimoh, and I. D. Oladipo, "An Efficient Lightweight Cryptographic Algorithm for IoT Security," 2021, pp. 444–456. doi: 10.1007/978-3-030-69143-1_34.

[8] K. C. Ravikumar, P. Chiranjeevi, N. Manikanda Devarajan, C. Kaur, and A. I. Taloba, "Challenges in the internet of things towards the security using deep learning techniques," *Meas. Sensors*, vol. 24, p. 100473, Dec. 2022, doi: 10.1016/j.measen.2022.100473.

[9] M. Steidl, M. Felderer, and R. Ramler, "The pipeline for the continuous development of artificial intelligence models—Current state of research and practice," *J. Syst. Softw.*, vol. 199, p. 111615, May 2023, doi: 10.1016/j.jss.2023.111615.

[10] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *J. Big Data*, vol. 9, no. 1, p. 102, Oct. 2022, doi: 10.1186/s40537-022-00652-w.

[11] A. W. Fazil, M. Hakimi, R. Akbari, M. M. Quchi, and K. Q. Khaliqyar, "Comparative Analysis of Machine Learning Models for Data Classification: An In-Depth Exploration," *J. Comput. Sci. Technol. Stud.*, vol. 5, no. 4, pp. 160–168, 2023.

[12] T. V. Khoa *et al.*, "Collaborative learning for cyberattack detection in blockchain networks," *arXiv Prepr. arXiv2203.11076*, 2022.

[13] H. Zhong, S. Yu, H. Trinh, Y. Lv, R. Yuan, and Y. Wang, "Fine-tuning transfer learning based on DCGAN integrated with self-attention and spectral normalization for bearing fault diagnosis," *Measurement*, vol. 210, p. 112421, Mar. 2023, doi: 10.1016/j.measurement.2022.112421.

[14] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *J. Biomol. Struct. Dyn.*, vol. 39, no. 15, pp. 5682–5689, Oct. 2021, doi: 10.1080/07391102.2020.1788642.

[15] E. Rodríguez *et al.*, "Transfer-Learning-Based Intrusion Detection Framework in IoT Networks," *Sensors*, vol. 22, no. 15, p. 5621, Jul. 2022, doi: 10.3390/s22155621.

[16] Y. Gao, "Author Name Disambiguation Using Co-training." University of Windsor (Canada), 2020.

[17] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, "Federated learning for the Internet of things: Applications, challenges, and opportunities," *IEEE Internet Things Mag.*, vol. 5, no. 1, pp. 24–29, 2022.

[18] Y. Wang *et al.*, "An exhaustive research on the application of intrusion detection technology in computer network security in sensor networks," *J. Sensors*, vol. 2021, pp. 1–11, 2021.

[19] S. Alosaimi and S. M. Almutairi, "An intrusion detection system using BoT-IoT," *Appl. Sci.*, vol. 13, no. 9, p. 5427, 2023.

[20] A. M. Banaamah and I. Ahmad, "Intrusion detection in IoT using deep learning," *Sensors*, vol. 22, no. 21, p. 8417, 2022.

[21] L. Vu, Q. U. Nguyen, D. N. Nguyen, D. T. Hoang, and E. Dutkiewicz, "Deep transfer learning for IoT attack detection," *IEEE Access*, vol. 8, pp. 107335–107344, 2020.

[22] S. Lahlou, Y. Moukafih, A. Sebbar, K. Zkik, M. Boulmalf, and M. Ghogho, "TD-RA policy-enforcement framework for an SDN-based IoT architecture," *J. Netw. Comput. Appl.*, vol. 204, p. 103390, 2022.

[23] F. Mazaed Alotaibi and Fawad, "A Multifaceted Deep Generative Adversarial Networks Model for Mobile Malware Detection," *Appl. Sci.*, vol. 12, no. 19, p. 9403, Sep. 2022, doi: 10.3390/app12199403.

[24] S. Benazzouza, M. Ridouani, F. Salahdine, and A. Hayar, "A Novel Prediction Model for Malicious Users Detection and Spectrum Sensing Based on Stacking and Deep Learning," *Sensors*, vol. 22, no. 17, p. 6477, Aug. 2022, doi: 10.3390/s22176477.

[25] U. Shafiq, M. K. Shahzad, M. Anwar, Q. Shaheen, M. Shiraz, and A. Gani, "Transfer Learning Auto-Encoder Neural Networks for Anomaly Detection of DDoS Generating IoT Devices," *Secur. Commun. Networks*, vol. 2022, pp. 1–13, May 2022, doi 10.1155/2022/8221351.

