# Intrusion Detection in IoT Using Traffic-to-Image Conversion and Lightweight Vision Transformers

Hajer Jamal Alatewish
*Faculty of Sciences*
*University of Tripoli*
Tripoli, Libya
h.alatewish@uot.edu.ly

Hala Shaari
*Faculty of Information Technology*
*University of Tripoli*
Tripoli, Libya
h.shaari@uot.edu.ly

***Abstract*— The ever-evolving nature of cyber-attacks within the Internet of Things (IoT) highlights the urgent need for developing advanced detection methods suitable for the limited resources of edge environments. Traditional detection methods face challenges in adapting to the rapidly changing and expanding landscape of IoT applications. This situation underscores the necessity for advanced detection methods that can achieve high detection accuracy while minimizing resource consumption. To address this issue, this paper proposes a lightweight framework for intrusion detection that combines the strengths of lightweight vision transformers with transfer learning. By incorporating a pre-trained MobileViT model into the domain of IoT intrusion detection, the proposed framework utilizes transfer learning to efficiently extract features and classify data, thereby enhancing detection performance. The approach involves converting traffic data into images by segmenting the data into blocks of successive samples and converting these blocks into grayscale images. Experimental results demonstrate the proposed framework's superior performance, achieving a high accuracy of 99.97% and an F1 score of 99.92% in a multi-class classification scenario on the Edge-IIoTset dataset, outperforming existing methods.**

***Keywords*— *Intrusion Detection, Traffic-to-image transformation, IoT, Lightweight Vision Transformers, TL***

## I. INTRODUCTION

The Internet of Things (IoT) stands as a cornerstone of the twenty-first century's technological revolution, profoundly transforming daily life and industrial operations by seamlessly connecting everyday devices to the internet. This technology establishes an expansive network of smart systems across diverse domains, ranging from smart homes and urban infrastructures to advanced manufacturing facilities, enabling unprecedented levels of automation, efficiency, and personalization.

With projections estimating growth to 39.6 billion connected devices by 2033 [1], this expansion has led to increased adoption of edge computing. Edge computing addresses challenges such as latency, bandwidth, and privacy by processing data locally, close to its source. This approach enables real-time capabilities and improved privacy, making it particularly beneficial for critical applications in smart cities, industrial settings, and healthcare systems [2] [3] [4].

Furthermore, this rapid expansion and the heterogeneous nature of IoT devices create significant security vulnerabilities, making IoT ecosystems highly vulnerable to cyber threats. These security risks extend beyond financial losses to potentially life-threatening scenarios, such as compromised sensors in autonomous vehicles [5]. As the adoption of IoT grows, addressing these vulnerabilities is crucial to ensure that security risks do not undermine the benefits of the technology. Intrusion Detection Systems (IDS) function as an essential security mechanism that monitors network traffic and device activities to detect unauthorized access and notify administrators of any suspicious behavior [6] [7].

Modern IDSs leverage deep learning techniques, which show significant potential for enhancing anomaly detection. Deep learning effectively learns complex patterns in IoT traffic and identifies zero-day attacks, where traditional methods often struggle [8] [9]. Deploying such systems at the network edge enables faster decision-making critical for sensitive applications. However, implementing advanced deep learning-based IDS on edge devices with limited resources is challenging because of their substantial computational and memory demands [10].

To address these challenges, an innovative and effective detection system is needed to overcome these limitations. This paper leverages the MobileViT architecture and transfer learning technique to provide a lightweight and robust intrusion detection system in IoT. MobileViT (Mobile Vision Transformer), a hybrid model combining convolutional neural networks and vision transformers, demonstrates high efficiency while maintaining high detection capabilities in many computer vision tasks, which makes it suitable for environments with limited resources [11] [12]. Transfer learning further addresses computational burdens by enabling the reuse of pre- trained models, significantly reducing training cost while improving detection performance [13]. Moreover, transfer learning can utilize existing knowledge even if the data characteristics of the source and target domains are significantly different.

The main contributions of this paper can be summarized as follows:

- The tabular data is converted into grayscale images by segmenting the data into overlapped blocks, enabling a suitable representation to enhance the learning process.
- A lightweight intrusion detection framework is proposed for identifying cyberattacks in IoT edge environments. This framework utilizes an advanced pretrained Mobile Vision Transformer (MobileViT) architecture.
- The proposed framework undergoes a comprehensive performance evaluation using the Edge-IIoTset dataset. This evaluation examines the framework's effectiveness in a multiclass classification scenario and compares its performance to a baseline model and other existing methods. The

findings provide valuable insights into the system's effectiveness in classifying complex attack patterns.

