Received 7 April 2025; revised 24 April 2025; accepted 6 May 2025. Date of publication 15 May 2025; date of current version 23 May 2025.

Digital Object Identifier 10.1109/OJCOMS.2025.3569301

Deep Transfer Learning Based on Hybrid Swin Transformers With LSTM for Intrusion Detection Systems in IoT Environment

IBRAHIM A. FARES [1], AHMED GAMAL ABDELLATIF IBRAHIM [2], MOHAMED ABD ELAZIZ [1], MANSOUR SHRAHILI [3], ADHAM AHMED ELMAHALLAWY [4], RANA MUHAMMAD SOHAIB [5], MAHMOUD A. SHAWKY [2,6], AND SYED TARIQ SHAH [7] (Member, IEEE)

[1]Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt
[2]Department of Electronics and Electrical Engineering, ADC EMA, Cairo 11841, Egypt
[3]Department of Statistics and Operations Research, College of Science, King Saud University, Riyadh 11451, Saudi Arabia
[4]Higher Institute of Engineering and Technology, King Mariout, Alexandria 23713, Egypt
[5]Department of Computer and Information Science, Northumbria University, NE1 8ST Newcastle upon Tyne, U.K.
[6]James Watt School of Engineering, University of Glasgow, G12 8QQ Glasgow, U.K.
[7]School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K.

Corresponding author: S. T. SHAH (e-mail: syed.shah@essex.ac.uk)

This work was supported by the Distinguished Scientist Fellowship Program (DSFP) at King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Extensive growth in the number of Internet Of Things (IoT) devices has significantly increased susceptibility to various cyber-attacks and hence emphasized the need for robust intrusion detection systems (IDS) for ensuring IoT network security. While deep learning (DL) methodologies have proven effective in the application of IDS, their success greatly depends on the availability of large datasets and significant computational resources during training. To overcome the limitations associated with this dependence on large datasets and significant computational capacity for training, the current work suggests employing the transfer learning (TL) mechanism by combining Swin Transformers with long short-term memory (LSTM) networks. Utilizing the beneficial properties of Swin Transformers in learning hierarchically structured data combined with the proficiency of LSTM in processing sequential dependencies, the hybrid model generates pre-trained weights in the first phase. These pre-trained weights are further transferred into another instance of the new model for subsequent fine-tuning. Experiments are carried out on several benchmarking datasets, namely NSL-KDD, ToN-IoT, BoTIoT, MQTTIoT, and CICIoT2023, which include both binary and multi-class classification scenarios. The proposed model outperforms state-of-the-art DL models, for example, the Autoencoders, ResNets, CNN, RNN, and LSTM models, and achieved an average of 98.97% in accuracy, of 98.97% in precision, of 99.02% in recall, of 98.97% in F1 score, across all datasets. Experimental results establish that the hybrid approach achieves better detection accuracy and better performance measures compared to the latest state-of-the-art methods, thus proving itself effective in increasing the scalability and adaptability of IDS in IoT.

INDEX TERMS Cyber-security, intrusion detection system (IDS), IoT, transformers, transfer learning.

I. INTRODUCTION

THE RAPID expansion of inter-connected devices originating from diverse domains such as healthcare, transportation, and urban infrastructures draws urgent attention to the need to secure the IoT ecosystem [1]. Intrusion detection system (IDS) in IoT networks encompasses the spectrum of identification to addressing unauthorized access to the network [1]. However, IoT devices are often characterized by limited computational power and are connected to publicly accessible networks; thus, they are susceptible to cyber-attacks. Accordingly, security and integrity stand out as critical security services in protecting the IoT network. These could be accomplished using IDS, which plays a vital role in controlling unauthorized access to protect sensitive data from possible threats [3].

As stated by [4], IoT systems, in general, create a huge volume of data that comes from different types of devices in heterogeneous formats. It is demanding to analyze the anomaly and trend that may indicate intrusion. Moreover, complex security schemes or resource-intensive algorithms of IDS are not possible to deploy on IoT devices due to their limited resources [5], [36]. These limitations have led to different IDS methods applied in IoT based on both standard machine learning (ML) as well as advanced techniques of deep learning (DL) algorithms, each with its merits and deficiencies. ML techniques such as decision tree (DT), support vector machines (SVMs), and K-nearest neighbors (KNN) have been used exclusively for IoT-based IDS [6], [51]. While these methods are already established and considered relatively easy to deploy, they may fail in efficiently handling large and complex datasets as are often received in IoT systems. To this end, DL methods have been employed to address these limitations [30]. The use of DL-based approaches for designing reliable and efficient IDSs can be applied to IoT network design. Among these, the DL algorithms, such as convolutional neural network (CNN) [7] and long short-term memory (LSTM) [8] learn sophisticated patterns and features from large datasets. These make them a potential technology that can enhance the scalability of the network [9]. The capability to learn complex patterns also makes the algorithm capable of identifying complex anomalies from a large dataset with promising results. However, most DL models require a great deal of training on large datasets usually at very high computational costs.

IDS integrated with transformer models may lead to improved detection accuracy and efficiency as transformer models are superior in learning complex interrelationships from data. However, DL techniques might suffer while handling that complexity [10]. In fact, transformers are one of the conceptual changes in IDS, which made it easier to extract features from data by minimizing the need for manual feature engineering [11]. Also, their resistance to a set of various attack scenarios further enhances the robustness of systems. However, their computational overhead still needs to be balanced against real-time processing. In addition, decisions by transformers need to be explainable; this is a critical factor for building trust and allowing transformers to integrate into security-critical systems. Integration of transformers into DL methods may lead to a fundamental rethink in terms of how IDS is defined and may offer effective, interpretable, adaptive methodologies for threat assessment in network security.

Recently, authors in [10] introduced a hybrid transformer deep learning (TDL) model that relies on transfer learning (TL) techniques, which enables it to discern patterns and irregularities indicative of potential intrusions. By fine-tuning a pre-trained TDL model on a specific dataset or task related to IDS in IoT, the system can enhance both the precision and speed of the IDS process. This study proposes the integration of the LSTM-Swin transformer, the Swin Transformer was recently developed by a Microsoft team in 2021 [13], [14]. Swin Transformer incorporates a shifted window mechanism and hierarchical structure, this enables it to adeptly capture both local and global patterns within data, thereby enhancing the system's capability for anomaly detection and intrusion identification (ID). Also, the LSTM addresses the vanishing gradient problem associated with traditional RNNs. It facilitates the capture of long-term dependencies in sequential data [15]. LSTMs have found extensive application in IDS due to their effectiveness in leveraging temporal dynamics within data, leading to superior detection of anomalous activities and intrusions.

To reduce the training time required for the proposed hybrid model, we employed a TL mechanism. TL enables us to influence pre-trained models by benefiting from the knowledge achieved from previous training on large datasets and improving the IDS's efficiency to new datasets [16]. The contributions of this work can be summarized as follows:

- This paper introduces an IDS architecture integrating LSTM and Swin Transformers, offering an alternative solution to traditional IDS architectures.
- Using TL, the model is trained on NSL-KDD [17], then tested on ToN-IoT [18], BoTIoT [19], MQTTIoT [20], and CICIoT2023 [21].
- Fine-tuning involves adjusting hyperparameters and regularization, with performance evaluated using various metrics.
- Comparisons analysis are also conducted against prominent state-of-the-art DL models, such as autoencoders, residual networks (ResNets), CNNs, recurrent neural networks (RNNs), and conventional LSTM-based approaches to confirm its competitive results.

II. RELATED WORK

Recently, IDS in IoT has attracted much attention since more and more sectors depend on the connectivity of devices, hence their potential risks and consequences of successful attacks also increase. There are some research has surveyed this area, underlining all the possible challenges and possible solutions for securing IoT systems. For example, Akyildiz et al. [22] include a review of some of the most important challenges and possible solutions on how to secure wireless sensor networks, including IDS. The authors identified several challenges and discussed limitations that were associated with conventional rule-based methods and anomaly-based methods. In the process, a new approach was proposed, which includes the fusion of both methods, emphasizing hybrid methods. The study also emphasized the need for improved feature selection and feature engineering methods to reinforce the performance of IDS in wireless sensor networks.

Authors in [23] present a state-of-the-art review of IDS in IoT based on associated security issues. The authors presented that the IoT nodes may be vulnerable to different cyber-attack types due to their constrained computational capability and communication. Their review establishes that there have been massive developments around IoT-based

IDS, but at the same time, it also brings out some challenges, which are continuous and some of the most promising future research directions. An example of this challenge, the complexity and heterogeneity in IoT systems, lack of sufficient labeled training data, the need for fast-running, and scalable algorithms.

The study conducted in [9] investigates the application of DL techniques for IDS in IoT environments, leveraging CNNs for feature extraction and LSTM networks for classification. The performance of the system was evaluated using the NSL-KDD and CICIDS2017 datasets showing increased accuracy with reduced false positive rates as compared to traditional ML methods. However, limitations include the labelled data made available, scarce in quantity, and challenges related to model generalization across diverse IoT environments. Muhammad et al. [24] proposed a DL-based IoT IDS, which relies on the strengths of the stacked autoencoder to learn features from data. The proposed system is effective in detecting various intrusion types in IoT systems. Thus, it was compared to traditional ML techniques concerning performance. Bhawana et al. [25] introduced a filter-based feature selection model integrated with a deep neural network, demonstrating high efficiency and practicality as a robust solution for safeguarding IoT networks against various attack types. The proposed model shows significant promise for datasets with imbalanced class labels, addressing this issue by utilising generative adversarial networks (GANs) to augment the number of packets in minority attack classes. However, the model has notable limitations, they achieved 90.9% of accuracy which considered low value, its dependency on large datasets and the potential for generating false alarms.

TL is thus one promising solution to tackle the challenge of low IDSs rates. In prior work, TL with CNN models has been conducted in a two-step process [26]. The model is first trained on a base dataset, MQTTIoT [27], and then knowledge learned will be transferred into the target dataset, NSL-KDD. The proposed system consists of two concatenated CNNs and was tested with the $KDDTest-21$ dataset in order to check its performance in identifying zero-day attacks. Their findings showed that the proposed system enhanced new attack detection by a rate of 2.86% compared to the traditional CNN methods to an accuracy of 81.94%.

Mehedi et al., [12] proposed a TL-based model for reducing the training time in IDS. Their model included two CNNs, trained with source and target datasets created using two subsets of a labeled dataset, developed for in-vehicle networks [28]. The mentioned dataset represented three types of attacks: flooding, fuzzing, and spoofing. Later, the results of this detection model turned out to be excellent and achieved 98.1% overall accuracy. Idrrissi et al. [31] presented their efforts to leverage TL to overcome most of the drawbacks of classic DL-based IDS methods, especially towards identifying new attacks within resource-constrained IoT environments with poorly available labeled data. They came up with the fine-tuning strategy of a pre-trained model by keeping most layers fixed and training only the last layers by using a CNN architecture. The authors used the source domain from the BoT-IoT dataset [32], which was a generic dataset for IoT systems, and enhanced it with a limited set of target domain data, specifically for IIoT acquired from the TON-IoT dataset. Surprisingly enough, the model was able to achieve an accuracy of 99% in identifying new attacks.

Transforming this further, therefore, has been a promising avenue in ID due to the adeptness of transformers in handling sequential data as provided by [33]. Using self-attention mechanisms, transformers have recorded an outstanding ability in capturing complex patterns and dependencies in network traffic sequences, which are important for the detection of anomalies. Because of the nature of this hierarchical architecture in intrusion data, it can support both short-term and long-term temporal relationships without any loss, hence enhancing precision in threat identification. An adaptive transformer model is developed for the detection of anomalies in WSN using spatio-temporal attention, as presented by the authors in [34]. The authors mention that the process of detecting anomalies in WSNs has become one of the major undertakings in recent times, particularly in applications like defense, health care, and home automation. Thus, the shortcomings of conventional models with respect to failing to consider spatio-temporal features in analysis can be overcome with the proposed model that incorporates spatio-temporal attention and an adaptive transformer model. Its strong point is the capability to learn and identify complex nonlinear relationships from the data. It has several advantages, including very high accuracy, real-time detection, and adaptability to different kinds of WSNs. Limitations that come with the model proposed herein include those involving high volumes of training data and the possibility of overfitting.

Recently, authors in [76] examines the application of deep transfer learning (DTL) in developing IDS for industrial control networks. It indicates the increasing connectivity of industrial control systems (ICS) to external networks, assuring the need for robust security measures against cyber threats. The study categorizes existing literature into DTL-only, IDS-only, and combined DTL and IDS approaches, providing outcomes into integrating DTL methods to enhance IDS performance in detecting novel attacks.

This comparative analysis (as summarized in Table 1) discovers challenges such as dataset limitations, computational resource constraints, and model generalization persist. The main difference between this work and the literature work lies in the introduction of a novel IDS model that integrates LSTM with Swin transformers based on TL techniques to increase scalability. The proposed model aims to utilize the unique strengths of both architectures to tackle the intricate and varied characteristics of IoT data effectively.

III. BACKGROUND

This section provides background information on LSTM, Swin Transformers, and TL techniques.

TABLE 1. Comparison of reviewed IDS approaches for IoT security.

Study	Methodology	Datasets Used	Key Findings
Akyildiz et al. [22]	Survey of challenges and solutions in securing WSN, hybrid IDS	Not applicable (survey)	Identified limitations of conventional IDS; emphasized improved feature selection and engineering.
Kumar et al. [23]	Survey of IDS in IoT, highlighting security issues and challenges.	Not applicable (survey)	Discussed complexity, and heterogeneity in IoT; emphasized the need for scalable algorithms.
Chawla et al. [9]	CNNs for feature extraction and LSTM for classification	NSL-KDD, CICIDS2017	increased accuracy, reduced false positive rates; limited labeled data.
Muhammad et al. [24]	stacked autoencoder-based IDS for IoT to detect various intrusion types.	KDDCup99, NSL-KDD, aegean Wi-Fi	effectiveness in detecting intrusions; compared favorably to traditional ML techniques.
Sharma et al. [25]	filter-based feature selection model integrated with DNN for IDS in IoT.	UNSW-NB15	high efficiency in safeguarding IoT networks; addressed imbalanced class labels using GANs.
Wu et al. [26]	Employed TL with CNNs, trained on the UNSW-NB15 dataset, and transferred knowledge to NSL-KDD dataset.	UNSW-NB15, NSL-KDD	Enhanced new attack detection by 2.86% compared to traditional CNN methods; achieved 81.94% accuracy.
Mehedi et al. [12]	TL-based model with two CNNs, trained on subsets of labeled datasets for in-vehicle networks.	Custom in-vehicle network datasets	Achieved 98.1% overall accuracy; reduced training time in IDS.
Idrissi et al. [31]	TL to update DL-based IDS for IoT, fine-tuning a pre-trained CNN model.	BoT-IoT, TON-IoT	Achieved 99% accuracy in identifying new attacks; addressed limited labeled data in IIoT environments.
Liu and Wu [33]	Adaptive transformer model using spatio-temporal attention for anomaly detection in WSNs.	NSL-KDD	Demonstrated high accuracy and real-time detection; noted limitations include high training data volume and potential overfitting.
Fares and Abd Elaziz [29]	Explainable Optimized TabNet for AIDS.	NSL-KDD, CICIoT2023, and RT_IoT2022	high accuracy with explainable results; complexity in implementation.
Kumar et al. [34]	Developed an adaptive transformer model using spatio-temporal attention for anomaly detection in WSNs.	SWaT	Demonstrated high accuracy and real-time detection; noted limitations include high training data volume and potential overfitting.
Fares et al. [73]	Transformer based on Kolmogorov–Arnold Networks	RT-IoT2022, IoT23, and CICIoT2023	high detection rate, lightweight model; high processing time.

