Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.051535

ARTICLE

# Abnormal Traffic Detection for Internet of Things Based on an Improved Residual Network

**Tingting Su[1], Jia Wang[1,*], Wei Hu[2,*], Gaoqiang Dong[1] and Jeon Gwanggil[3]**

[1]School of Computer Science and Technology, Xinjiang University, Urumqi, 830046, China

[2]School of Cyber Science and Engineering, Jinling Institute of Technology City, Nanjing, 210000, China

[3]College of Information Technology, Incheon National University, Incheon, 22012, Korea

*Corresponding Authors: Jia Wang. Email: jw1024@xju.edu.cn; Wei Hu. Email: huwei@jit.edu.cn

Received: 07 March 2024 Accepted: 19 April 2024 Published: 20 June 2024

**ABSTRACT**

Along with the progression of Internet of Things (IoT) technology, network terminals are becoming continuously more intelligent. IoT has been widely applied in various scenarios, including urban infrastructure, transportation, industry, personal life, and other socio-economic fields. The introduction of deep learning has brought new security challenges, like an increment in abnormal traffic, which threatens network security. Insufficient feature extraction leads to less accurate classification results. In abnormal traffic detection, the data of network traffic is high-dimensional and complex. This data not only increases the computational burden of model training but also makes information extraction more difficult. To address these issues, this paper proposes an MD-MRD-ResNeXt model for abnormal network traffic detection. To fully utilize the multi-scale information in network traffic, a Multi-scale Dilated feature extraction (MD) block is introduced. This module can effectively understand and process information at various scales and uses dilated convolution technology to significantly broaden the model's receptive field. The proposed Max-feature-map Residual with Dual-channel pooling (MRD) block integrates the maximum feature map with the residual block. This module ensures the model focuses on key information, thereby optimizing computational efficiency and reducing unnecessary information redundancy. Experimental results show that compared to the latest methods, the proposed abnormal traffic detection model improves accuracy by about 2%.

**KEYWORDS**

Abnormal network traffic; deep learning; residual network; multi-scale feature extraction; max-feature-map

## 1 Introduction

With the continuous advancement of Internet of Things (IoT) technology, network terminals have become more intelligent. It has been widely applied in manufacturing, healthcare, and transportation, profoundly influencing both our work and daily lives. By the end of June 2022, global IoT connections rose to 14.4 billion [1]. However, the rapid development and wide application of IoT have made IoT security face an increasingly severe situation. For example, in 2016, Dyn Inc. in the United States was attacked by a Distributed Denial of Service (DDoS), which infected nearly 65,000 IoT devices and

led to the disruption of services of many Uniform Resource Locator (URL) [2]. IoT attacks pose a significant challenge to IoT security, and it is a crucial component of IoT security.

In smart manufacturing factories, many devices such as robotic arms, sensors, and control systems are interconnected. Meanwhile, vast different kinds of production figures, equipment status reports, and environmental monitoring information are generated, which contain many redundancies and irrelevant data. In other words, these data increase computational complexity and complicate the monitoring and detection of abnormal traffic. Furthermore, the distinguishing between abnormal and regular patterns becomes challenging due to the variety and intricacy of the network traffic patterns. In general, the precision of feature selection directly impacts the effectiveness of abnormal detection. Therefore, it is crucial to identify the feature representation of abnormal traffic.

In recent years, machine learning has provided new ways to detect abnormal traffic with its powerful data analysis and pattern recognition capabilities. Existing machine learning methods rely on manual feature extraction, which is not only inefficient in high-dimensional space but also easily leads to insufficient and inaccurate features. Considering the significant advantages of deep learning in automatic feature extraction and pattern recognition, current abnormal traffic detection technology is mainly based on deep learning methods. Although deep learning has advantages in automatic feature extraction. It may lead to incomplete feature extraction and computational inefficiency when processing multidimensional data.

Although machine learning and deep learning have made certain progress in abnormal traffic detection, they still have problems such as incomplete feature extraction, low computational efficiency, and insufficient accuracy when processing multi-dimensional complex data. In order to detect complex features of different scale data in abnormal traffic. This paper proposes a Multi-scale Dilated feature extraction (MD) block. At the same time, in order to effectively process high-dimensional data and maintain key features, a Max-feature-map Residual with Dual-channel pooling (MRD) block is proposed. Therefore, we propose an MD-MRD-ResNeXt model, to solve the problem of insufficient feature extraction in abnormal traffic detection.