The remainder of this paper is organized as follows. Section II provides an overview of related research on IoT intrusion detection. Section III introduces the proposed framework. Section IV presents the experimental results and briefly discusses the findings. Section V concludes the paper and suggests directions for future research.

## II. Related Work

Deep learning has emerged as a powerful paradigm for advancing intrusion detection in IoT environments. This section provides an overview of some of the state-of-the-art approaches that leverage deep learning techniques to enhance attack detection.

A recent study [14] introduced L2D2, an intrusion detection system in Internet of Medical Things (IoMT) environments, utilizing an enhanced version of the LSTM deep learning algorithm. The model leverages stacked LSTMs to capture temporal dependencies in sequential traffic data and is evaluated on the CICIoMT2024 dataset, achieving an accuracy of 98% across multiple attack categories. Authors in [15] employed seven deep learning models, including the Transformer, to analyze network traffic and detect potential intrusions through binary and multi-class classifications, utilizing the CIC-IoT-2023 dataset. In the binary classification task, the accuracy of the Transformer model was lower than that of the DNN and the CNN + LSTM hybrid models. However, it outperformed these models in multi-class classification, achieving an accuracy of 99.40%.

In addition, transformed traffic data into visual images by reshaping the data to create a 2D array, which was then used as input for several deep learning models, including CNN, MobileNet, ResNet50, and VGG16, to classify benign and malicious activities in [16]. The TON IoT dataset was used for evaluation, and the results showed an accuracy of 99.1%. Another study [17] proposed a lightweight ensemble transfer learning model using CNNs to detect common attacks in cloud IoT environments. They leveraged five pre-trained CNNs, which are VGG16, VGG19, Inception, MobileNet, and EfficientNets. They convert the feature vectors to images via Quantile Transformer for CNN compatibility. The models were optimized using Bayesian optimization (BO-TPE), with the top three models combined through model averaging. Evaluated on CIC-IDS2017 and CSE-CICIDS2018 datasets, the method achieved an accuracy of 100% and 99.98%.

Moreover, authors in [18] highlighted the feasibility of transfer learning in environments with limited computational resources and proposed a transfer learning-based framework for detecting network intrusion on the Internet of Battlefield Things (IoBT). The authors trained a one-dimensional CNN combined with a random forest model using the UNSW-NB15 and CICIDS2017 datasets. The experimental results showed an attack detection accuracy of 96.80% using 5,000 training samples. The training time on edge devices was approximately 67 seconds.

Furthermore, a study [19] proposed a ViT-based network intrusion detection system that converts raw network flows into RGB images, enabling the model to capture global spatial traffic patterns via self-attention. Evaluated on CICIDS2017 and UNSW-NB15 datasets, the system achieved 96% and 98% accuracy for binary and multiclass classification. On the other hand, authors in [20] developed an Intrusion Detection System using tree-based Support Vector Machines (SVM), ensemble methods, Long Short-Term Memory (LSTM) networks, and Vision Transformers (ViT), with hyperparameters optimized through Bayesian optimization. The models evaluated using the NSL-KDD dataset achieved 99.90% accuracy with the Random Forest and Ensemble Bagged Tree models. The LSTM model reached an accuracy of 99.97%, while the ViT model achieved a validation accuracy of 78.70%.

Another recent study proposed an optimized intrusion detection system (IDS) for heterogeneous IIoT networks using deep transfer learning (DTL) [21]. The system incorporates seven pre-trained CNN architectures (Xception, VGG16, VGG19, Inception, InceptionResNetV2, EfficientNetB7, and EfficientNetV2L) with hyperparameters fine-tuned via a genetic algorithm (GA). The top five models were combined using a bootstrap aggregation ensemble. Tested on the Edge-IIoTset dataset, the model achieved 100% accuracy in detecting 14 attack classes. Also, authors in [22] proposed a hybrid IDS for IoT that combines autoencoders, LSTMs, and CNNs to enhance feature extraction and capture both temporal and spatial characteristics of network traffic. To address class imbalance and improve the detection of minority attack types, SMOTE was applied. Evaluated on the CICIoT2023 dataset, the model achieved 99.15% accuracy and an F1-score of 99.19%, outperforming several existing IDS approaches across diverse attack categories.

Moreover, a combined dataset framework for intrusion detection systems (IDS) using a hybrid PCA-Transformer model has been proposed in [23]. In this approach, PCA performs feature extraction and dimensionality reduction, while the Transformer handles classification. Class imbalance was addressed with class weights, ADASYN, and ENN, and enhanced preprocessing was applied to the vertically concatenated CSE-CIC-IDS2018 and CICIDS2017 datasets. Experimental evaluation on the combined dataset demonstrated that the model achieved 99.80% accuracy for binary classification and 99.28% for multi-class classification. Differently, an intrusion detection system using a transformer-based transfer learning approach with BERT for extracting data insights and SMOTE to balance minority attacks has been developed by authors in [24]. A hybrid CNN-LSTM model detects attack types from the extracted deep features. Tested on UNSW-NB15, CIC-IDS2017, and NSL-KDD datasets, IDS-INT achieved 99.21% accuracy.