A. LONG SHORT-TERM MEMORY (LSTM)

A variation on RNNs, LSTM networks were created specifically to capture long-term dependencies in sequential data; Hochreiter and Schmidhuber [39] first proposed this idea. This paper gives an overview of the major traits and capabilities inherent in LSTM networks, as well as their relevance, applications, and problems within the area of IDS in the context of the IoT. Artificial neural networks with a recurrent loop topology that are similar to RNNs but improved with extra gating mechanisms make up LSTMs. The function of these gates is to update and selectively save data while it's concealed. The input gate, forget gate, output gate, and cell state are the four essential components of an LSTM. The information that enters and exits the concealed states–which serve as archives for information preservation throughout time steps–is regulated by these gate mechanisms.

These fundamental equations provide a mathematical definition of the behavior of the LSTM cell [39]:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$
$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$
$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{3}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{4}$$
$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{5}$$
$$h_t = o_t \odot \tanh(C_t) \tag{6}$$

where $\odot$ means element-wise multiplication, tanh stands for the hyperbolic tangent function, σ for the sigmoid activation function, and W and b for the weight matrices and bias vectors, respectively. Every equation relates to a crucial facet of the functioning of the LSTM cell. To be more precise, f_t is the forget gate, i_t is the input gate, $\tilde{C}_t$ is the candidate cell

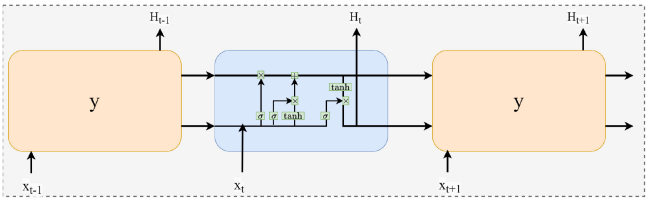


FIGURE 1. Flowchart of the LSTM model.

state, C_t is the updated cell state, o_t is the output gate, and h_t is the hidden state. These gates, in conjunction with their corresponding equations, provide the LSTM cell the ability to update and output data selectively, which in turn allows it to catch relevant patterns and correlations among sequential data. Figure 1 presents the flowchart of the LSTM model.

B. SWIN TRANSFORMERS

Transformers introduced a remarkable development in various fundamental DL domains such as targeted sequential data, such as natural language [13]. Applications of transformers to image data are hampered by significant computational complications due to the quadratic complexity relative to image resolution. Swin transformers introduce a novel architectural paradigm toward overcoming this limitation by an effective methodology for the processing of image data while paying appropriate attention to context information [13]. The key insight of Swin transformers is found in their hierarchical design-they allows for much more effective scaling to higher image resolutions. This hierarchy consists of dividing the image into non-overlapping patches, subsequent to which are successive hierarchical stages of processing. Probably one of the most salient features of the Swin transformer is the ingenious "shift operation". Conventionally, convolutional layers or mechanisms of attention make use of overlapping patches; this leads to considerable redundancy in computation. Such avoidance is done by the shift operation, using non-overlapping patches and aggregating the shifted local information in a manner that reduces computation by a large amount. It could be mathematically represented as [13]:

$$\mathbf{y}_{i,j} = \sum_{\Delta i, \Delta j} \mathbf{W}_{\Delta i, \Delta j} * \mathbf{x}_{i+\Delta i, j+\Delta j}, \tag{7}$$

where $\mathbf{x}_{i+\Delta i, j+\Delta j}$ indicates input features at the shifted location, $\mathbf{W}_{\Delta i, \Delta j}$ are learnable weights, and $\mathbf{y}_{i,j}$ represents the aggregated feature at position (i, j).

Additionally, tokenization and local self-attention methods are included in swin transformers. The input picture patches are projected onto query, key, and value spaces in these procedures, and then attention scores and weighted values are computed. In terms of math, this is expressed as:

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{X}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{X}\mathbf{W}_V, \tag{8}$$

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \tag{9}$$

where $\mathbf{X}$ represents the input tensor, $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are learnable weight matrices, d_k denotes the dimensionality of keys, and $\mathbf{A}$ signifies the attention output. The architectural diagram of the Swin transformers model is presented in Figure 2. The figure detailing the input processing via Patch Partition and the hierarchical structure composed of sequential Swin Transformer Blocks containing attention (W-MSA/SW-MSA) and MLP layers.

C. TRANSFER LEARNING

TL is a technique that enables ML models to use information from one domain to improve performance on a domain that is connected to a variation [42]. The notion is that, instead of starting from zero and building a new model, an existing model that has been prepared and pre-trained on large datasets may be fine-tuned for a new job using a smaller dataset. This can enhance performance and save a substantial amount of processing time and resources, especially when working with small amounts of data. Figure 3 displays the TL flowchart.

Probably the greatest advantage of TL is that the model learns general representations of data, that is, features that are going to be useful across different domains. For example, a model pre-trained on image classification can easily adapt for object detection or segmentation with some fine-tuning. In a similar vein, it would be pretty easy to fine tune a model that is already pre-trained on natural language processing for sentiment analysis. This is attributed to the fact that during pre-training, the features learned are often general and transfer across tasks rather than being induced from a certain task at hand [43].

TL can be used in the context of IDS on the IoT to enhance the performance of DL models [44]. Assuming a model is pre-trained first on a large network traffic dataset, fine-tuning could take place using a smaller dataset of IoT traffic for it to get used to learning features of particular patterns and anomalies indicative of probable intrusions.

IV. THE PROPOSED METHOD

The current section presents the hybrid LSTM-Swin transformer model based on TL mechanism for IDS in IoT. This strategic integration leverages the intrinsic strengths of the Swin transformer in the effective salient feature extraction in IoT data and promotes efficiency and performance in the model. For adapting Swin transformers from a pure 2D image processing framework to one-dimensional network traffic data for intrusion detection, a preliminary transformation is necessary. All the relevant 2D operations within Swin transformers need to be replaced by appropriate one-dimensional counterparts [45]. In other words, the 2D attention used in Swin transformers' building blocks will be substituted with the corresponding one-dimensional convolution operation. This conversion is essential for Swin transformers to be best integrated into the ID domain and tap their full strengths in performing one-dimensional data analysis. Swin transformers leverage meaningful features in pre-processed

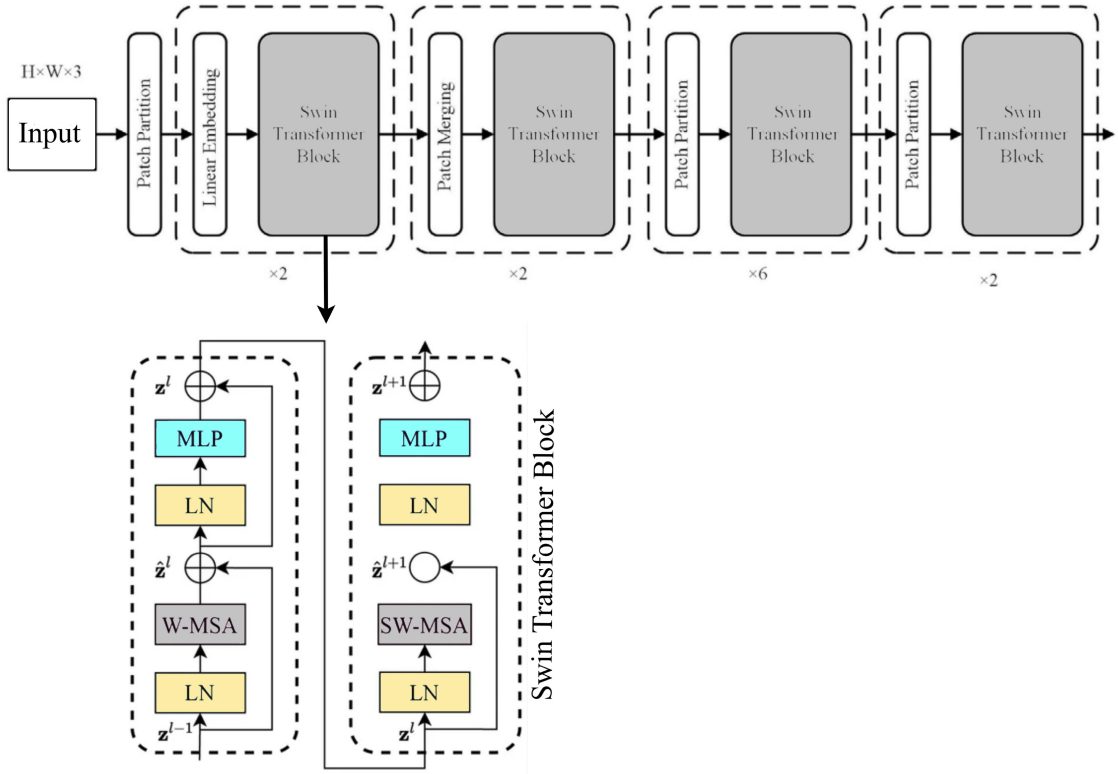


FIGURE 2. Architectural diagram of the Swin transformer model.

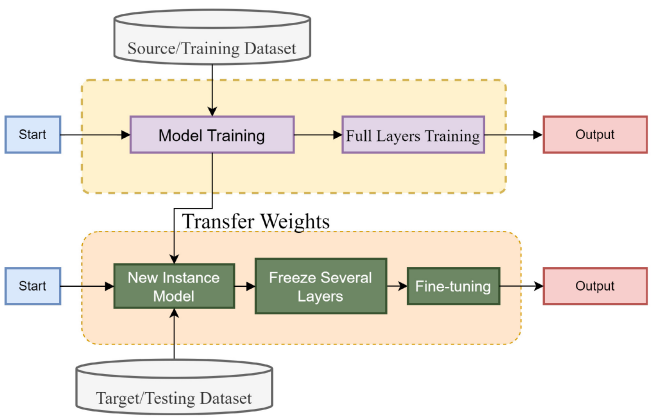


FIGURE 3. Flowchart of the transfer learning mechanism.

IoT data. The self-attention mechanisms in the transformer layers of Swin transformers dig deeper for complex patterns and characteristics that manifest intrusions. The global representation by Swin transformers is then passed on to the LSTM layers. Specifically, LSTM layers are designed to capture long-range dependencies in sequences. Given recurrent neurons with memory, LSTMs add loops to track information from past time steps so that informed predictions can be made for subsequent steps. Gate mechanisms within the LSTMs regulate information flow, thus making selective storage and updating over time possible.

In the proposed architecture, LSTM is placed strategically to capture temporal dependencies while modeling the likelihood of intrusion. The aggregate input to the LSTM includes features resulting from Swin Transformer along with other relevant sources. This aggregate input is fed sequentially to the LSTM, which, through internal hidden states, represents and updates information related to data. The output of the LSTM conveys intrusion predictions extracted from the learned phase. In general, the proposed model involves a few steps: First of all, the execution of the experiment is performed using the source dataset, which helps the model to learn basic information and calculate weights. These weights then get transferred by using the TL mechanism to a new instance (non-trained) model. This is an important step in maintaining continuity in knowledge learned across the change between different datasets. To maximize performance, fine-tuning is done in an attempt to upgrade the model's capabilities while simultaneously cutting down on the number of training parameters, thereby encouraging efficiency and generalization. In fine-tuning, for example, we reduce the number of trainable parameters from 1, 018, 243 at the training phase to 66, 307 to allow very fast and simple training in the target dataset to be able to gain foundational insights and patterns. This makes the model very fast. Finally, in the last phase, the model is tested on various datasets.

Figure 4 shows the general architecture of the proposed model intended for IDS in IoT. The model is trained

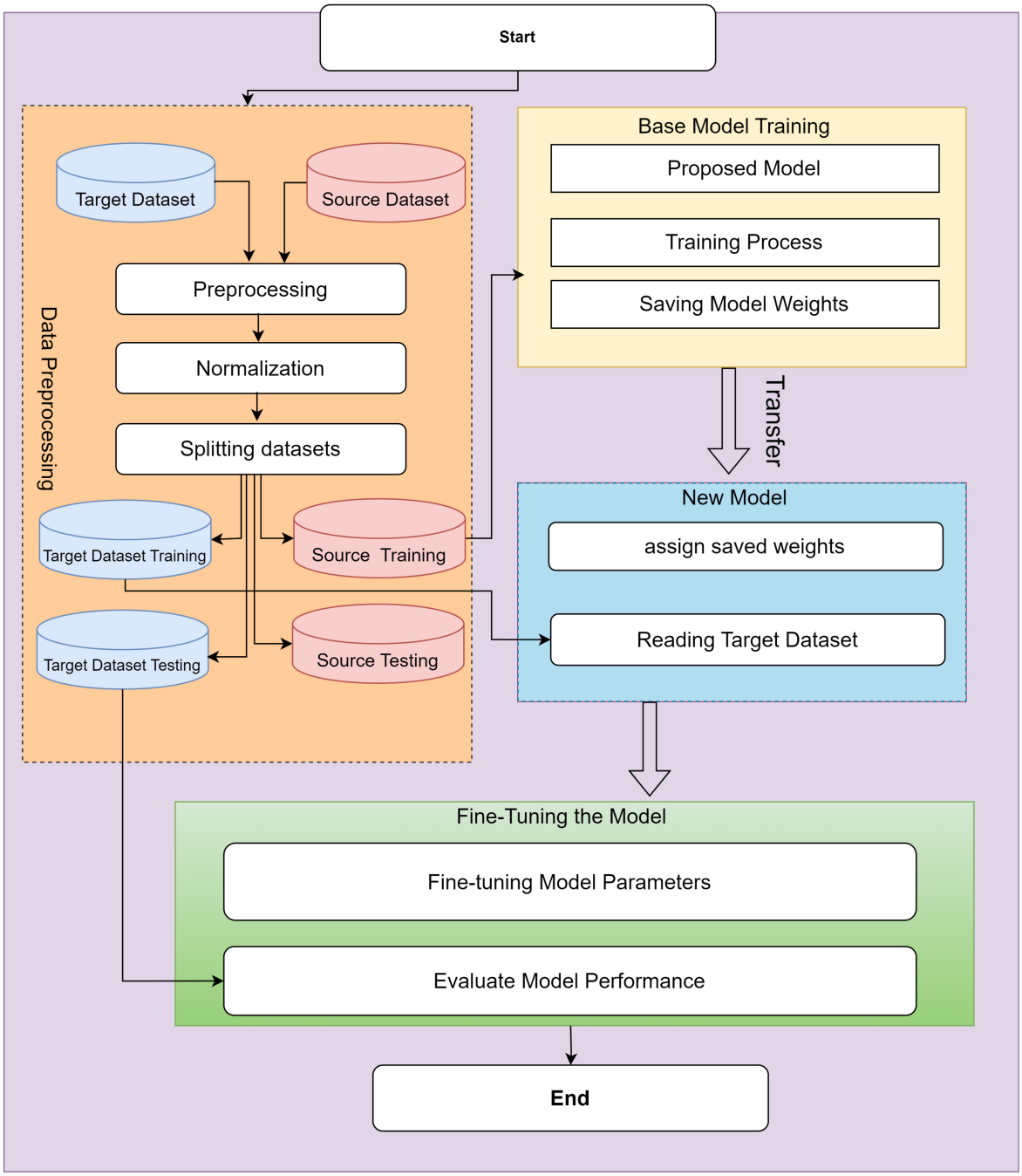


FIGURE 4. The architecture of the proposed model.