The main contributions of this paper are described as follows:

(1) This paper introduces the MD block, a feature integration mechanism that addresses the limitations of existing deep learning models in capturing the multidimensionality of data. Unlike methods that rely solely on single-scale feature extraction, MD block combines multi-scale feature extraction with dilated convolution techniques. MD block can significantly enlarge the model's receptive field, enabling it to more effectively understand and process information across various scales for capturing details and broad context.

(2) In accordance with the diverse features and complex relationships in high-dimensional data of IoT data, a MRD block is proposed to combine maximum feature mapping and residual blocks. In order to accurately extract key features and maintain their continuity and stability in deep networks. The method alleviates overfitting and improves generalization ability. Furthermore, computational efficiency is optimized through the use of parallel dual-channel pooling technology.

The paper is organized as follows: Section 2 reviews related work. Section 3 introduces the details of the proposed methods. Section 4 reports the experimental results and analyses. Section 5 concludes our work and looks forward to future research directions.

## 2 Related Work

Identifying and distinguishing normal and abnormal network traffic is one of the important tasks in the field of network security. In order to achieve this goal efficiently, machine learning and deep learning methods have been widely used in the analysis and identification process of network traffic.

### *2.1 Machine Learning-Based Methods*

In abnormal traffic detection, the application of machine learning technology has become a key research direction. For example, Autoencoder (AE), eXtreme Gradient Boosting (XGBoost) and XGBoost combined with Principal Component Analysis (PCA) are advantageous in identifying and processing spatial data features.

In network traffic detection, an AE as a neural network architecture is employed for dimensionality reduction or feature extraction [3–5]. Ieracitano et al. [3] proposed an AE-based Intrusion Detection System (IDS). It combines statistical analysis with AE to extract more optimized and strongly correlated features. Andresini et al. [4] introduced a deep metric learning strategy that learns feature embeddings through triplet networks combined with Hnamte et al. [5] introduced a two-stage deep learning model (LSTM-AE) by combining Long Short-Term Memory networks (LSTM) with AE. This model aims to effectively identify anomalous behaviors in complex network data. In abnormal traffic detection, XGBoost applications address complex data challenges [6–8]. Kasongo et al. [6] used the XGBoost algorithm to reduce the dimensionality of the feature space, which improves the performance of various Machine Learning (ML) models and addresses challenges such as high-dimensional data spaces and dataset imbalance. On the other hand, the model combining particle swarm optimization and XGBoost is proposed by Jiang et al. [7]. This method focuses on improving the parameter settings of XGBoost through the Particle Swarm Optimization (PSO) algorithm, which enables the model to find optimal solutions within a wider parameter space. This approach significantly improves the performance of XGBoost on classification problems. Further research has improved model performance by combining PCA with XGBoost [9–11]. Bhattacharya et al. [9] demonstrated the effectiveness of PCA in spatial feature extraction and dimensionality reduction, especially when it is combined with the Firefly algorithm and XGBoost. This approach lays the foundation for data classification by reducing data dimensions while retaining important information. Pan et al. [10] further expanded the application of PCA on this basis, especially in dealing with class imbalance problems. By integrating with the Adaptive Synthetic Sampling (ADASYN) algorithm, PCA not only reduces the complexity of the data but also enhances the balance of the dataset, which provides a more optimized feature set for XGBoost. With the issues of feature redundancy and the neglect of feature mean, Chen et al. [11] proposed an optimized feature extraction algorithm. This method initially applies Kernel PCA (KPCA) to project the original data into a high-dimensional space, removing redundant and irrelevant features. Subsequently, it utilizes Linear Discriminant Analysis (LDA) to perform secondary feature extraction in the new feature space, taking into account the mean differences between and within classes, thereby improving the effect of feature extraction. Diwan et al. [12] proposed a novel, lightweight feature selection method for IoT intrusion detection, which leverages rank-based chi-square, Pearson correlation, and score correlation to identify key dataset features. Similarly, Jhansi et al. [13] used Ant Lion Optimization, Cuckoo Search Optimization, and Firefly Optimization alongside autoencoders for efficient Application Programming Interface (API) scheduling in malware detection.