A novel Network Intrusion Detection System designed to secure IoT-enabled Smart Agriculture has been introduced in [25]. It first converts tabular network traffic data into images. Then it leverages deep transfer learning by fine-tuning five CNN models (MobileNet, EfficientNet, Xception, VGG19, and Inception) pre-trained on the ImageNet dataset. Furthermore, the Black Kite Algorithm (BKA) was utilized to optimize the hyperparameters of these models, significantly enhancing their performance. The top three optimized models are then combined using a confidence averaging ensemble strategy to produce the final classification. Evaluated on three datasets, ToN-IoT, WSN-DS, and Edge-IIoTset, the method achieved an accuracy exceeding 99% on all of them. Also, authors in [26] proposed a BERT-based model for cyber threat detection in IoT networks called SecurityBERT. It uses a Privacy-Preserving Fixed-Length Encoding (PPFLE) and a Byte-level Byte-Pair Encoder (BBPE) Tokenizer to represent

structured network data. SecurityBERT achieved 98.2% accuracy using the Edge-IIoTset dataset, SecurityBERT achieved 98.2% accuracy across 14 attack types, with an average inference time under 0.15 seconds and a model size of 16.7MB, making it suitable for real-time, resource-limited IoT devices. A metamorphic malware detection model in IoT devices has been introduced by [27]. They used six pre-trained models, including VGG16, InceptionV3, CNN, ResNet50, MobileNet, and EfficientNetB0, on the Malimg malware image dataset. The findings of their study indicated that the EfficientNetB0 model outperformed all other machine learning and deep learning models, achieving an accuracy of 99% and an F1-score of 97%.

Moreover, authors in [28] proposed HiViT-IDS, an intrusion detection based on Vision Transformer to address the trade-off found in traditional ML/DL and deep transfer learning approaches between accuracy and computational cost. The method converts traffic data into RGB images and leverages self-attention in ViT for efficient classification. HiViT-IDS achieved 99.7% and 100% accuracy, using ToN-IoT and Edge-IIoTset datasets, respectively, while reducing training time compared to state-of-the-art DTL models. Also, a study [29] proposed sSecure Net, a hybrid CNN-LSTM intrusion detection framework that combines CNN's spatial feature extraction with LSTM's temporal learning. The authors preprocess the data using statistical filtering techniques to remove noise and outliers. Evaluated on the comprehensive Edge-IIoTset dataset, the model achieved a high accuracy of 94.99% and 94.50% of F1-score.

## III. The Prposed Framework

The general research approach for the suggested framework is presented in this section. It begins by describing the dataset, followed by an overview of data preprocessing and data conversion. Finally, the section provides a brief discussion of the transfer learning strategy. A block representation of the suggested framework may be seen in Figure 1.

### *A. Dataset*

This work utilized the modern and comprehensive Edge-IIoTset dataset. To create the Edge-IIoTset dataset, Ferrag et al. [30] used a specialized IoT/IIoT testbed to simulate a varied and realistic environment. The testbed had more than 10 different kinds of IoT devices, including sensors for temperature, humidity, ultrasonic, pH, water level, heart rate, soil moisture, and flames, in addition to a variety of protocols, sensors, and cloud/edge setups. 14 different attack types, as well as normal traffic, are included in the dataset. The distribution of data categories in the Edge-IIoTset dataset is presented in Figure 2.

### *B. Dataset Preprocessing*

For deep learning models to train and perform at their best, effective data preparation is crucial. The raw Edge-IIoTset dataset consists of 63 features and 15 classes. In the first stage of preprocessing, 309530 duplicate rows were eliminated. 16 unnecessary features that did not significantly impact output predictions were removed in the following step. These features included metadata such as 'frame.time','ip.dst_host', 'ip.src_host', 'arp.src.proto_ipv4', 'http.file_data', 'arp.dst.proto_ipv4', 'http.request.full_uri', 'http.request.uri.query', 'mqtt.msg', 'icmp.transmit_timestamp', 'tcp.payload', 'tcp.options', 'tcp.srcport', 'udp.port', 'tcp.dstport', and 'Attack_label'. Then, using dummy encoding, feature columns with object data types were transformed into numerical values. The final preprocessed dataset contains 96 columns and 1,909,671 rows.