on the source dataset, and performance testing and evaluation are conducted on the target datasets. The following procedures can be used to clarify the flowchart:

- *Data preprocessing:* Raw data undergoes preprocessing, transforming it into a model-compatible format and splitting the source and target datasets into training and testing. This reduces the time needed for training processing.
- *Model training:* The processed data is directed through the proposed model to extract features relevant to ID. Once training is finished, the model weights are saved locally.
- *Transfer learning:* This phase loads saved weights from the previous phase into a new instance of the model (non-trained model).
- *Partial fine-tuning model:* The model parameters are fine-tuned to optimize the overall performance of the model. Input features between datasets, we modified

the input layer of the model to match the feature size of each target dataset. This allowed us to adapt our model to different feature sets without needing full retraining. The model parameters are reduced by 90% – 95% to make a lightweight model and simple training in the target dataset to acquire foundational insights and patterns.

- *Evaluation:* Model performance evaluation is conducted through metrics including accuracy, precision, recall, and F1 score.

Also, during the implementation phase, some of the technical challenges are related to the computational demands of the Swin transformers, while at the same time, the dearth of high-quality training data for LSTM results in extremely time-consuming processes. This also poses a challenge by the incorporation of LSTM layers that may handle sequential data. Optimizations are done to mitigate these issues. Architectural changes include adding dropout and changing feed-forward dimensions based on the model complexity to reduce computational costs for Swin transformer components. This will enable the transformer to handle large-scale data effectively. Since the availability of quality training data for the LSTM is limited, the integration with the Swin transformer is made easy by ensuring that the output of the Decoder feeds into the input layer of the LSTM. This optimization involves treating the decoder output with adaptive average pooling and dropout layers before passing it into the LSTM for better flow between different model components. On the other hand, in the model, the input is treated both with the transformer and LSTM to capture contextual and sequential dependencies effectively. The final output is further refined using a convolutional layer with the *sigmoid* activation that gives the ultimate prediction.

V. EXPERIMENTAL ANALYSIS AND DISCUSSION

In this section, we perform several experiments where we use the NSL-KDD dataset as a source for training and the ToN-IoT, BoTIoT, MQTTIoT, and CICIoT2023 datasets as a target for evaluation. Note that, all experiments are run on a Dell *G*5 laptop powered by an *Intel* Core *i*7 processor, 32 GB of *RAM*, and an *Nvidia* 1650TI graphic card and software used includes Python 3.10, *TensorFlow*, and *Pandas*. First, we provide details for the utilized datasets in this work in the next subsection.

A. DATASET DESCRIPTION

This section presents a detailed overview of the datasets utilized in this study.

1) NSL-KDD DATASET

The NSL-KDD dataset is widely recognized and highly used to evaluate ML models in IDS [17]. It is an enhanced version of the original KDD Cup 1999 dataset, in which additional variability in attack types was added and several redundant records were removed to make it more reliable and useful.

TABLE 2. Description of the NSL-KDD dataset.

Property	Value
Number of Features	41
Number of Samples	148,517
Number of Classes	Normal+ 4 attacks classes
Samples per Class	Normal: 78,646 DoS: 54,022 Probe: 13,813 R2L: 1,959 U2R: 77

TABLE 3. Descriptions of the ToN-IoT dataset.

Property	Value
Features	43
Samples	16,977,496 (combined across four subsets)
Classes	Normal + 9 attacks classes
Samples per Class	Normal: 6,099,469 Scanning : 3,781,419 XSS : 2,455,020 DDoS : 2,026,234 Password : 1,153,323 DoS : 712,609 Injection : 684,465 Backdoor : 16,809 MITM : 7,723 Ransomware : 3,425

The NSL-KDD dataset contains approximately 148, 517 records of both normal traffic and several types of attacks. Detailed with benign and diverse attack vectors categorized into four groups: Normal, DoS, Probe, R2L, U2R. The NSL-KDD dataset is well-known for its research value, is now a standard dataset for IDS ML model benchmarking, also is frequently used in academic publications. A summary of these details is provided in Table 2.

2) TON-IOT DATASET

The ToN-IoT dataset is a group of datasets focused on privacy and security risks that are common in the IoT space [18]. The main goal of curating this dataset was to investigate how heterogeneity affects IoT network intrusion datasets and to highlight the importance of standardizing attack classifications and attribute sets. The ToN-IoT dataset consists of nine classes (Scanning, XSS, DDoS, Password, Injection, DoS, Backdoor, MITM, Ransomware). The content of the dataset covers a wide range of attack types, including malware infections, denial-of-service assaults, and unauthorized access, providing a complete view of the security difficulties associated with IoT networks. A summary of these details is provided in Table 3.

3) BOTIOT DATASET

Due to the lack of publicly available botnet datasets specifically designed for the IoT, the BoTIoT dataset, represents a significant addition to the area of IoT security [19], [37].

TABLE 4. Description of the BoTIoT dataset.

Property	Value
Features	46
Samples	4,292,0000
Classes	Normal + 7 attacks classes
Samples per Class	Normal: 477,000 DDoS: 1,300,000 DoS: 1,100,000 Reconnaissance: 1,000,000 Theft (Data Exfiltration): 50,000 Keylogging: 15,000 OS Scan: 200,000 Service Scan: 150,000

This environment incorporated a combination of normal and botnet traffic. The dataset's source files are provided in various formats, including the original pcap files, generated argus files, and CSV files. The dataset includes DDoS, DoS, OS and Service Scan, Keylogging, and Data Exfiltration attacks, with DDoS and DoS attacks further categorized based on the protocol used. It consists of $72,000,000$ instances with 46 features, due to the large size, this work was conducted on 5% of the total dataset size. Table 4 lists a summary of the reduced and preprocessed BotIoT dataset.

4) MQTTIOT DATASET

The MQTTIoT dataset represents a premium resource to enable practical IDS within the ecosystem of IoT. Introduced by Hindy et al. [20], it meets an increasing need for robust IDS solutions, given the rapid growth not only in IoT devices but also in the communication protocols between them. This pioneering dataset emulates an MQTT-based IoT network architecture comprising twelve sensors. This dataset includes five different attack scenarios: normal operation, aggressive scan, UDP scan, Sparta SSH brute-force, and MQTTIoT brute-force attacks. This dataset allows full feature extraction at three levels of abstraction: packet features, unidirectional flow features, and bidirectional flow features. The basic packet characteristics include flags, length, and MQTT message parameters, while other flow-specific information may be differentiated by forward and backward prefixes. In this work, we used a preprocessed version introduced by [38], Table 5 lists a details summary of this dataset.

5) CICIOT2023 DATASET

CICIoT2023 [21], [29] is a dataset that is provided by Neto et al. to aid in the development and assessment of intrusion detection systems. This dataset is quite thorough and provides a broad and realistic testbed for analyzing the efficiency of security solutions adapted to the vast spectrum of IoT-specific cyber threats. As a summary, Table 6 lists the seven classes in which it hosts network behavior from 105 IoT devices that are under attack. These classes include DoS, DDoS, Web-based, Reconnaissance, Spoofing, Mirai, and Brute Force, and together they capture the complexity and variability of such conditions with 46 unique features.

TABLE 5. Descriptions of the MQTTIoT dataset.

Property	Value
Features	23
Samples	495,134
Classes	Normal + 5 attacks classes
Samples per Class	Normal: 330,204 Flood: 58,356 Slowite: 44,864 Malformed Data: 35,295 DoS: 16,455 Bruteforce: 9,960

TABLE 6. Description of the CICIoT2023 dataset.

Property	Value
Features	46
Samples	46,686,579
Classes	Normal + 7 attacks classes
Samples per Class	Benign: 1,098,195 DDoS: 33,984,560 DoS: 8,090,738 Mirai: 2,634,124 Spoofing: 486,504 Recon: 354,565 Web: 24,829 BruteForce: 13,064

B. EVALUATION METRICS

Several metrics are used in IDS for the IoT to evaluate model performance. These measurements shed light on how well the model detects hazards. These measures consist of the F1 score, recall, accuracy, and precision. The percentage of corrected forecasts for each prediction is determined by accuracy. The following formula is used to compute it:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

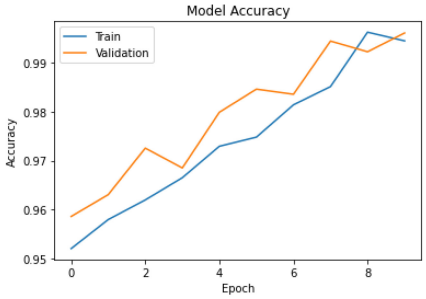
TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

The percentage of true positive forecasts among all positive predictions is known as Precision. It is defined as follows and signifies the model's predictive accuracy for positive instances:

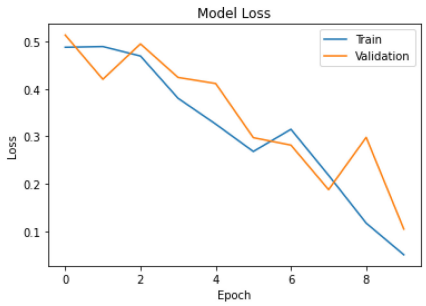
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

The true positive rate, or Recall, gauges how well the model can detect all true positive cases. The formula is used to calculate it:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

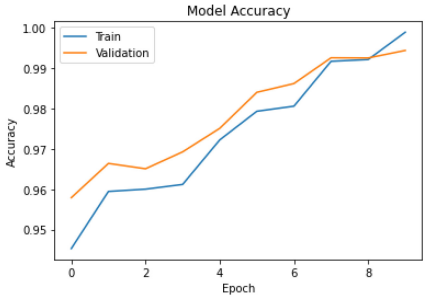


(a) Accuracy curve

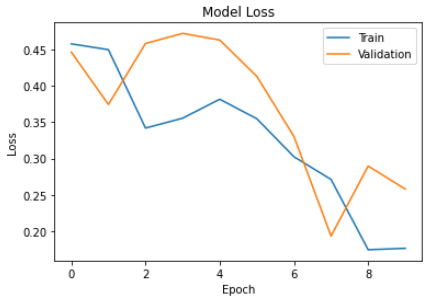


(b) Loss curve

FIGURE 5. The accuracy and loss curves of the proposed model for the ToN-IoT dataset.

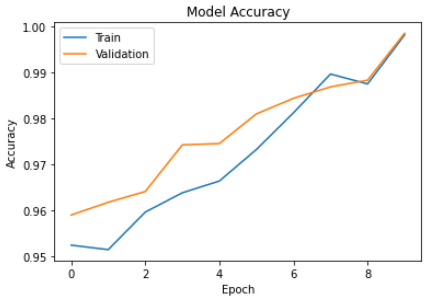


(a) Accuracy curve

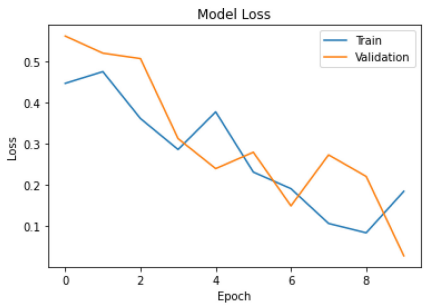


(b) Loss curve

FIGURE 6. The accuracy and loss curves of the proposed model for the BoTIoT dataset.



(a) Accuracy curve



(b) Loss curve

FIGURE 7. The accuracy and loss curves of the proposed model for the MQTTIoT dataset.

The F1 score combines recall and accuracy to produce a fair assessment. It is these two measures' harmonic mean. It provides a thorough assessment of the model's functionality. The following formula determines the F1 score:

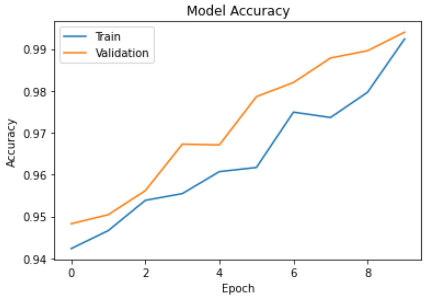
$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

These metrics provide a thorough insight of how well a model performs in IoT environment intrusion classification.

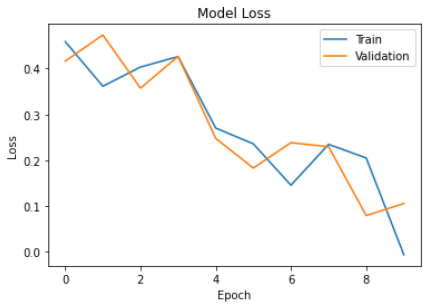
C. COMPARATIVE RESULTS AND DISCUSSION

The evaluation of the proposed model is based on testing it on different datasets and comparing it with state-of-the-art(SOTA) DL models. The proposed model was trained for 10 epochs to ensure sufficient learning. This number was determined empirically based on performance on the validation phase. One standard method by which it can be achieved is through the use of the learning curve method by which the accuracy of the model with respect to a chosen parameter is plotted [46]. Further, the loss curve approach can be used to monitor the training curves of a model and estimate its performance as emphasized in [47]. Figures 5(a)-8(a) show accuracy curves that describe the efficiency of the proposed model across the four datasets.