Although existing machine learning methods have achieved good results in detecting anomalous traffic, many models highly depend on manually extracted features. This approach is not only time-consuming but may also overlook some important complex features in the data. Furthermore, manual feature extraction can lead to models struggling to adapt to new or unknown attack patterns, limiting their generalization and practicality. Therefore, employing methods that automatically learn and extract features is particularly important to overcome the limitations of current approaches in feature extraction and processing.

### *2.2 Deep Learning-Based Methods*

Compared with machine learning methods, deep learning performs better in terms of learning accuracy and portability because it does not require manual design of features. Abnormal traffic detection typically always relies on spatial and temporal features, as well as a combination of both. Spatial features are usually extracted using Convolutional Neural Network (CNN) [14–16]. Li et al. [14] proposed a multi-CNN fusion method. This method divides feature data into four parts for processing and fusion, aiming to enhance the precision and efficiency of network intrusion detection. In order to further optimize the problem of feature extraction caused by sample data differences, Shi et al. [15] proposed the Deep Abnormal Network Traffic Detection (DANTD) method for effective spatial feature extraction. This model uses deep convolutional autoencoders for high-order feature extraction and employs Generative Adversarial Networks (GAN) for data augmentation. The extraction of temporal features relies on the temporal convolution model [17,18]. Li et al. [17] proposed a method using dynamic chaotic Cross-optimized bidirectional residual-gated recurrent unit and Wasserstein generative adversarial network with generated feature domains. This approach leverages the strengths of GRU for processing time series data, which optimizes weights to achieve more efficient feature extraction and reduced time complexity. Cai et al. [18] developed a method using Bidirectional Temporal Convolutional Network (BiTCN) and Multi-Head Self-Attention (MHSA) mechanism. The method employs BiTCN to capture bidirectional semantic features of network traffic and uses MHSA to assign varying weights to different subsequence segments. Recent studies highlight the significant advantages of hybrid models that focus on extracting both spatial and temporal features [19–23]. Kanna et al. [19] introduced a model combining an Optimized CNN and Hierarchical Multi-Scale LSTM (HMLSTM). This model employs Lion Swarm Optimization (LSO) to enhance CNN spatial feature extraction, while HMLSTM handles temporal feature extraction. Anitha et al. [20] developed a network integrating Bidirectional Long Short-Term Memory (BiLSTM) with a CNN, where the BiLSTM captures long-term dependencies in time series data and the CNN processes and classifies the data. Zhu et al. [21] proposed a model that combines 1D-CNN and BiLSTM. This method effectively extracts time series and spatial features. It also employs a cost penalty matrix and an improved cross-entropy loss function to enhance the recognition of minority class samples. Wang et al. [22] further proposed a model for spatial-temporal feature fusion, using a simplified CNN for spatial learning and BiLSTM for temporal feature learning, incorporating an attention mechanism for effective feature integration. To address overfitting in training, Hassan et al. [23] developed a hybrid deep learning model that combines CNN with Weighted Decreasing LSTM (WDLSTM) to extract key features efficiently. Since the accurate extraction of spatial features directly affects the sensitivity and accuracy of abnormal traffic detection. It is a core link to ensure network security. We mainly focus on the extraction of spatial features in this paper.

Multi-scale feature extraction methods are crucial for improving model accuracy, generalization, capturing data features at various levels [24–27]. Duan et al. [24] introduced a Multi-Scale Residual Classifier (MSRC) for anomaly traffic detection, utilizing wavelet transform to effectively process

multi-scale network traffic features. This enhances the accuracy of detecting network traffic anomalies. Yu et al. [25] developed a high-precision intrusion detection system using a Multi-Scale CNN, which extracts features from disordered data to increase the accuracy of intrusion detection. He et al. [26] proposed a method combining a Variational Gaussian Model with a One-Dimensional Pyramid Depth Separable Convolution (PyDSC) network. This approach simplifies complex features using PyConv with added DSC to reduce network complexity. Addressing high-dimensional and complex datasets, Zhang et al. [27] firstly analyzed spatial features using a Multi-Scale Convolutional Neural Network, and processed temporal features with LSTM. Sathya et al. [28] introduced a classification method, which utilizes a dual weight update mechanism to differentiate between attack and non-attack data in IoT devices. Ravi Kiran Varma et al. [29] proposed a software-defined IoT intrusion attack detection method based on enhanced Elman spiking neural network.