### *C. Date Conversion*

To ensure compatibility with the MobileViT model and utilize its superior performance on image data, converting the

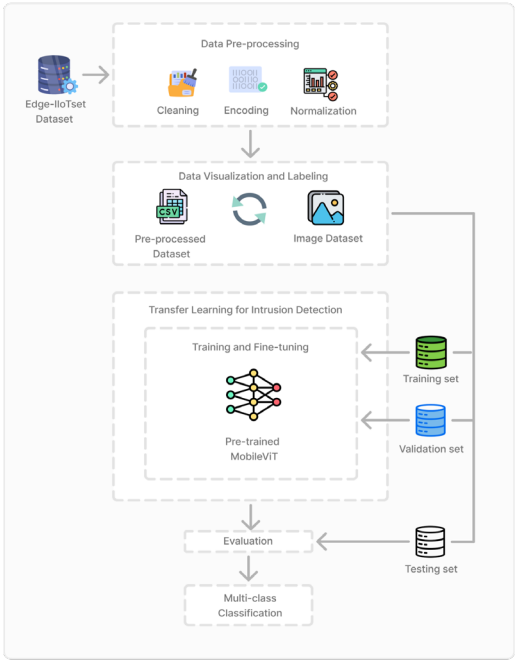


Fig. 1. The Proposed Framework of Intrusion Detection.

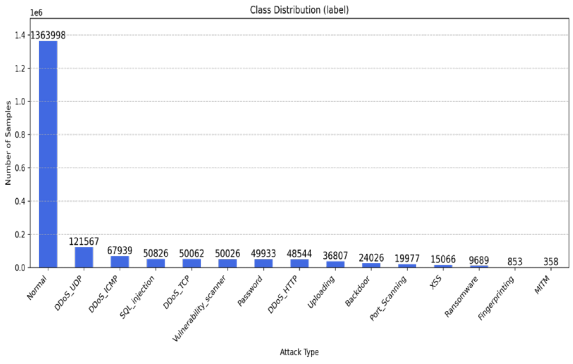


Fig. 2. The Distribution of Attack Categories in Edge-IIoTset Dataset.

tabular data from the Edge-IIoTset dataset into image representations is essential. In this data conversion procedure, normalization is the initial stage. The values in the dataset must be translated to the image pixel values, which range from 0 to 255. We use the Quantile Transformation method [31] to rescale the network traffic data onto this predefined scale, ensuring it is suitable for processing with MobileViT. Following feature scaling, the samples are divided into blocks according to the dataset's feature size and the timestamps. There are 95 unique features in the preprocessed Edge-IIoTset dataset. Overlapping data blocks are created using a sliding window technique with a stride of 1 and a sequence length of 95. 95 consecutive samples, each with 95 features, make up each block, which is then transformed into a 95x95 grayscale

image. This timestamps x features matrix keeps the temporal order of each feature throughout the sequence. To assign a single representative label to each generated image, we employ the following labeling strategy. Images that contain all normal samples are referred to be "Normal." If there are any attack samples in the image, they are labeled with the specific attack type. The image is labeled according to the most common attack type in the sequence when there are many attack types present. Images are rescaled into an input format that is appropriate for the MobileViT model as the last stage of the data conversion process. While the generated images are 95x95 grayscale images, the pre-trained MobileViT model requires images in the 256x256x3 format. Therefore, we resize the images to 256x256 and replicate them across RGB channels. Representative examples of the generated images are shown in Figure 3.

### *D. Transfer Learning with MobileViT*

A MobileViT is a recent lightweight hybrid vision transformer that has demonstrated strong performance in various vision tasks [12]. A typical MobileViT model comprises several components: standard convolution layers,

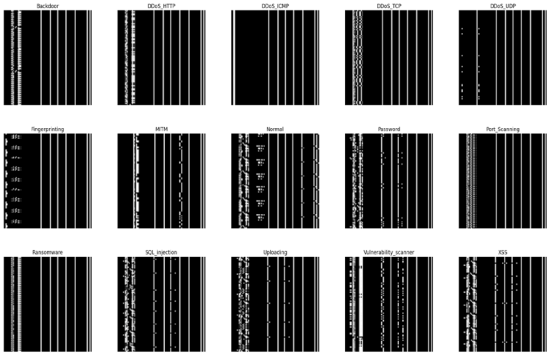


Fig. 3. Distribution of Edge-IIoTset Generated Images.

MobileNetV2 blocks, MobileViT blocks, global pooling, and a fully connected layer. The convolution layers facilitate automatic feature extraction of local spatial information from the input images. The MobileViT component utilizes a standard transformer to divide images into patches, processing them as sequences and applying self-attention to capture the global relationships among all patches. The pooling layer reduces complexity while preserving important information by leveraging local correlations. In order to produce predictions for the input images, the fully connected layer lastly combines all of the retrieved information. While keeping the number of parameters low, this hybrid architecture combines the benefits of lightweight convolutional neural networks (CNNs) with the global context modeling capabilities of vision transformers (ViTs).