From the curves, the training accuracy is first a bit lagging behind the validation accuracy but catches up quickly and achieves a steady lead. This small lag may indicate an appropriate learning rate where the model doesn't overfit at all, as the validation accuracy closely tracks the training accuracy, hence generalisability of the model. These curves represent a smooth increase in both training and validation accuracies with epochs. Both curves tend towards very high accuracy, further suggesting a high model efficacy. Figures 5(b)-8(b) give the loss curves of model performance across epochs; it is observable from both graphs that, in general, both training



(a) Accuracy Curve



(b) Loss Curve

FIGURE 8. The accuracy and loss curves of the proposed model for CICIoT2023 dataset.

and validation loss are moving downwards. However, for the presented curves, there are fluctuations, especially in the case of the validation loss, which may indicate variability in the generalisation capability of the model across subsets of data.

D. COMPARISON WITH OTHER MODELS

In this section, we undertake a comparative analysis of the performance of the developed IDS model in contrast to other models such as Autoencoders [48], RecNets [49], RecNNs [50], CNN, RNN [52], and LSTM [53]. In the experiments, both binary and multi-classification tasks are employed to assess the performance of the models, inclusive of the proposed model. The parameter settings for these models are detailed in Table 7 with the proposed model. It is interesting to note the diversity in architectural complexity, as seen in the varying number of layers, from a simple 3-layer setup in Autoencoders and RNNs to a more complex hierarchical structure in the proposed model characterised by a $2-2-6-2$ arrangement. The table also highlights the model-specific settings, like the number of units, where the proposed model utilizes a multi-head attention mechanism denoted by 3×96, indicating 3 heads with a dimension of 96 each. The proposed model uses the Gaussian error linear unit (GELU) instead of the more widely used hyperbolic tangent (tanh) and rectified linear unit (ReLU) activation functions. The learning rate and batch size parameters are adjusted to each model's requirements. All models were trained for 10 epochs to ensure sufficient learning. The parameter settings are adjusted by the try and error method.

Tables 8-11 collectively provide a comprehensive assessment of the proposed model and other DL models. For binary classification, as exemplified in Table 8 using the ToN-IoT dataset as an illustration, the proposed model shows exceptionally high performance across all metrics, nearing perfect scores with over 99% in each. LSTMs also perform very well, especially with regard to recall and F1 Score, both surpassing the 99% threshold. RNNs are close behind, with slightly lower scores than LSTMs but still above 98% across all metrics. CNNs show strong precision and F1 Scores, indicating good positive predictive value and balance between precision and recall. Autoencoders and RecNNs have comparable performance, with Autoencoders slightly ahead, especially in accuracy and precision. ResNets have the lowest performance in binary classification among the models listed.

Additionally, multi-class categorization results are shown in Table 8. With the greatest results, particularly in accuracy, precision, and F1 Score, the proposed model comes in first, demonstrating its excellent ability in accurately classify occurrences across several classes. LSTMs perform exceptionally well, especially in recall and F1 Score, indicating that they can find pertinent cases in a variety of classes. RNNs perform somewhat better than LSTMs but still have high scores comparable to them. Once more displaying a solid balance between recall and precision, CNNs show high precision and F1 Scores. Autoencoders perform relatively well, with consistent scores across all measures. ResNets also perform the worst in multi-class classification, with all metrics hovering around the mid-80s.

A detailed summary of the model performance metrics for binary and multi-class classification scenarios tested on the BoTIoT dataset can be found in Table 9. The proposed model remarkably outperforms other conventional designs, obtaining an accuracy of 98.02% in multi-class classification and 99.93% in binary classification. Similarly, using the MQTTIoT dataset, Table 10 compares the proposed model's performance to those of other models. The proposed model delivers excellent results, boasting an accuracy of 97.93% in binary classification and 96.46% in multi-class classification. Also, the results presented in Table 11 are the metrics values for implementing the proposed model on the CICIoT2023 dataset compared to other DL models. The proposed model achieves an accuracy of 98.78% for binary classification and 97.44% for multi-class classification.

The proposed model's strong recall, F1 scores, and accuracy throughout Tables 8-11 highlight how well it detects genuine positives while keeping a low false positive rate. This feature is essential for guaranteeing an IDS's dependability and credibility, especially in the dynamic and frequently unpredictable world of IoT security. All things considered, these findings highlight how crucial the proposed model is for differentiating between typical and unusual occurrences. Thereby validating its robustness and reliability for both binary and multi-classification tasks. The exceptional performance can be explained by the

TABLE 7. Parameter settings for all models.

Parameter	Autoencoders	ResNets	RecNNs	CNN	RNN	LSTM	Proposed Model
Number of Layers	3	18	3	5	3	3	2-2-6-2
Number of Units (Heads×Dim)	64	N/A	N/A	32	100	100	3 × 96
Activation Function	ReLU				tanh		GELU
Optimizer	Adam						AdamW
Learning Rate	0.001						0.0005
Batch Size	32						16
Loss Function	Cross-Entropy						
Epochs	10						

TABLE 8. Performance of the proposed model for the ToN-IoT dataset compared to other DL models.

Model	Binary Classification				Multi Classification			
	Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
Autoencoders	94.87	94.65	94.65	94.65	92.47	92.41	92.4	92.40
ResNets	87.57	85.87	85.32	85.59	84.75	84.7	83.5	84.10
RecNNs	90.07	88.28	88.25	88.26	88.3	87.7	88.3	88.00
CNN	94.60	90.74	90.21	90.47	92.4	92.74	93.01	92.57
RNN	98.23	98.02	98.01	98.01	97.5	98.04	97.5	97.77
LSTM	99.02	99.30	99.01	99.15	98.04	97.6	98.6	98.10
Proposed model	**99.97**	**99.86**	**99.76**	**99.81**	**98.8**	**98.24**	**98.67**	**98.45**

TABLE 9. Performance of the proposed model models for BoTIoT dataset compared to others DL models.

Model	Binary Classification				Multi Classification			
	Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
Autoencoders	94.43	94.21	94.41	94.31	92.32	93.02	92.32	92.67
ResNets	86.07	84.44	83.82	84.13	84.7	81.7	81.7	81.70
RecNNs	87.07	85.25	85.25	85.25	82.71	81.78	81.5	81.64
CNN	93.6	93.42	89.21	91.27	91.32	90.87	91.32	91.09
RNN	97.23	98.02	97.23	97.62	96.07	96.02	96.11	96.06
LSTM	97.86	98.11	97.11	97.61	96.08	96.01	96.02	96.01
Proposed model	**99.93**	**99.74**	**99.72**	**99.73**	**98.02**	**98.01**	**98.02**	**98.01**

TABLE 10. Performance of the proposed model for MQTTIoT dataset compared to other DL models.

Model	Binary Classification				Multi Classification			
	Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
Autoencoders	92.43	92.01	92.21	92.11	89.45	89.74	89.41	89.57
ResNets	84.07	84.45	83.82	84.13	81.87	81.94	81.87	81.90
RecNNs	85.07	84.87	85.02	84.94	80.67	80.78	80.62	80.70
CNN	94.6	94.04	94.23	94.13	90.74	91.01	90.87	90.94
RNN	95.23	95.04	95.01	95.02	94.05	94.07	94.05	94.06
LSTM	95.82	96.11	95.79	95.95	95.02	94.98	94.8	94.89
Proposed model	**97.93**	**97.72**	**97.84**	**97.78**	**96.46**	**96.47**	**96.41**	**96.44**

distinctive architecture and learning mechanisms inherent to the proposed model which have been optimised for high-dimensional and complex datasets.

Moreover, Figures 9–12 present the graphics results for the model compared to DL models for binary classification tasks while Figures 13–16 multi-classification tasks for the four datasets. As observed from these figures, the proposed model shows superior performance across all metrics, with particularly high scores in accuracy, precision, and F1 score, indicating its accuracy as well as its effectiveness in classifying anomaly instances without many false positives and a balanced trade-off between precision and recall. While still performing well, autoencoders lag behind the proposed model. While RNNs and LSTM display competitive results but are marginally surpassed by the proposed model, ResNets, RecNNs, and CNNs provide modest performance. The results given unequivocally demonstrate that, in comparison to the other models, our model functions as a more reliable and accurate classifier.

TABLE 11. Performance of the proposed model for CICIoT2023 dataset compared to other DL models.

Model	Binary Classification				Multi Classification			
	Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
Autoencoders	93.53	93.21	93.31	93.28	90.55	90.84	90.51	90.63
ResNets	85.17	85.55	84.92	85.23	82.97	83.04	82.97	83.00
RecNNs	86.17	85.97	86.12	86.04	81.77	81.88	81.72	81.80
CNN	95.70	95.14	95.33	95.23	91.84	92.11	91.97	92.04
RNN	96.33	96.14	96.11	96.12	95.15	95.17	95.15	95.16
LSTM	96.92	97.21	96.89	97.05	96.12	96.08	95.90	95.99
Proposed model	**98.78**	**98.54**	**98.77**	**98.56**	**97.44**	**97.67**	**97.37**	**97.76**

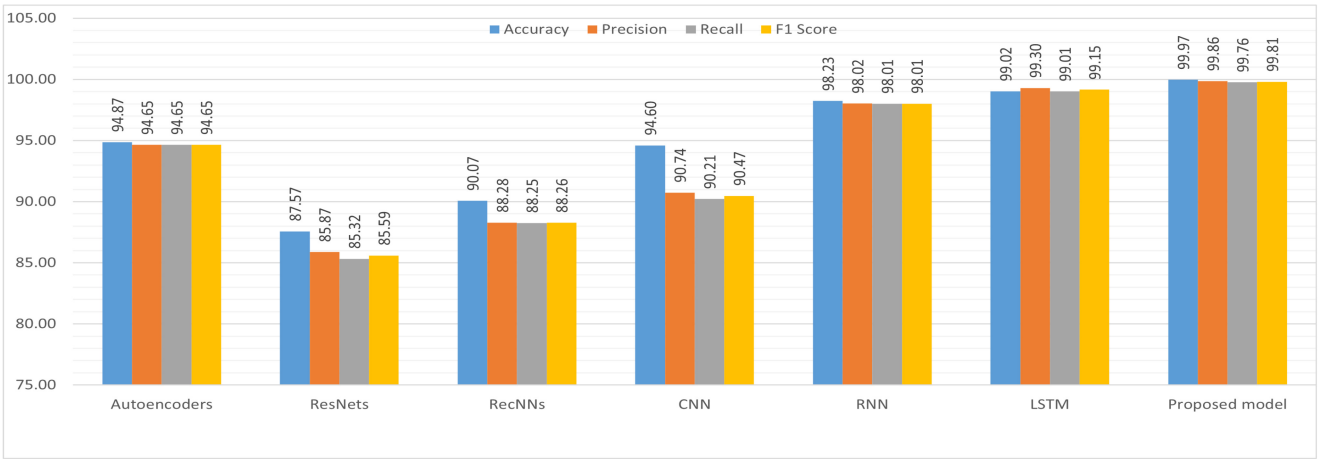


FIGURE 9. Results of binary classification for the ToN-IoT dataset.

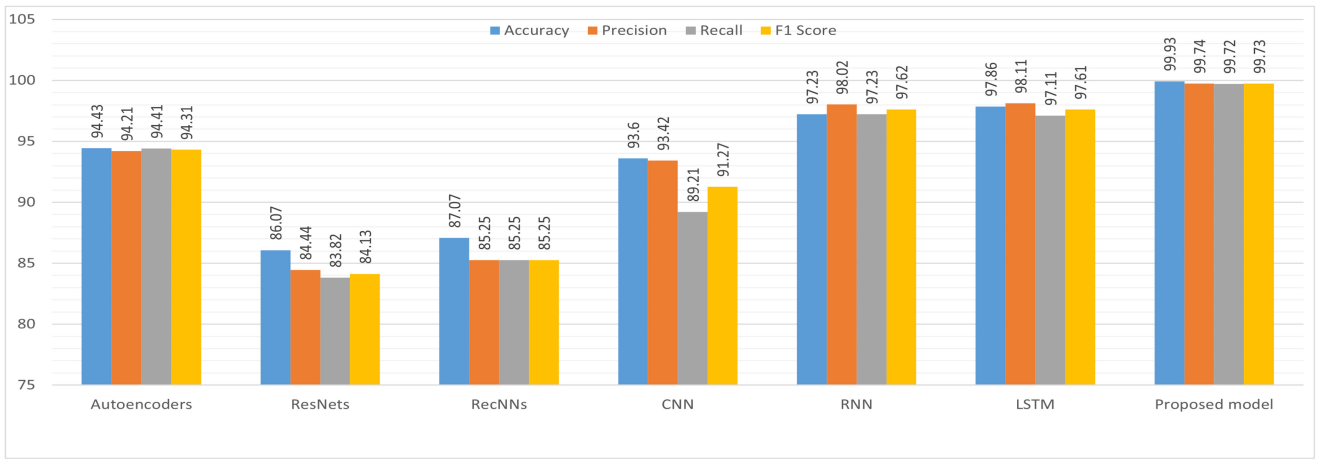


FIGURE 10. Results of binary classification for the BoTIoT dataset.

Finally, the Friedman test is used to check the significant differences between treatments within multiple test attempts of the four datasets of both binary and multi-classification tasks discussed in [54]. A model with the lowest rank overall is the most proficient performer for all metrics across the tasks. Figure 12 presents the overall rankings across all datasets and metrics placing the proposed model at the top, followed by LSTM, RNN, Autoencoders, CNN, RecNNs, and ResNets in descending order, indicating its superior performance.

Thus, the proposed model outperforms all the other compared DL models in terms of all metrics on all four datasets effectively, revealing its dominance in both binary and multi-class classification. The outstanding performance signifies that the proposed model is effective in detecting and classifying anomalies and represents a very good candidate for IDS in different IoT scenarios. Tables and figures also reveal that although some of the models have competitive performances with high accuracy, precision, and recall, the proposed model always has the top value. This is the ability

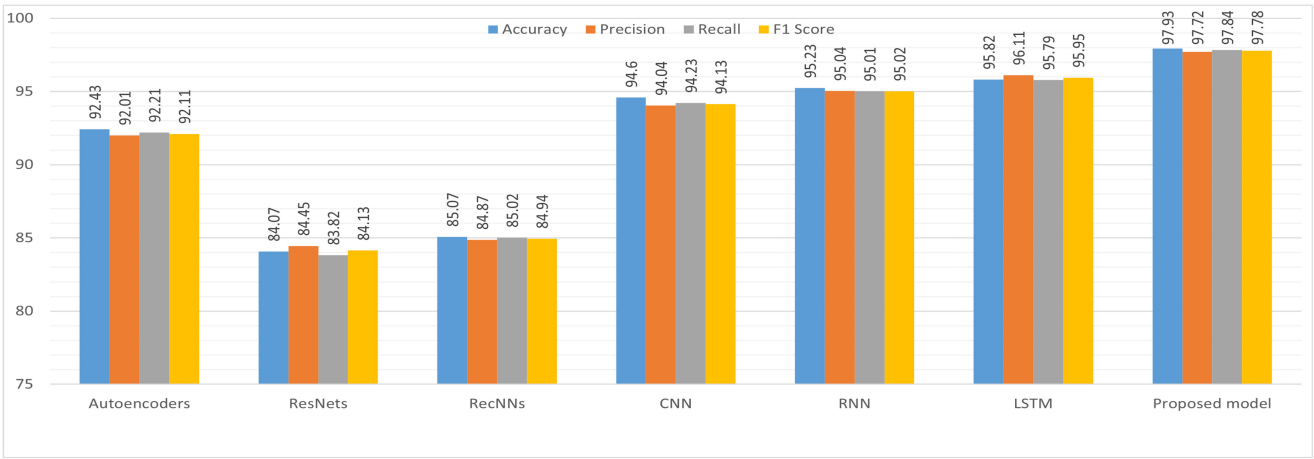


FIGURE 11. Results of binary classification for the MQTTIoT dataset.