[26] H. Elubeyd, D. Yiltas-Kaplan, and Ş. Bahtiyar, "A Multi-Modal Deep Transfer Learning Framework for Attack Detection in Software-Defined Networks," *IEEE Access*, vol. 11, pp. 114128–114145, 2023, doi: 10.1109/ACCESS.2023.3324878.

[27] D. A. Bierbrauer, M. J. De Lucia, K. Reddy, P. Maxwell, and N. D. Bastian, "Transfer learning for raw network traffic detection," *Expert Syst. Appl.*, vol. 211, p. 118641, Jan. 2023, doi: 10.1016/j.eswa.2022.118641.

[28] I. Debicha, R. Bauwens, T. Debatty, J.-M. Dricot, T. Kenaza, and W. Mees, "TAD: Transfer learning-based multi-adversarial

detection of evasion attacks against network intrusion detection systems," *Futur. Gener. Comput. Syst.*, vol. 138, pp. 185–197, Jan. 2023, doi: 10.1016/j.future.2022.08.011.

[29] A. Singh, Z. Mushtaq, H. A. Abosaq, S. N. F. Mursal, M. Irfan, and G. Nowakowski, "Enhancing Ransomware Attack Detection Using Transfer Learning and Deep Learning Ensemble Models on Cloud-Encrypted Data," *Electronics*, vol. 12, no. 18, p. 3899, Sep. 2023, doi: 10.3390/electronics12183899.

[30] O. D. Okey, D. C. Melgarejo, M. Saadi, R. L. Rosa, J. H. Kleinschmidt, and D. Z. Rodriguez, "Transfer Learning Approach to IDS on Cloud IoT Devices Using Optimized CNN," *IEEE Access*, vol. 11, pp. 1023–1038, 2023, doi: 10.1109/ACCESS.2022.3233775.

[31] Ü. Çavuşoğlu, D. Akgun, and S. Hizal, "A Novel Cyber Security Model Using Deep Transfer Learning," *Arab. J. Sci. Eng.*, vol. 49, no. 3, pp. 3623–3632, Mar. 2024, doi: 10.1007/s13369-023-08092-1.

[32] C. Hazman, A. Guezzaz, S. Benkirane, and M. Azrour, "lIDS-SIoEL: intrusion detection framework for IoT-based smart environments security using ensemble learning," *Cluster Comput.*, vol. 26, no. 6, pp. 4069–4083, Dec. 2023, doi: 10.1007/s10586-022-03810-0.

[33] Shukla, Alok Kumar, and Shubhra Dwivedi. "Discovery of Botnet activities in Internet-of-Things system using dynamic evolutionary mechanism." *New Generation Computing* 40.1 (2022): 255-283.

[34] I. A. Abdulmajeed and I. M. Husien, "Machine Learning Algorithms and Datasets for Modern IDS Design," in *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, IEEE, Jun. 2022, pp. 335–340. doi 10.1109/CyberneticsCom55287.2022.9865255.

[35] I. A. Abdulmajeed and I. M. Husien, "MLIDS22- IDS Design by Applying Hybrid CNN-LSTM model on Mixed-Datasets," *Informatica*, vol. 46, no. 8, Nov. 2022, doi: 10.31449/inf.v46i8.4348.

[36] Y. I. Aljanabi, A. A. Majeed, K. H. Jihad, and B. A. Qader, "Detect and Mitigate Blockchain-Based DDoS Attacks Using Machine Learning and Smart Contracts," *Informatica*, vol. 46, no. 7, Oct. 2022, doi: 10.31449/inf.v46i7.4033.

[37] M. El Ghazi and N. Aknin, "Optimizing Deep LSTM Model through Hyperparameter Tuning for Sensor-Based Human Activity Recognition in Smart Home," *Informatica*, vol. 47, no. 10, Jan. 2024, doi: 10.31449/inf.v47i10.5268.

[38] B. A. Qader, K. H. Jihad, and M. R. Baker, "Evolving and training of Neural Network to Play DAMA Board Game Using NEAT Algorithm," *Informatica*, vol. 46, no. 5, Mar. 2022, doi: 10.31449/inf.v46i5.3897.

[39] A. D. Khaleefah and H. M. Al-Mashhadi, "Detection of IoT Botnet Cyber Attacks using Machine Learning," *Informatica*, vol. 47, no. 6, May 2023, doi: 10.31449/inf.v47i6.4668.

[40] Mustafa, Duaa Haider, and Idress Mohammed Husien. "Adaptive DBSCAN with Grey Wolf Optimizer for Botnet Detection." International Journal of Intelligent Engineering & Systems 16.4 (2023). DOI: 10.22266/ijies2023.0831.33

[41] Ahmed, Mohammed, and Idress Hussein. "Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review." Journal of Robotics and Control (JRC) 5.3 (2024): 884-892. DOI: 10.18196/jrc.v5i3.21606.