Reusing and adjusting pre-trained weights from one dataset to another is known as transfer learning (TL), a machine learning approach that reduces training costs and potentially enhances detection performance [32], [33]. Deep learning models are structured as layered architectures that learn different features at various levels. The initial layers capture low-level, general features, progressively narrowing down to high-level, task-specific features as the network deepens. The final layers, usually fully connected layers, produce the output [32]. A pre-trained model can be employed as a feature extractor by freezing its backbone while only training a new classifier head. It can also be utilized by unfreezing some or all of the pre-trained layers and fine-tuning them on the new dataset.

For the core of the proposed framework, MobileViT-xx-small is selected. The ImageNet dataset was used to pre-train this architecture, and it has demonstrated encouraging outcomes for image classification tasks. In this approach, the entire pre-trained backbone was frozen and employed as a feature extractor. The backbone's output was fed into a custom classifier head, incorporating two hidden layers, each followed by GELU activation, Layer Normalization, Dropout, and an output layer tailored for 15 classes.

## IV. Experiments and Performance Analysis

The conducted experiment and performance evaluation of the proposed framework are presented in this section. The implementation environment, performance metrics, and the hyperparameters of the training process are described first. Following that, the section represents an analysis of the results and a comparison of the proposed scheme with state-of-the-art methods.

### *A. Implementation environment and performance metrics*

PyTorch for model development, Pandas for data manipulation, Scikit-learn for preprocessing and metrics computations, and MLflow for experiment tracking were used in the implementation and evaluation of the framework. A personal workstation equipped with an Intel Core i7 CPU and 16GB of RAM was used for the experiments.

To enable exact comparisons with results from different models, we use traditional measurements like as accuracy, precision, recall, F1-score, and confusion matrices. We consider the model's size and parameter count while evaluating its efficacy. These metrics are critical for comprehensively evaluating the model's performance and capabilities.

### *B. Hyperparameters of the Training Process*

Using the hyperparameters listed in Table I, the MobileViT model was trained.

**TABLE I.** Hyperparameters of the of MobileViT model

| Hyperparameter | Value |
|---|---|
| Learning Rate | 0.0001 |
| Epochs | 25 |
| Batch Size | 32 |
| Optimizer | AdamW |
| Weight decay | 0.01 |
| Dropout rate | 0.3 |

### *C. Performance Analysis*

The experimentation and performance analysis were conducted as follows. We trained the model for 25 epochs with the selected hyperparameters, selecting the final model based on the maximum macro F1-score achieved on the validation set. Metrics for accuracy, precision, recall, F1-score, macro average, and the confusion matrix were then obtained as part of the model's prediction findings. Figure 4

displays the training process' accuracy and loss history. Table II displays the model's prediction results in summary form.

On the unseen test set, the suggested model had a macro F1-score of 99.92% and an overall accuracy of 99.97%. The strong and balanced performance is indicated by the high macro F1-score, which effectively classified 14 different forms of cyberattacks, including the rarest ones like fingerprinting and MITM. Figure 5's study of the confusion matrix demonstrates the model's remarkable performance in spite of the data's imbalance.

Notably, no attack samples were misclassified as normal traffic, which signifies an extremely low false-negative rate. We assessed the suggested model's efficacy by contrasting its results with those of a baseline model (MobileNetV2) that used the identical classifier architecture and data format. MobileNet has been successfully implemented for network intrusion detection in the existing literature and showed strong performance with TL. MobileNetV2 also represents a pure Lightweight CNN that MobileViT hybrid architecture leverages its capabilities. Table III presents the findings of the comparison. In comparison to the baseline model, the proposed model showed highly competitive results.

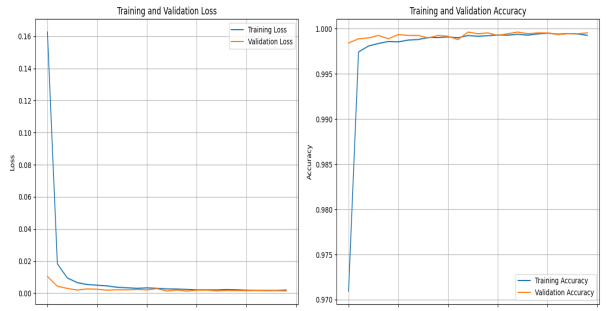


Fig. 4. Training Results of MobileViT.