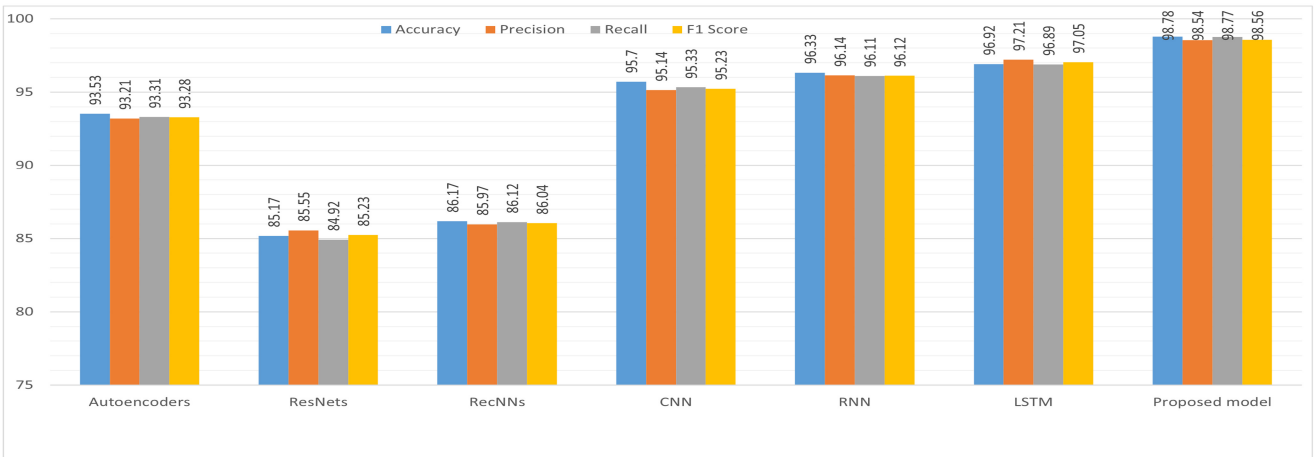


FIGURE 12. Results of binary classification for the CICIoT2023 dataset.

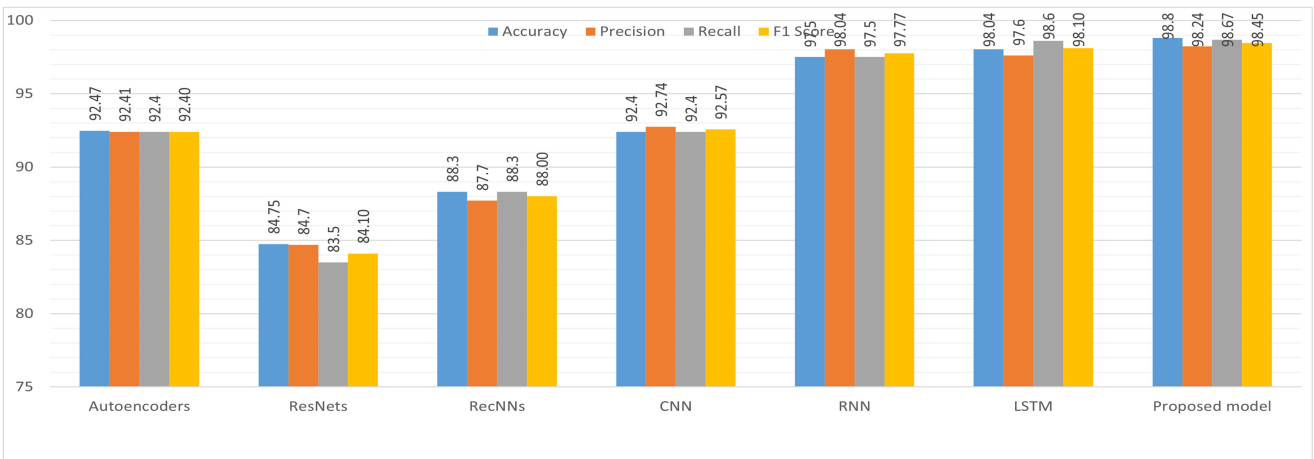


FIGURE 13. Results of multi-classification for the ToN-IoT dataset.

of the model to balance between intrusion identification and false alarm avoidance, which is the most important characteristic in keeping IoT systems cybersafe and intact. What it implies is that the proposed model is strong and shows its robustness when compared with its variants on different datasets-a lot about generalisability and adaptability, the essential ingredients when it comes to future deployment in IoT environments.

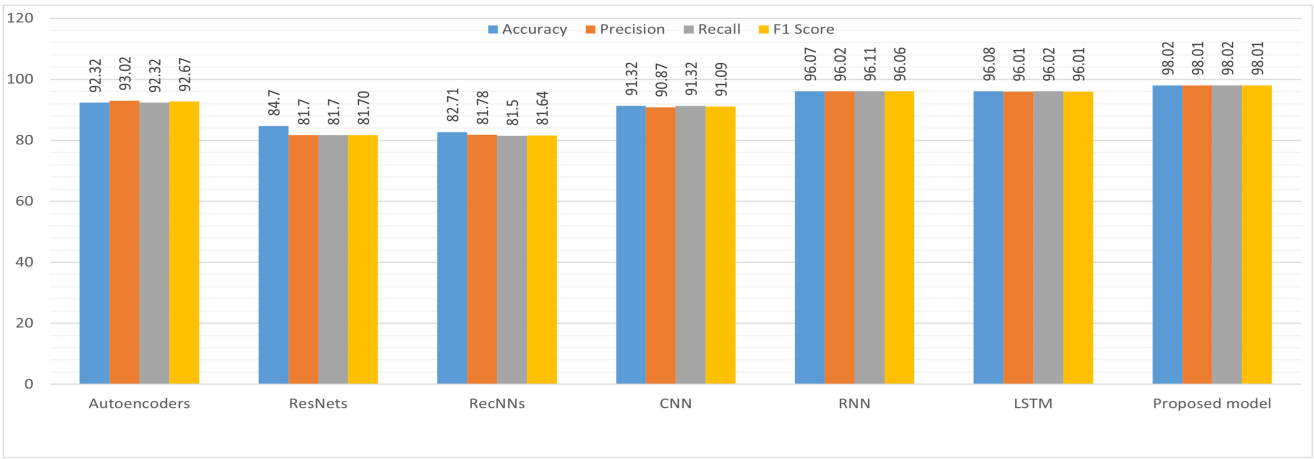


FIGURE 14. Results of multi-classification for the BoTIoT dataset.

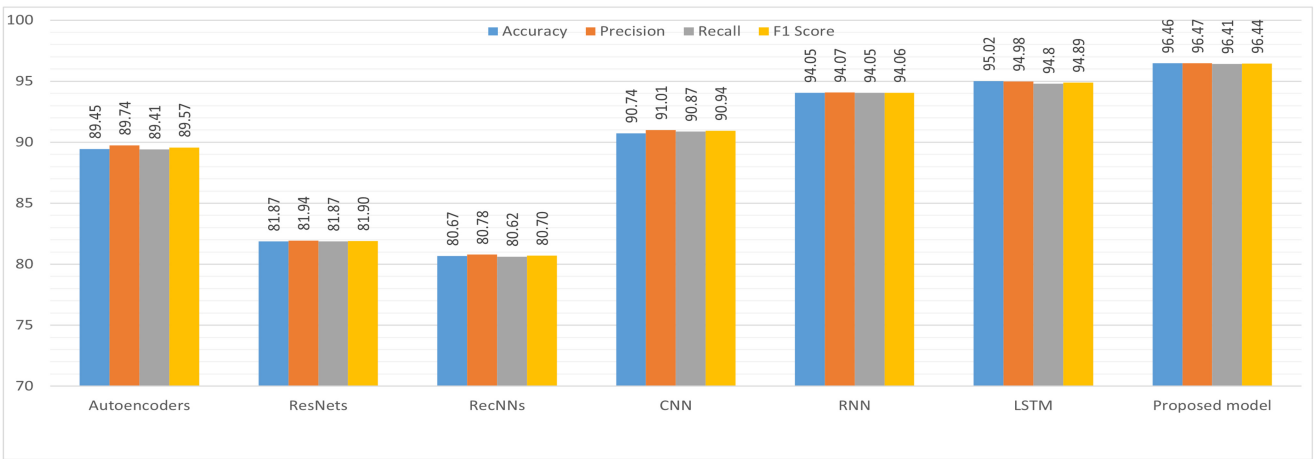


FIGURE 15. Results of multi-classification for the MQTTIoT dataset.

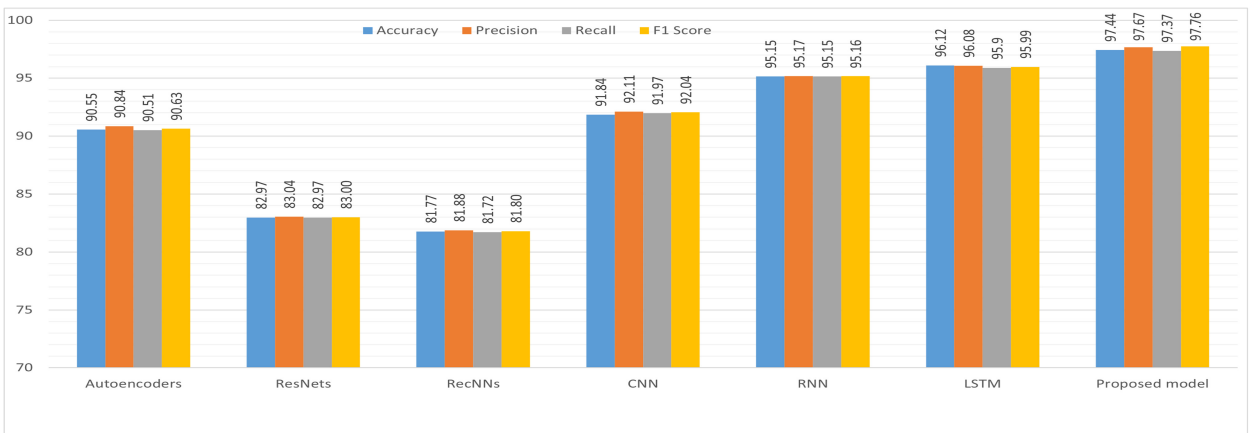


FIGURE 16. Results of multi-classification for the CICIoT2023 dataset.

The tables, Friedman test, and figures above succinctly illustrate the importance of careful model evaluation and selection when doing research work in the area of IDS in IoT. The consistent superior performance from the proposed model shows its potential for providing a strong and efficient solution in enhancing the security of IoT systems; it therefore constitutes an important building block of the general discourse on the protection of IoT networks from malicious actions.

TABLE 12. Performance of the proposed model across CICIoT2023, NSL-KDD, ToN-IoT, BoTIoT, and MQTTIoT datasets.

#	Dataset	Binary Classification				Multi Classification			
		Acc.	Prec.	Recall	F1	Acc.	Prec.	Recall	F1
Source	CICIoT2023	99.01	98.97	99.00	98.99	98.40	98.65	98.23	98.52
Target	NSL-KDD	99.50	99.40	99.35	99.45	99.10	99.05	98.95	99.07
	ToN-IoT	99.97	99.74	99.76	99.85	99.00	98.59	98.96	98.79
	BoTIoT	99.87	99.84	99.67	99.85	98.14	98.10	98.00	98.12
	MQTTIoT	98.54	98.23	98.53	98.38	97.89	97.00	97.89	97.44

E. EFFICIENCY ANALYSIS

To fine-tune the proposed model for use as a target dataset like NSL-KDD and IDS-based-IoT datasets like ToN-IoT, BoTIoT, and MQTTIoT, it was trained on a variety of sources (i.e., IDS datasets) such as the CICIoT2023 dataset. The performance of the proposed model is discussed in this subsection.

The performance of the suggested model on the CICIoT2023, NSL-KDD, ToN-IoT, BoTIoT, and MQTTIoT datasets is summarized in Table 12. For both multi-classification and binary metrics. These findings demonstrate the model's resilience and its capacity to handle the variety of IDS data that are present in an IoT setting. The model, for example, achieves remarkable accuracy of 98.40% in multi-classification and 99.01% in binary for the CICIoT2023 dataset, which serves as the training dataset. The outcomes of the target datasets, which include NSL-KDD, ToN-IoT, BoTIoT, and MQTTIoT, demonstrate how well the model generalizes across a range of IoT IDS situations. The best accuracy scores are reached on the ToN-IoT dataset (99.97%) thus indicating how exact the model is in an environment focused towards IoT. However, even if the scores in BoTIoT and MQTTIoT are slightly lower, they are still rather high, indicating that the suggested model is highly versatile and capable of maintaining high levels of detection accuracy in a wide range of scenarios. The variations in these datasets' properties and the kinds of assaults they contain can be used to explain the modest variations in performance metrics between them.

Figure 17 represents the binary evaluation results of the proposed model across all four datasets across all metrics. The model displays exceptional proficiency, with a special peak in performance on the ToN-IoT and CICIOT2023 datasets reflecting its capability to identify and classify ID with near perfection. Also for the MQTTIoT datasets, the model achieved 98.54% in accuracy which is a high performance but a bit lower than other datasets. While in Figure 18, it represents the multi-classification evaluation across all datasets. It can be noted from the figure that the best performance for our proposed model is achieved on the ToN-IoT, CICIOT2023, and NSL-KDD datasets but slightly low for the MQTTIoT dataset.

To clear the efficiency of the proposed model, Table 13 presents the performance metrics across all datasets and compares them with DL models, both for binary and multi-classification. The metrics assessed include detection rate (DR), false positive rate (FPR), and processing time (in seconds), whereby the resultant analysis still concludes that the proposed model is superior to other DL models.

The proposed model performs exceptionally well on the binary classification test, obtaining the highest DRs with the lowest FPRs across all datasets. Specifically, in the ToN-IoT dataset, the suggested model achieves a DR of 99.9% with an exceptionally low FPR of 0.05%, demonstrating the model's effectiveness in accurately identifying pertinent risks. In scenarios where multi-classes need to be determined, the proposed model again proves better than other models because it always maintains high DRs and low FPRs in general and in all implemented datasets such as CICIOT2023, NSL-KDD, and BoTIoT. This also shows that the model is robust and capable of handling intricate and diverse scenarios in ID.