In summary, existing abnormal traffic detection methods have obvious limitations in multi-dimensional data feature extraction, identification and maintenance of high-level features. Specifically, (1) some existing methods usually only focus on feature extraction at a single scale, ignoring other levels of information in the data. This limits the model's ability to understand and process complex data structures. (2) With the complex high-dimensional data, it is difficult to identify subtle differences and high-dimensional features. Therefore, this paper proposes an MD block to effectively capture and integrate features of different scales in multi-dimensional data. And a MRD block is designed to identify complex patterns in high-level features with subtle differences.

## 3 Methodology

### *3.1 Overall Network Model*

Raw data with pcap format generated by industrial IoT need to be converted to images with grayscale format by method in [30]. This paper adopts the ResNeXt network [31] as its main architecture. The preprocessed data are fed into MD block with different scales to cover more global features. Dilated convolutions are used in the MD block to enhance the receptive field. The output of the MD block serves as the input for the MRD block, which combines the Max-Feature-Map (MFM) with residual blocks and dual-channel pooling for the enhancement of precision and efficiency. The MR block of MRD is used to enhance the extraction of key features and maintain feature continuity in deep networks. Finally, the classification results are obtained through the fully connected layer. The overall structure of the MR-MRD-ResNeXt model is shown in Fig. 1a. Meanwhile the detailed MD and MR block are shown in Figs. 1b and 1c, respectively.

### *3.2 Multi-Scale Dilated Feature Extraction*

With the introduction of deep learning, feature extraction has been automated and the ability to process complex data has been improved. However, some feature extraction methods in deep learning are still limited to a single scale, which often focuses only on local details or overall patterns and neglects other important dimensions of the data.

Inspired by the multi-scale feature processing method [32], a four-layer multi-scale feature extraction module is used as shown in Fig. 1b. In this structure, a $1 \times 1$ convolution is used in the first branch to enhance the feature representation capabilities of the model, improve the detail capture capability of the networks, and optimize computational efficiency. In the last three branches, dilated convolution is added to the $3 \times 3$ convolution. The dilation rate of the dilated convolution is adjusted to 1, 2, and 3. A convolution with a dilation rate of 1 is equivalent to a regular convolution and it can capture the potential local features. Gradually increasing the expansion rate to 2 and 3, the

receptive field's distribution is optimized. It captures multi-scale information without changing the size of the feature map, thereby capturing a wider range of spatial features. Consequently, this approach avoids excessive sparseness in feature maps and enhances model sensitivity to small-scale features. The features extracted at different scales are integrated to ensure them work collaboratively in the final decision-making process. Afterwards, by integrating batch normalization and ReLU activation function in the post-convolution stage, the training stability and nonlinear expression ability of the model are improved.