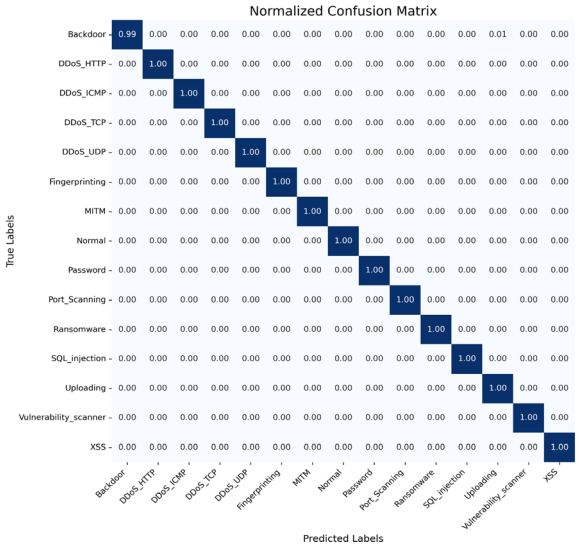


Fig. 5. Confusion Matrix of the Proposed Approach.

The performance gap between the two models was minimal across evaluation metrics. Notably, the proposed model achieved these results while maintaining parameter efficiency and integrating advanced techniques.

Additionally, a comparison with state-of-the-art models is carried out and shown in Table IV. The proposed model obtained 99.97%, 99.93%, 99.92%, and 99.92% of accuracy, precision, recall, and F1-score, respectively. The proposed model performs better than all other models on every performance metric, even if these metrics are slightly lower than those of DTL-IDS [21] and HiViT-IDS [28]. In terms of efficiency, these models consist of ensembles of multiple CNNs, LLMs, hybrid CNN-LSTM, and traditional ViT. These models may introduce complexity and tend to have a larger parameter count than the proposed model, which is designed to maintain efficiency with fewer parameters and reduced complexity. This makes the proposed model more suitable for real-time detection in edge environments.

**TABLE II.** PERFORMANCE SCORES OF THE TRAINED MODEL

| Class | Precision | Recall | F1-score |
|---|---|---|---|
| Backdoor | 1.0 | 0.9942 | 0.9971 |
| DDoS_HTTP | 1.0 | 1.0 | 1.0 |
| DDoS_ICMP | 1.0 | 1.0 | 1.0 |
| DDoS_TCP | 1.0 | 0.9968 | 0.9984 |
| DDoS_UDP | 1.0 | 1.0 | 1.0 |
| Fingerprinting | 1.0 | 1.0 | 1.0 |
| MITM | 1.0 | 1.0 | 1.0 |
| Normal | 1.0 | 1.0 | 1.0 |
| Password | 0.9968 | 1.0 | 0.9984 |
| Port Scanning | 1.0 | 1.0 | 1.0 |
| Ransomware | 1.0 | 1.0 | 1.0 |
| SQL Injection | 0.9968 | 1.0 | 0.9984 |
| Uploading | 0.9958 | 1.0 | 0.9979 |
| Vulnerability Scanner | 1.0 | 0.9968 | 0.9984 |
| XXS | 1.0 | 1.0 | 1.0 |
| **Macro Average** | 0.9993 | 0.9992 | **0.9992** |
| **Accuracy** | **0.9997 (99.97%)** | | |

**TABLE III.** COMPARISON OF THE PROPOSED MODEL PERFORMANCE AGAINST THE BASELINE MODEL

| Model | Accuracy (%) | F1-score (%) | Model Size | No. Total Parameters |
|---|---|---|---|---|
| MobileNetV2 | 99.99 | 99.94 | 10.11 MB | ~ 2.5 Million |
| **Proposed Model** | **99.97** | **99.92** | **4.24 MB** | **~ 1 Million** |

**TABLE IV.** PERFORMANCE EVALUATION OF MAINSTREAM MODELS AND THE PROPOSED MODEL ON THE EDGE_IIOTSET DATASET

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|
| DTL-IDS [21] | 100 | 100 | 100 | 100 |
| CBCTL-IDS [25] | 99.90 | 99.88 | 99.90 | 99.89 |
| SecurityBERT [26] | 98.20 | 87.0 | 84.0 | 84.0 |
| HiViT-IDS [28] | 100 | 100 | 100 | 100 |
| sSecure Net [29] | 94.99 | 96.21 | 94.95 | 94.50 |
| **Proposed Model** | **99.97** | **99.93** | **99.92** | **99.92** |

## V. CONCLUSION

Using the MobileViT architecture and transfer learning, this work presents a robust intrusion detection system that offers an advanced approach to secure resource-constrained IoT and edge environments. The model outperformed several existing methods with an accuracy of 99.97% when evaluated on the Edge-IIoTset dataset. Using MobileViT's hybrid capabilities, the system effectively detected a wide spectrum of IoT intrusions with excellent accuracy, recall, and F1 scores. Furthermore, the proposed method demonstrates its competitive advantage in edge scenarios by significantly

reducing training time and resource consumption. Although the use of hybrid vision transformers in cybersecurity has advanced significantly as a result of this study, there are a number of possible avenues for further investigation that might build on these encouraging results. Increasing the resilience of the proposed approach against different kinds of threats, such as adversarial attacks and more sophisticated threats, is one potential direction. The suggested model must also be updated and retrained using the most recent real-world datasets in order to retain its efficacy, considering the dynamic nature of cyberthreats.