The LSTM model follows as the second best, especially in datasets like NSL-KDD and ToN-IoT, which really shows the aptitude of the model towards sequence data processing which is crucial in network traffic analysis. The RNN and CNN models, although good, are somehow overshadowed by the LSTM because LSTM is better able to capture long-term dependencies than others. In the case of models like Autoencoders, RecNNs, and ResNets, while providing crucial insights into the ID field, the levels of efficacy of the models are really diverse, and that can explain why the performance of the models is really decreased in such challenging datasets as MQTTIoT and CICIoT2023. This could be due to the architectural nuances of the models and the inherent complexities of the datasets that may affect the models' ability to generalise across different types of network behaviours and attack vectors.

The presented results in Table 13 prove that the proposed model has advanced detection capabilities, especially with regard to maintaining high detection rates and low false positives, which are paramount in an effective IDS implementation. The processing times across all models indicate the practicality of deploying these models, where the proposed model also performs best in operational efficiency. These results show that the model performs efficiently in the detection and handling of almost every class of intrusion. Also, in this subsection, the results presented underline the adaptability to the various types of attacks and scenarios. This underlines the potential of the model as an adaptive tool in IoT security that is ready to be deployed in real-world applications for various scenarios.

TABLE 13. Efficiency comparison of the proposed model across various datasets.

Model	Dataset	Binary Classification			Multi Classification		
		DR (%)	FPR(%)	Time	DR(%)	FPR(%)	Time
Proposed model	CICIoT2023	99.14	0.17	950	97.87	0.2	1312
	NSL-KDD	99.07	0.09	100	99.00	0.17	134
	ToN-IoT	99.9	0.05	460	98.75	0.2	520
	BoTIoT	99.81	0.14	375	98.47	0.2	415
	MQTTIoT	98.35	0.22	24	97.5	0.3	58
LSTM	CICIoT2023	97.29	0.27	1075	96.09	0.38	3011
	NSL-KDD	99.01	0.20	110	98.59	0.31	140
	ToN-IoT	99.08	0.17	585	98.17	0.2	615
	BoTIoT	97.8	0.35	432	96.75	0.37	605
	MQTTIoT	96.41	0.32	52	95.2	0.38	75
RNN	CICIoT2023	95.97	0.3	2055	95.54	0.35	3020
	NSL-KDD	99.3	0.2	115	98.34	0.27	135
	ToN-IoT	97.7	0.28	604	96.67	0.35	689
	BoTIoT	97.55	0.31	553	96.47	0.37	630
	MQTTIoT	96.3	0.31	48	95.12	0.41	69
CNN	CICIoT2023	96.01	0.35	2080	91.7	0.45	3070
	NSL-KDD	99.32	0.22	124	98.23	0.25	133
	ToN-IoT	94.57	0.32	652	93.5	0.4	725
	BoTIoT	94.7	0.34	632	92.6	0.37	750
	MQTTIoT	94.5	0.4	62	91.1	0.57	80
Autoencoders	CICIoT2023	93.5	0.5	2072	90.7	0.55	3065
	NSL-KDD	99.47	0.19	115	98.23	0.34	145
	ToN-IoT	94.59	0.42	547	93.4	0.52	635
	BoTIoT	94.6	0.44	505	92.5	0.47	610
	MQTTIoT	93.7	0.57	56	90.0	0.67	74
RecNNs	CICIoT2023	86.3	0.7	3018	81.5	0.75	4321
	NSL-KDD	98.88	0.32	140	98.01	0.55	170
	ToN-IoT	90.27	0.72	615	89.2	0.7	645
	BoTIoT	88.04	0.64	495	82.3	0.92	540
	MQTTIoT	85.2	.80	53	80.8	1.15	80
ResNets	CICIoT2023	85.4	0.73	3201	82.6	0.85	4602
	NSL-KDD	97.74	0.32	135	98.12	0.45	165
	ToN-IoT	87.60	0.52	585	84.3	0.82	640
	BoTIoT	86.5	0.75	510	85.32	1.01	587
	MQTTIoT	84.3	1.04	55	81.9	1.12	76

F. THREAT MODELLING ANALYSIS

The primary security objectives when working with the IDS models would be accuracy to maximise the correct identification of normal and malicious activities, Efficiency to minimise computational overhead while processing the dataset, and scalability to ensure the solution can handle varying sizes of data inputs and potentially more complex attack vectors. Table 14 summarises threat modelling associated with NSL-KDD, ToN-IoT, Bot-IoT, MQTTIoT, and CICIoT2023 datasets. It summarises the datasets by identifying the common threat agents, the potential attacks they are susceptible to, and the vulnerabilities that could be exploited.

The NSL-KDD dataset, as an example, contains a set of data points that are labelled as either normal or as one of several different types of attacks as presented in Table 2. Threat agents for the NSL-KDD dataset include hackers such as amateur and professional, script kiddies, malicious insiders, and automated scripts and bots that execute predefined attack patterns [55]. The potential attacks modelled in this dataset enclose (DoS, R2L, U2R, Probing). The NSL-KDD dataset shows several major network vulnerabilities related to cybersecurity. These include buffer overflows that enable unauthorised code execution through U2R attacks, weak authentication mechanisms exploited by R2L attackers, and ineffective data filtering, which exposes the systems to DoS attacks. Risk assessment incorporates both the frequency and the impact of these attacks: While DoS attacks are responsible for service disruption, U2R and R2L attacks, although less frequent, often lead to great data breaches and

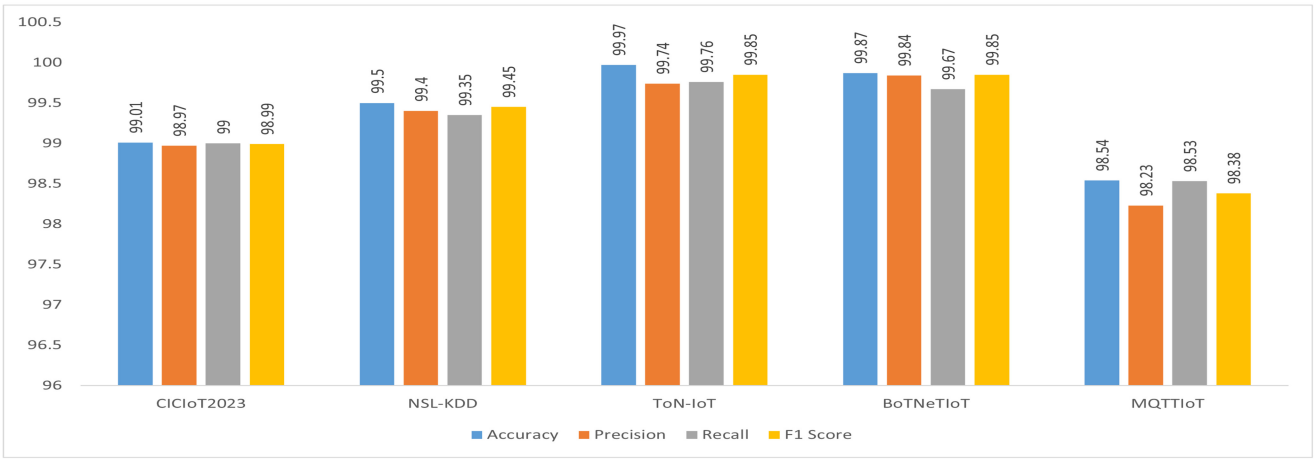


FIGURE 17. Binary classification results of the proposed model across all datasets.

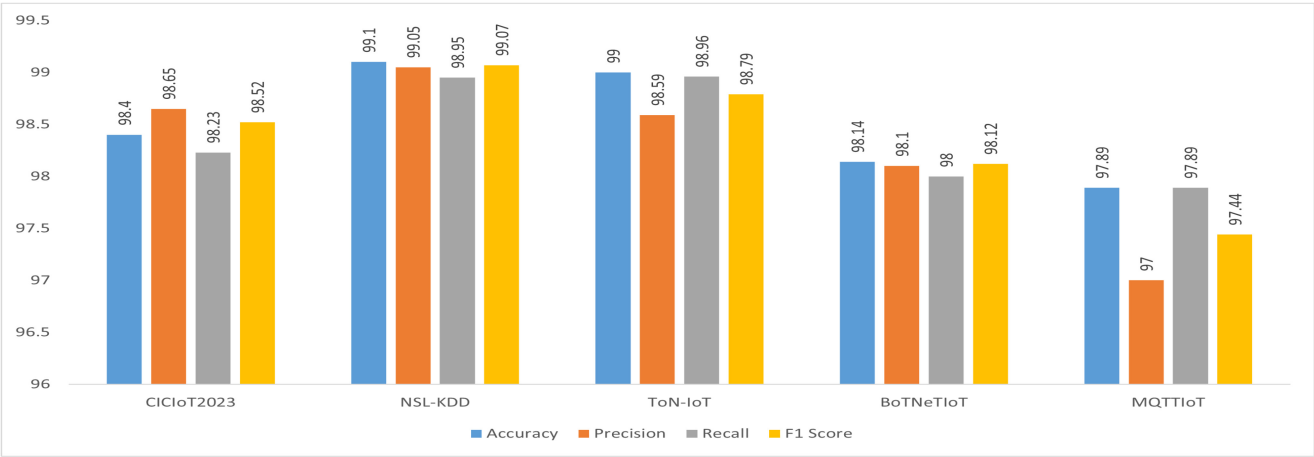


FIGURE 18. Multi-classification results of the proposed model across all datasets.

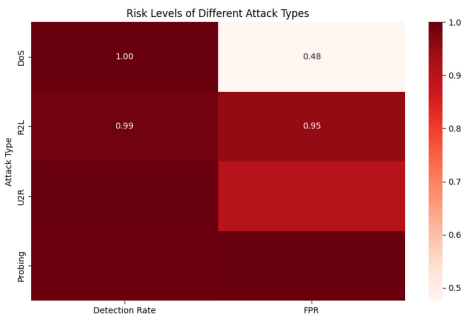
TABLE 14. Threat modelling for various IoT datasets.

Dataset	Threat Agents	Potential Attacks	Vulnerabilities
NSL-KDD	Hackers, Automated scripts	DoS, R2L, U2R, Probing	Protocol vulnerabilities, Improper configurations
ToN-IoT	Hackers, Malware	DDoS, Scan, PortScan, Botnet, Infiltration	Weak security protocols
Bot-IoT	Botnets	DoS, Info Theft	Exploitable IoT device vulnerabilities
MQTTIoT	External Attackers	Brute-force, DoS	Inherent MQTT protocol vulnerabilities
CICIoT2023	Cyber Attackers	Advanced persistent threats, Phishing	IoT device firmware vulnerabilities

severe damage. Proper mitigation efforts include rate-limiting and robust authentication against DoS attacks, enhanced monitoring and strong authentication against U2R and R2L attacks, and sophisticated ID systems to combat probing.

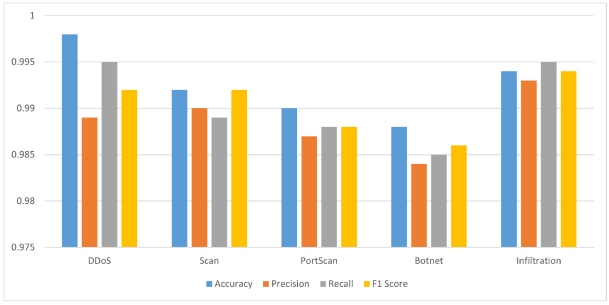
Also from Table 14, the ToN-IoT dataset highlights cybercriminals exploiting IoT vulnerabilities with attacks such as DDoS, Scan, PortScan, Botnet, and Infiltration. This demands robust cybersecurity measures such as secure communication protocols and enhanced authentication. The Bot-IoT dataset thus involves botnets; hence, there is a need to secure IoT devices against unauthorized access and frequent software updates. MQTTIoT has weaknesses in the MQTT protocol that make it prone to brute-force and DoS attacks, which can be reduced using strong encryption and proper authentications. Advanced persistent threats and phishing in regard to the CICIoT2023 dataset require

(a) Performance metrics per each attack type

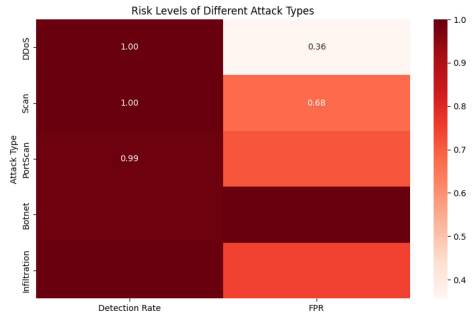


(b) Threat modelling heatmap

FIGURE 19. The threat modelling analysis of the proposed model on the NSL-KDD dataset.



(a) Performance metrics per each attack type



(b) Threat modelling heatmap

FIGURE 20. The threat modelling analysis of the proposed model on the ToN-IoT dataset.

advanced threat detection systems and a full comprehensive security framework of IoT devices. Each of these varied datasets reveals specific cybersecurity challenges in IoT and, as such, underlines the need for continuous updating of threat models and security means so as to efficiently counteract the dynamic cyber threat landscape. Continuous updating and reviewing of the threat model shall be performed with the view of addressing newly appearing threats and maintaining efficacy of the applied security means.

Figure 19(a) illustrates the performance metrics (accuracy, precision, recall, F1 Score) for four main attack types of the NSL-KDD dataset for the proposed model. This breakdown highlights the IDS's robustness, with high values across all metrics, underscoring its effectiveness in identifying and responding to these attacks. Figure 19(b) presents the heatmap of focused comparison of DR and FPR for the same attack types. This visualisation efficiently conveys the IDS's sensitivity and specificity in detecting each attack, with darker shades indicating higher detection rates and lighter shades representing lower false positive rates. The heatmap format succinctly summarises the risk levels associated with each attack type, this facilitates a quick assessment of which attacks are more likely to bypass the IDS and which are consistently identified. We can observe from these Figures (19(a) and 19(b)) that the DoS attacks might be considered the key attack type as they exhibit the highest DRs and lower FPR for the NSL-KDD dataset. Similarly, Figures 20, 21, 22, and 23 present the performance metrics and the heatmap for the main attack types in ToN-IoT, BoTIoT, MQTTIoT, and CICIoT2023 datasets, respectively. The key attacks achieved by the proposed model are summarised in table 15.

TABLE 15. Key attacks overall datasets.