## REFERENCES


[1] "IoT connections worldwide 2022-2033," Statista. Accessed: Nov. 03, 2024.[Online].Available:https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[2] P. Grzesik and D. Mrozek, "Combining Machine Learning and Edge Computing: Opportunities, Challenges, Platforms, Frameworks, and Use Cases," *Electronics*, vol. 13, no. 3, Art. no. 3, Jan. 2024, doi: 10.3390/electronics13030640.

[3] N. Quy, L. Ngoc, N. Ban, V.-H. Nguyen, and Q. Vu Khanh, "Edge Computing for Real-Time Internet of Things Applications: Future Internet Revolution," *Wirel. Pers. Commun.*, vol. 132, pp. 1–30, July 2023, doi: 10.1007/s11277-023-10669-w.

[4] F. C. Andriulo, M. Fiore, M. Mongiello, E. Traversa, and V. Zizzo, "Edge Computing and Cloud Computing for Internet of Things: A Review," *Informatics*, vol. 11, no. 4, Art. no. 4, Dec. 2024, doi: 10.3390/informatics11040071.

[5] P. Malhotra, Y. Singh, P. Anand, D. K. Bangotra, P. K. Singh, and W.-C. Hong, "Internet of Things: Evolution, Concerns and Security Challenges," *Sensors*, vol. 21, no. 5, Art. no. 5, Jan. 2021, doi: 10.3390/s21051809.

[6] A. Heidari and M. A. Jabraeil Jamali, "Internet of Things intrusion detection systems: a comprehensive review and future directions," *Clust. Comput.*, vol. 26, no. 6, pp. 3753–3780, Dec. 2023, doi: 10.1007/s10586-022-03776-z.

[7] E. Gyamfi and A. Jurcut, "Intrusion Detection in Internet of Things Systems: A Review on Design Approaches Leveraging Multi-Access Edge Computing, Machine Learning, and Datasets," *Sensors*, vol. 22, no. 10, p. 3744, May 2022, doi: 10.3390/s22103744.

[8] H. Liao *et al.*, "A Survey of Deep Learning Technologies for Intrusion Detection in Internet of Things," *IEEE Access*, vol. 12, pp. 4745–4761, 2024, doi: 10.1109/ACCESS.2023.3349287.

[9] M. A. Alsoufi *et al.*, "Anomaly-Based Intrusion Detection Systems in IoT Using Deep Learning: A Systematic Literature Review," *Appl. Sci.*, vol. 11, no. 18, Art. no. 18, Jan. 2021, doi: 10.3390/app11188383.

[10] D. Liu, H. Kong, X. Luo, W. Liu, and R. Subramaniam, "Bringing AI to edge: From deep learning's perspective," *Neurocomputing*, vol. 485, pp. 297–320, May 2022, doi: 10.1016/j.neucom.2021.04.141.

[11] H.-I. Liu *et al.*, "Lightweight Deep Learning for Resource-Constrained Environments: A Survey," *ACM Comput. Surv.*, p. 3657282, May 2024, doi: 10.1145/3657282.

[12] S. Mehta and M. Rastegari, "MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer," Mar. 04, 2022, *arXiv*: arXiv:2110.02178. doi: 10.48550/arXiv.2110.02178.

[13] W. Wu *et al.*, "Deep Transfer Learning Techniques in Intrusion Detection System-Internet of Vehicles: A State-of-the-Art Review," *Comput. Mater. Contin.*, vol. 80, pp. 1–29, Aug. 2024, doi: 10.32604/cmc.2024.053037.

[14] G. Akar, S. Sahmoud, M. Onat, Ü. Cavusoglu, and E. Malondo, "L2D2: A Novel LSTM Model for Multi-Class Intrusion Detection Systems in the Era of IoMT," *IEEE Access*, vol. 13, pp. 7002–7013, 2025, doi: 10.1109/ACCESS.2025.3526883.

[15] S.-M. Tseng, Y.-Q. Wang, and Y.-C. Wang, "Multi-Class Intrusion Detection Based on Transformer for IoT Networks Using CIC-IoT-2023 Dataset," *Future Internet*, vol. 16, no. 8, Art. no. 8, Aug. 2024, doi: 10.3390/fi16080284.

[16] A. Hattak, F. Martinelli, F. Mercaldo, and A. Santone, "Leveraging Deep Learning for Intrusion Detection in IoT Through Visualized Network Data:," in *Proceedings of the 21st International Conference on Security and Cryptography*, Dijon, France: SCITEPRESS - Science and Technology Publications, 2024, pp. 722–729. doi: 10.5220/0012768400003767.

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

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

[19] C. M. K. Ho, K.-C. Yow, Z. Zhu, and S. Aravamuthan, "Network Intrusion Detection via Flow-to-Image Conversion and Vision Transformer Classification," *IEEE Access*, vol. 10, pp. 97780–97793, 2022, doi: 10.1109/ACCESS.2022.3200034.

[20] L. Sana *et al.*, "Securing the IoT Cyber Environment: Enhancing Intrusion Anomaly Detection With Vision Transformers," *IEEE Access*, vol. 12, pp. 82443–82468, 2024, doi: 10.1109/ACCESS.2024.3404778.

[21] S. Latif, W. Boulila, A. Koubaa, Z. Zou, and J. Ahmad, "DTL-IDS: An optimized Intrusion Detection Framework using Deep Transfer Learning and Genetic Algorithm," *J. Netw. Comput. Appl.*, vol. 221, p. 103784, Jan. 2024, doi: 10.1016/j.jnca.2023.103784.

[22] B. Susilo, A. Muis, and R. F. Sari, "Intelligent Intrusion Detection System Against Various Attacks Based on a Hybrid Deep Learning Algorithm," *Sensors*, vol. 25, no. 2, p. 580, Jan. 2025, doi: 10.3390/s25020580.

[23] H. Kamal and M. Mashaly, "Combined Dataset System Based on a Hybrid PCA–Transformer Model for Effective Intrusion Detection Systems," *AI*, vol. 6, no. 8, p. 168, Aug. 2025, doi: 10.3390/ai6080168.

[24] F. Ullah, S. Ullah, G. Srivastava, and J. C.-W. 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, Feb. 2024, doi: 10.1016/j.dcan.2023.03.008.

[25] H. Zhou, H. Zou, P. Zhou, Y. Shen, D. Li, and W. Li, "CBCTL-IDS: A Transfer Learning-Based Intrusion Detection System Optimized With the Black Kite Algorithm for IoT-Enabled Smart Agriculture," *IEEE Access*, vol. 13, pp. 46601–46615, 2025, doi: 10.1109/ACCESS.2025.3550800.

[26] M. A. Ferrag *et al.*, "Revolutionizing Cyber Threat Detection with Large Language Models: A privacy-preserving BERT-based Lightweight Model for IoT/IIoT Devices," *IEEE Access*, vol. PP, Feb. 2024, doi: 10.1109/ACCESS.2024.3363469.

[27] F. Habib, S. H. Shirazi, K. Aurangzeb, A. Khan, B. Bhushan, and M. Alhussein, "Deep Neural Networks for Enhanced Security: Detecting Metamorphic Malware in IoT Devices," *IEEE Access*, vol. 12, pp. 48570–48582, 2024, doi: 10.1109/ACCESS.2024.3383831.

[28] H. Zhou, H. Zou, W. Li, D. Li, and Y. Kuang, "HiViT-IDS: An Efficient Network Intrusion Detection Method Based on Vision Transformer," *Sensors*, vol. 25, no. 6, Art. no. 6, Jan. 2025, doi: 10.3390/s25061752.

[29] O. A. Hussain, Z. Chen, and H. Zhu, "sSecure Net: A Hybrid CNN-LSTM-based Intrusion Detection System for Securing IoT Networks," in *Proceedings of the 4th International Conference on Computer, Artificial Intelligence and Control Engineering*, New York, NY, USA: Association for Computing Machinery, 2025, pp. 537–544. Accessed: Aug. 25, 2025. [Online]. Available: https://doi.org/10.1145/3727648.3727736

[30] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022, doi: 10.1109/ACCESS.2022.3165809.

[31] S.-F. Lokman, A. T. Othman, M. H. A. Bakar, and S. Musa, "The Impact of Different Feature Scaling Methods on Intrusion Detection for in-Vehicle Controller Area Network (CAN)," in *Advances in Cyber Security*, M. Anbar, N. Abdullah, and S. Manickam, Eds., Singapore: Springer, 2020, pp. 195–205. doi: 10.1007/978-981-15-2693-0_14.

[32] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies*, vol. 11, no. 2, Art. no. 2, Apr. 2023, doi: 10.3390/technologies11020040.

[33] "Attention-Driven Transfer Learning Model for Improved IoT Intrusion Detection." Accessed: Nov. 03, 2024. [Online]. Available: https://www.mdpi.com/2504-2289/8/9/116.