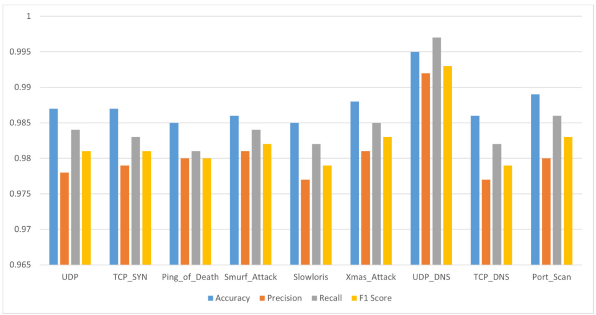
Dataset	First Attack	DR (%)	Second Attack	DR (%)
NSL-KDD	DoS	99.7	U2R	99.5
ToN-IoT	DDoS	99.8	Infiltration	99.4
BoTIoT	UDP_DNS	99.5	Port_Scan	98.9
MQTTIoT	Dos Attack	98.7	Legitimate	98.5
CICIoT2023	DDoS	99.4	Dos	99.0

G. COMPARATIVE ANALYSIS

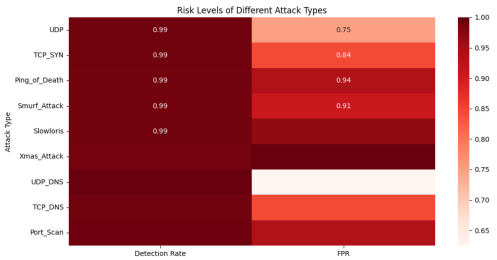
This section presents a critical analysis of various DL models used in developing IDS for IoT. From here, the significance of contemporary DL techniques in IoT security systems has been highlighted. The relevant works related to ToN-IoT, BoTIoT, MQTTIoT, and CICIoT2023 datasets used in this research are summarized in Table 16. It has implemented various models, which range from decision trees to XGBoost, DenseNet, LSTM-RNN, CNN, and the latest hybrids, such as Dugat-LSTM and CNNLSTM, among others. This table effectively summarises the trends in IDS approaches taken by researchers in the last years, from 2019 to 2024. The accuracy percentage of these studies is also quite high, lying mostly above 97%, which indicates the effectiveness of the usage of DL techniques to identify malicious activities within the IoT. Also, binary

TABLE 16. Comparative literature of deep learning models for IDS.

Model (Year)	Dataset Used	Classification Type	Accuracy (%)
DT (2021) [56]	ToN-IoT	Multi	97.29
XGBoost (2021) [57]	ToN-IoT	Multi	98.3
DenseNet (2022) [58]	ToN-IoT	Multi	98.57
ANN (2019) [59]	BoTIoT	NA	99
CNN (2021) [60]	BoTIoT	Multi	99.44
DNN (2021) [61]	MQTTIoT	Multi	98.12
CNN (2023) [62]	BoTIoT	Multi	98.70
CNNLSTM (2024) [63]	MQTTIoT	Multi	98.94
Ensemble LSTM (2021) [64]	BoTIoT	Both	99.9
LSTM-RNN (2023) [65]	ToN-IoT	Binary	99.62
Dugat-LSTM (2024) [66]	ToN-IoT	Binary	99.65
CNN (2021) [67]	BoTIoT	Binary	99.63
LSTM-RNN (2023) [65]	BoTIoT	Binary	88.32
ARTEMIS (2019) [68]	MQTTIoT	Binary	99.41
GAN-AE (2023) [69]	MQTTIoT	Binary	97
TNN-IDS (2023) [70]	MQTTIoT	Binary	99.9
TBF-MD (2024) [71]	CICIoT2023	Both	79.07
LIDS-MV (2024) [72]	CICIoT2023	Both	98.35
DCNN (2024) [74]	CICIoT2023	Both	99.50
CKAN (2024) [75]	ToN-IoT	Both	93.3
	CICIoT2023	Both	99.22
	NSL-KDD	Both	98.71
Proposed Model	ToN-IoT	Both	**99.97**
	BoTIoT	Both	**99.93**
	MQTTIoT	Both	**97.93**

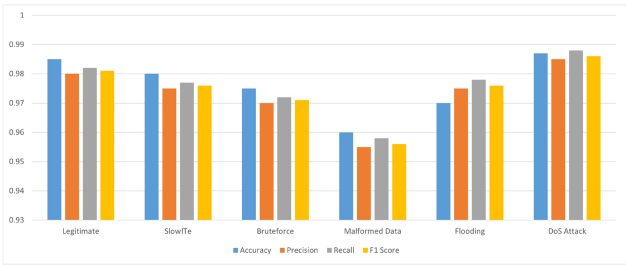


(a) Performance metrics per each attack type

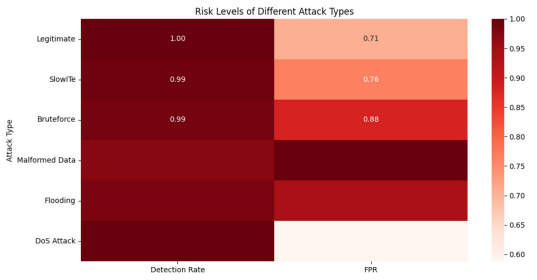


(b) Threat modelling heatmap

FIGURE 21. The threat modelling analysis of the proposed model on the BoTIoT dataset.



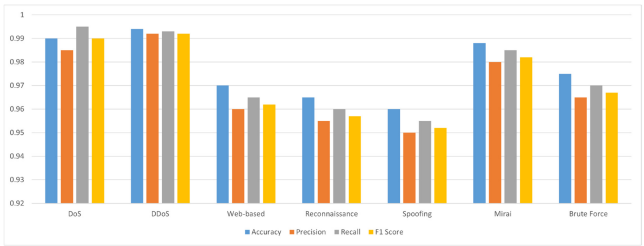
(a) Performance metrics per each attack type



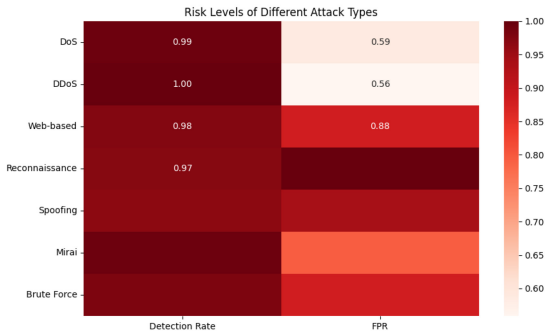
(b) Threat modelling heatmap

FIGURE 22. The threat modelling analysis of the proposed model on the MQTTIoT dataset.

and multi-class classification types were discussed, whose usage ranges from a wide range of intrusion scenarios. This increasingly points, especially in the near future, to hybrid models such as Dugat-LSTM and CNNLSTM, which will further refine the detection capabilities by leveraging strengths from multiple architectures.



(a) Performance metrics per each attack type



(b) Threat modelling heatmap

FIGURE 23. **The threat modelling analysis of the proposed model on the CICIoT2023 dataset.**

That is the interesting observation that can be drawn from this table: the same ToN-IoT and BoTIoT datasets appear in different studies, hence making them important in IDS research for IoT. It would appear they are a comprehensive benchmark to test the strength of various models on different attack scenarios. Again, as demonstrated by GAN-AE and TNN-IDS, the introduction of GAN and temporal neural network methods within the framework of IDS shows a new trend in this area. These proposed approaches are designed to realize high accuracy and address the challenges posed by dynamic and sophisticated cyber threats in IoT environments. While the table indicates that continuous research and development in deep learning-based IDS are well positioned to help make IoT systems resilient to the ever-increasingly complex nature of the cyber threat landscape, there is also a strong focus on accuracy and adaptability to new and emerging types of attacks.

Therefore, the comparative analysis between the proposed model's accuracy and those in the literature table shows a critical trend toward the efficacy of DL methodologies in IDS. Needless to say, the proposed model has quite remarkable accuracy, which often surpasses high-performance benchmarks left by other benchmark models in the field. A comparison with such a performance underlines the robustness of the suggested model and its good performance in identifying and mitigating security threats within IoT environments. The performance, when compared against a wide array of models such as MLP, CNN, LSTM, and their variants, brings into light not only the proficiency of the model in handling difficult data patterns typical of IoT networks but also puts in focus on ongoing enhancement and increased effectiveness of DL approaches with respect to facing dynamics and challenges that face network security.

VI. CONCLUSION AND FUTURE WORK

This paper proposed an enhanced IDS method in the IoT environment, which is mainly a hybrid LSTM-Swin transformer based on the TL mechanism. The results of experiments have, in turn, fully confirmed the excellence of the proposed model. Our proposed model outperforms state-of-the-art DL models such as Autoencoders, ResNets, CNN, RNN, and LSTM and attained an average accuracy of 98.97% across all datasets. Its consistent superiority testifies to the potential of the proposed model for a reliable tool in bolstering IoT system security, hence adding knowledge to the bigger discussion about how best to protect IoT networks against malicious activities. The proposed hybrid model provides an optimistic approach toward IDS in the domain of IoTs, though it might be adapted by other domains where labelled data availability is limited. The use of TL makes the proposed model more scalable and adaptable. This could be further developed in future research from the interpretability of the proposed model with information carried in the presented Tables and Figures by bringing out the underlying patterns and features that make it superior in performance across diverse datasets. This is especially true concerning the scalability of the model to larger, more complex IoT networks and for adapting to the ever-changing threat landscape, which constitutes another area worth investigating in order to obtain certain insights into the feasibility and effectiveness of the model in a real-world environment. All in all, hybrid approaches using the pooled capabilities of several DL models working together may achieve even more robust and accurate IDS IoT environments, using Transformers with domain-specific knowledge.

REFERENCES

[1] B. I. Farhan and A. D. Jasim, "A survey of intrusion detection using deep learning in the Internet-of-Things," *Iraqi J. Comput. Sci. Math.*, vol. 3, pp. 83–93, Dec. 2022.

[2] A. Sedrati and A. Mezrioui, "A survey of security challenges in Internet of Things," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 3, no. 1, pp. 274–280, 2018.

[3] K. Kimani, V. Oduol, and K. Langat, "Cyber security challenges for IoT-based smart grid networks," *Int. J. Crit. Infrastruct. Protect.*, vol. 25, pp. 36–49, Jun. 2019.

[4] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[5] S. Khanam, I. Ahmedy, M. Idris, M. Jaward, and A. Sabri, "A survey of security challenges, attacks taxonomy and advanced countermeasures in the Internet of Things," *IEEE Access*, vol. 8, pp. 219709–219743, 2020.

[6] C. Zhang, D. Jia, L. Wang, W. Wang, F. Liu, and A. Yang, "Comparative research on network intrusion detection methods based on machine learning," *Comput. Security*, vol. 121, Oct. 2022, Art. no. 102861.

[7] R. Vinayakumar, K. Soman, and P. Poornachandran, "Applying convolutional neural network for network intrusion detection," in *Proc. Int. Conf. Adv. Comput. Commun. Informat. (ICACCI)*, 2017, pp. 1222–1228.

[8] F. Laghrissi, S. Douzi, K. Douzi, and B. Hssina, "Intrusion detection systems using long short-term memory (LSTM)," *J. Big Data*, vol. 8, pp. 1–16, May 2021.

[9] S. Chawla, "Deep learning based intrusion detection system for Internet of Things," Ph.D. dissertation, Dept. Comput. Sci., Univ. Washington, Washington, DC, USA, 2017.

[10] F. Ullah, S. Ullah, G. Srivastava, and J. Lin, "IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic," *Digit. Commun. Netw.*, vol. 10, no. 1, pp. 190–204, 2023.

[11] Z. Wu, H. Zhang, P. Wang, and Z. Sun, "RTIDS: A robust transformer-based approach for intrusion detection system," *IEEE Access*, vol. 10, pp. 64375–64387, 2022.

[12] S. Mehedi, A. Anwar, Z. Rahman, and K. Ahmed, "Deep transfer learning based intrusion detection system for electric vehicular networks," *Sensors*, vol. 21, no. 14, p. 4736, 2021.

[13] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[14] Z. Liu et al., "Swin transformer V2: Scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12009–12019.

[15] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.

[16] Z. Zhu, K. Lin, A. Jain, and J. Zhou, "Transfer learning in deep reinforcement learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13344–13362, Nov. 2023.

[17] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "NSL-KDD dataset: A new generation dataset for network-based intrusion detection systems," *Exp. Syst. Appl.*, vol. 36, no. 1, pp. 2935–2943, 2009.

[18] T. Booij, I. Chiscop, E. Meeuwissen, N. Moustafa, and F. Hartog, "ToN-IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 485–496, Jan. 2022.

[19] A. Alhowaide, I. Alsmadi, and J. Tang, "Towards the design of real-time autonomous IoT NIDS," *Clust. Comput.*, vol. 26, pp. 2489–2502, Jan. 2021.

[20] H. Hindy, C. Tachtatzis, R. Atkinson, E. Bayne, and X. Bellekens. "MQTT-IDS2020: MQTT Internet of Things intrusion detection dataset." 2020. [Online]. Available: https://paperswithcode.com/dataset/mqtt-iot-ids2020#: :text=MQTT%2DIoT%2DIDS2020%20is%20the,simulated%20camera%2C%20and%20an%20attacker

[21] E. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, p. 5941, 2023.

[22] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[23] N. Kumar, J. Madhuri, and M. ChanneGowda, "Review on security and privacy concerns in Internet of Things," in *Proc. Int. Conf. IoT Appl. (ICIOT)*, 2017, pp. 1–5.

[24] G. Muhammad, M. Hossain, and S. Garg, "Stacked autoencoder-based intrusion detection system to combat financial fraudulent," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2071–2078, Feb. 2023.

[25] B. Sharma, L. Sharma, C. Lal, and S. Roy, "Anomaly based network intrusion detection for IoT attacks using deep learning technique," *Comput. Elect. Eng.*, vol. 107, Apr. 2023, Art. no. 108626.

[26] P. Wu, H. Guo, and R. Buckland, "A transfer learning approach for network intrusion detection," in *Proc. IEEE 4th Int. Conf. Big Data Anal. (ICBDA)*, 2019, pp. 281–285.

[27] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, 2015, pp. 1–6.

[28] H. Kang, B. Kwak, Y. Lee, H. Lee, H. Lee, and H. Kim. "Car Hacking: Attack & defense challenge 2020 dataset." 2021. [Online]. Available: https://dx.doi.org/10.21227/qvr7-n418

[29] I. A. Fares and M. A. Elaziz, "Explainable TabNet transformer-based on Google vizier optimizer for anomaly intrusion detection system," *Knowl.-Based Syst.*, vol. 316, May 2025, Art. no. 113351.

[30] H. Asgharzadeh, A. Ghaffari, M. Masdari, and F. S. Gharehchopogh, "An intrusion detection system on the Internet of Things using deep learning and multi-objective enhanced gorilla troops optimizer," *J. Bionic Eng.*, vol. 21, no. 5, pp. 2658–2684, 2024.

[31] I. Idrissi, M. Azizi, and O. Moussaoui, "Accelerating the update of a DL-based IDS for IoT using deep transfer learning," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 1059–1067, 2021.

[32] N. Moustafa. "The Bot-IoT dataset." 2019. [Online]. Available: https://dx.doi.org/10.21227/r7v2-x988

[33] Y. Liu and L. Wu, "Intrusion detection model based on improved transformer," *Appl. Sci.*, vol. 13, no. 10, p. 6251, 2023.

[34] A. Kumar, S. Raja, N. Pritha, H. Raviraj, R. Lincy, and J. Rubia, "An adaptive transformer model for anomaly detection in wireless sensor networks in real-time," *Meas. Sensors*, vol. 25, Feb. 2023, Art. no. 100625.

[35] S. Wang, W. Xu, and Y. Liu, "Res-TranBiLSTM: An intelligent approach for intrusion detection in the Internet of Things," *Comput. Netw.*, vol. 235, Nov. 2023, Art. no. 109982.

[36] M. S. Bonab, A. Ghaffari, F. S. Gharehchopogh, and P. Alemi, "A wrapper-based feature selection for improving performance of intrusion detection systems," *Int. J. Commun. Syst.*, vol. 33, no. 12, 2020, Art. no. e4434.

[37] R. Alsulami, B. Alqarni, R. Alshomrani, F. Mashat, and T. Gazdar, "IoT protocol-enabled IDS based on machine learning," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 6, pp. 12373–12380, 2023.

[38] I. Vaccari, G. Chiola, M. Aiello, M. Mongelli, and E. Cambiaso, "MQTTset, a new dataset for machine learning techniques on MQTT," *Sensors*, vol. 20, no. 22, p. 6578, 2020.

[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[40] S. Althubiti, E. Jones, and K. Roy, "LSTM for anomaly-based network intrusion detection," in *Proc. 28th Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, 2018, pp. 1–3.

[41] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using LSTM networks," *Comput. Ind.*, vol. 131, Oct. 2021, Art. no. 103498.

[42] F. Zhuang et al., "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[43] H. Liang, W. Fu, and F. Yi, "A survey of recent advances in transfer learning," in *Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT)*, 2019, pp. 1516–1523.

[44] E. Mahdavi, A. Fanian, A. Mirzaei, and Z. Taghiyarrenani, "ITL-IDS: Incremental transfer learning for intrusion detection systems," *Knowl.-Based Syst.*, vol. 253, Oct. 2022, Art. no. 109542.

[45] G. Zhao et al., "Lightweight intrusion detection model of the Internet of Things with hybrid cloud-fog computing," *Security Commun. Netw.*, vol. 2023, p. 6, Jan. 2023.

[46] O. Cohen, O. Malka, and Z. Ringel, "Learning curves for over-parametrized deep neural networks: A field theory perspective," *Phys. Rev. Res.*, vol. 3, Apr. 2021, Art. no. 23034. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.3.023034

[47] J. Gilmer et al., "A loss curvature perspective on training instabilities of deep learning models," in *Proc. Int. Conf. Learn. Rep.*, 2022, p. 46.

[48] S. Zavrak and M. Iskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108346–108358, 2020.

[49] G. Kumar, R. Kumar, K. Kumar, N. Sai, and M. Brahmaiah, "Deep residual convolutional neural network: An efficient technique for intrusion detection system," *Exp. Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121912.

[50] S. Singh, M. Lohakare, K. Sayar, and S. Sharma, "RecNN: A deep neural network based recommendation system," in *Proc. Int. Conf. Artif. Intell. Mach. Vis. (AIMV)*, 2021, pp. 1–5.

[51] F. S. Gharehchopogh, B. Abdollahzadeh, S. Barshandeh, and B. Arasteh, "A multi-objective mutation-based dynamic Harris Hawks optimization for botnet detection in IoT," *Internet Things*, vol. 24, Dec. 2023, Art. no. 100952.

[52] M. Sheikhan, Z. Jadidi, and A. Farrokhi, "Intrusion detection using reduced-size RNN based on feature grouping," *Neural Comput. Appl.*, vol. 21, pp. 1185–1190, Nov. 2012.

[53] H. Altunay and Z. Albayrak, "A hybrid CNN+ LSTM-based intrusion detection system for Industrial IoT networks," *Eng. Sci. Technol. Int. J.*, vol. 38, Feb. 2023, Art. no. 101322.

[54] K. Howlader et al., "Machine learning models for classification and identification of significant attributes to detect type 2 diabetes," *Health Inf. Sci. Syst.*, vol. 10, p. 2, Feb. 2022.
[55] A.A. Wardana, and P. Sukarno, "Taxonomy and survey of collaborative intrusion detection system using federated learning," *ACM Comput. Surveys*, vol. 57, no. 4, pp. 1–36, Apr. 2025.
[56] M. Sarhan, S. Layeghy, N. Moustafa, M. Gallagher, and M. Portmann, "Feature extraction for machine learning-based intrusion detection in IoT networks," *Digit. Commun. Netw.*, vol. 10, no. 1, pp. 205–216, 2024.
[57] A. Gad, A. Nashat, and T. Barkat, "Intrusion detection system using machine learning for vehicular ad hoc networks based on ToN-IoT dataset," *IEEE Access*, vol. 9, pp. 142206–142217, 2021.
[58] I. Tareq, B. Elbagoury, S. El-Regaily, and E. El-Horbaty, "Analysis of TON-IoT, UNW-NB15, and edge-IIoT datasets using DL in cybersecurity for IoT," *Appl. Sci.*, vol. 12, no. 19, p. 9572, 2022.
[59] Y. Soe, P. Santosa, and R. Hartanto, "DDoS attack detection based on simple ANN with SMOTE for IoT environment," in *Proc. 4th Int. Conf. Informat. Comput. (ICIC)*, 2019, pp. 1–5.
[60] I. Idrissi, M. Boukabous, M. Azizi, O. Moussaoui, and H. El Fadili, "Toward a deep learning-based intrusion detection system for IoT against botnet attacks," *IAES Int. J. Artif. Intell.*, vol. 10, no. 1, p. 110, 2021.
[61] M. Khan et al., "A deep learning-based intrusion detection system for MQTT enabled IoT," *Sensors*, vol. 21, no. 21, p. 7016, 2021.
[62] A. Sharma, V. Mansotra, and K. Singh, "Detection of Mirai botnet attacks on IoT devices using deep learning," *J. Sci. Res. Technol.*, vol. 1, no. 6, pp. 174–187, 2023.
[63] P. Vijayan and S. Sundar, "IoT intrusion detection system using ensemble classifier and hyperparameter optimization using tuna search algorithm," *J. Auton. Intell.*, vol. 7, no. 2, p. 962, 2024.
[64] M. Ge, N. Syed, X. Fu, Z. Baig, and A. Robles-Kelly, "Towards a deep learning-driven intrusion detection approach for Internet of Things," *Comput. Netw.*, vol. 186, Feb. 2021, Art. no. 107784.
[65] S. Khanday, H. Fatima, and N. Rakesh, "Implementation of intrusion detection model for DDoS attacks in lightweight IoT Networks," *Exp. Syst. Appl.*, vol. 215, Apr. 2023, Art. no. 119330.
[66] R. Devendiran, and A. Turukmane, "Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy," *Exp. Syst. Appl.*, vol. 245, Jul. 2024, Art. no. 123027.
[67] A. Fatani, M. Abd Elaziz, A. Dahou, M. Al-Qaness, and S. Lu, "IoT intrusion detection system using deep learning and enhanced transient search optimization," *IEEE Access*, vol. 9, pp. 123448–123464, 2021.
[68] E. Ciklabakkal, A. Donmez, M. Erdemir, E. Suren, M. Yilmaz, and P. Angin, "ARTEMIS: An intrusion detection system for MQTT attacks in Internet of Things," in *Proc. 38th Symp. Rel. Distrib. Syst. (SRDS)*, 2019, pp. 369–3692.
[69] T. Boppana, and P. Bagade, "GAN-AE: An unsupervised intrusion detection system for MQTT networks," *Eng. Appl. Artif. Intell.*, vol. 119, Mar. 2023, Art. no. 105805.
[70] S. Ullah et al., "TNN-IDS: Transformer neural network-based intrusion detection system for MQTT-enabled IoT networks," *Comput. Netw.*, vol. 237, Dec. 2023, Art. no. 110072.
[71] K. Stein, A. Mahyari, G. Francia, and E. El-Sheikh, "A transformer-based framework for payload malware detection and classification," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, 2024, pp. 105–111.
[72] A. Aguru and S. Erukala, "A lightweight multi-vector DDoS detection framework for IoT-enabled mobile health informatics systems using deep learning," *Inf. Sci.*, vol. 662, Mar. 2024, Art. no. 120209.
[73] I. A. Fares, M. A. Elaziz, A. O. Aseeri, H. S. Zied, A. G. Abdellatif, "TFKAN: Transformer based on Kolmogorov–Arnold networks for intrusion detection in IoT environment," *Egypt. Informat. J.*, vol. 30, Jun. 2025, Art. no. 100666.
[74] A. Shebl, S. Elsedimy, A. Ismail, A. Salama, and M. Herajy, "DCNN: A novel binary and multi-class network intrusion detection model via deep convolutional neural network," *EURASIP J. Inf. Security*, vol. 2024, p. 36, Dec. 2024.
[75] M. A. Elaziz, I. A. Fares, and A. Aseeri, "CKAN: Convolutional Kolmogorov–Arnold networks model for intrusion detection in IoT environment," *IEEE Access*, vol. 12, pp. 134837–134851, 2024.
[76] H. Kheddar, Y. Himeur, and A. I. Awad, "Deep transfer learning applications in intrusion detection systems: A comprehensive review," *J. Netw. Comput. Appl.*, vol. 220, Nov. 2023, Art. no. 103760.

IBRAHIM A. FARES received the B.Sc. degree in mathematics and computer science and the M.Sc. degree from Zagazig University, Egypt, in 2016 and 2021, respectively, where he is currently pursuing the Ph.D. degree. He is an Assistant Lecturer for teaching and mentoring students with the Department of Computer Science, Faculty of Science, Zagazig University. His research interests include metaheuristic techniques, IoT, cybersecurity, intrusion detection, deep learning, and transformers.

AHMED GAMAL ABDELLATIF IBRAHIM was born in Sharkia Governorate, Egypt, in 1985. He received the B.Sc. degree (with Hons.) in electronics and electrical engineering from the Air Defense College, Alexandria University in 2007, the master's degree in electronics and electrical engineering from Alexandria University in 2017, and the Ph.D. degree in electronics and communications engineering in 2022. He is a Dedicated Lecturer with the Department of Communications and Electronics Engineering, Air Defense College, Alexandria, Egypt. He has also demonstrated his commitment to expanding his knowledge and gaining international experience. In 2019, he became a Ph.D. student visitor with the prestigious Research Center of Geomatics (CIRGEO), University of Padua, Italy. This valuable experience allowed him to broaden his horizons and enrich his research pursuits. His research interests reflect his diverse background and multidisciplinary approach, including navigation, indoor positioning, tracking, filtering, information security, and image processing. Finally, He was a reviewer for several journals and major conferences.

MOHAMED ABD ELAZIZ received the B.S. and M.S. degrees in computer science and the Ph.D. degree in mathematics and computer science from Zagazig University, Egypt, in 2008, 2011, and 2014, respectively. From 2008 to 2011, he was an assistant lecturer for top scientists. He is an Associate Professor with Zagazig University. He is the author of more than 430 articles. His computer science department is one of the 2% influential scholars, which depicts the 100,000 top scientists in the world. He is one of the highest-cited researchers according to WOS 2022–2023. His research interests include metaheuristic techniques, medical applications, digital twins, renewable energy, security IoT, cloud computing, machine learning, signal processing, image processing, and evolutionary algorithms.

MANSOUR SHRAHILI received the bachelor's degree in mathematics from the Teachers College, King Saud University, Saudi Arabia, in 2003, the master's degree in mathematical statistics from the Department of Statistics and Operations Research, College of Science, King Saud University in 2008, and the Ph.D. degree in statistics from Salford University, U.K., in 2014. He was an Assistant Lecturer with the Department of Statistics and Operations Research, College of Science, King Saud University from 2015 to 2020. He was the Chairman of the Department of Statistics and Operation Research, College of Science, King Saud University from 2015 to 2017 and the Director of the Statistics and Information Department from 2018 to 2021. He has been an Associate Professor with King Saud University since 2021. He is the Vice Dean of the College of Science for Postgraduate Studies and Scientific Research. His research interests include mathematical statistics, applied statistics, reliability, characterization of distributions, and regression analysis.

ADHAM AHMED ELMAHALLAWY received the bachelor's degree (with Distinction and Hons.), the master's degree in electrical engineering, and the Doctor of Philosophy degree in electrical engineering from the Department of Electronics and Communications Engineering, University of Alexandria in 1994, 2001, and 2010, respectively. He is a highly qualified lecturer with a strong background in communications, electronics, and electro physics engineering. Also, he has published different research papers in highly-rated journals. He is a Lecturer with the Higher Institute of Engineering and Technology, King Mariout, Alexandria, Egypt.

RANA MUHAMMAD SOHAIB received the Ph.D. degree in electronics and electrical Engineering from the University of Glasgow, U.K., in 2024. He is currently working as a Lecturer with the Northumbria University, Newcastle upon Tyne, U.K. He has been actively involved in various Open RAN projects worth over £4 million. His research interests include radio resource management, AI-enabled wireless, energy-efficient, and vehicular networks.

MAHMOUD A. SHAWKY was born in Saudi Arabia in 1990. He received the B.Sc. degree in electronics and electrical engineering from the Air Defence College, Alexandria University, Alexandria, Egypt, in 2012, the M.Sc. (Eng.) degree in authentication mechanisms in computer network protocols from Alexandria University, and the Ph.D. degree from the James Watt School of Engineering, University of Glasgow, U.K. His research interests are in the areas of cryptography and number theory, digital signatures, authentication in wireless communications, and cyber security.

SYED TARIQ SHAH (Member, IEEE) received the master's and Ph.D. degrees from Sungkyunkwan University, South Korea, in 2015 and 2018, respectively, specializing in electrical and electronic engineering (telecommunications). He is a distinguished academic with a rich background in electrical and electronic engineering. His journey in academia has been marked by a blend of rigorous research and impactful teaching. He has joined the University of Essex as an Assistant Professor (Lecturer), following his tenure as a Postdoctoral Fellow with the University of Glasgow, U.K. Prior to this, he contributed significantly to the Department of Electrical Engineering, BUITEMS, Pakistan, in the capacity of an Associate Professor. His research domains are at the forefront of technological advancement, focusing on cutting-edge areas, such as 5G and beyond networks, open RAN, AI-enabled wireless networks, RF energy harvesting, and intelligent reflecting surfaces. His expertise is not only recognized in academia but also in the industry, as evidenced by his role as an Editor of the *Electronics* and his contributions as a reviewer for various prestigious IEEE Transactions, Letters, and Magazines.