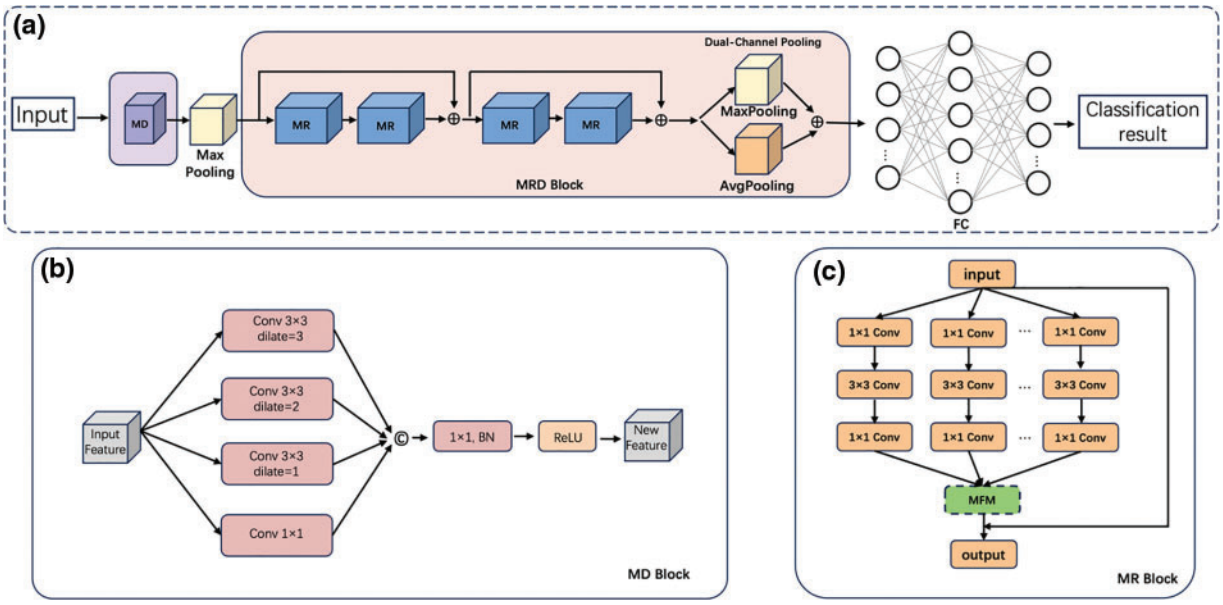


**Figure 1:** (a) Overall architecture of the proposed model (b) Multi-scale dilated feature extraction (c) Detailed information about the MR module

### *3.3 Max-Feature-Map Residual with Dual-Channel Pooling*

When dealing with high-dimensional data with multiple characteristics and complex relationships, traditional methods are often difficult to effectively distinguish and maintain these key features. In addition, as the network depth increases, the model may lose sensitivity to important features, which results in insufficient feature recognition and generalization capabilities. In order to avoid the fuzzy identification of high-dimensional data with multiple characteristics, in this paper we propose an MRD block. The feature extraction is optimized by combining MFM and residual blocks. As illustrated in Fig. 1c. MFM selects the most significant feature responses across channels to enhance key features. Assume $X_{h_i,w_j,c_k}$ represents the value of channel $c_k$ at position $(h_i, w_j)$ in feature map $X \in R^{C\times H\times W}$, then $MFM(X)_{h_i,w_j,c_k}$ is calculated as Eq. (1).

$$MFM(X)_{h_i,w_j,c_k} = max_{c_k\in C}\left(X_{h_i,w_j,c_k}, X_{h_i,w_j,c_{k+\frac{C}{2}}}\right) \tag{1}$$

The residual block helps maintain the continuity and stability of features in deep networks through its skip connections. In each layer of the residual block, through skip connections. In each layer of the residual block, skip connections allow the model to retain the feature information from the previous layer, to reduce information loss, and to improve the accuracy of feature recognition. This combination

strategy not only improves the model's ability to maintain features in complex data, but also effectively reduces overfitting and enhances generalization capabilities.

In addition, because dual-channel pooling can capture features more comprehensively in feature extraction, it combines the advantages of the significant feature response of max pooling and the global information capture of average pooling. Therefore, we adopt it to process key information in complex data and improve the performance of the model in abnormal traffic detection. Assume $D(X)$ and $MaxPool(X)$ represent the feature map obtained after dual-channel pooling and the result after the maximum pooling, then $D(X) = MaxPool(X) \oplus AvgPool(X)$. $AvgPool(X)$ is the result after using the average pooling and $\oplus$ represents the element-wise phase.

Through the combination of MFM, residual block and dual-channel pooling. The MRD block enhances the model's feature capture capabilities and optimizes computational efficiency to achieve more efficient and accurate abnormal traffic detection. The detailed MRD is shown in Algorithm 1.

### *3.4 Model Training Process*

The preprocessed dataset is divided the training set and validation set into a ratio of 8:2. The training set is used for the learning process of the model, while the validation set is used for the evaluation of model performance. In the first stage of the model, the MD block is responsible for capturing features of different scales through dilated convolution technology, which helps the model learn more detailed data representation. Next, in the MRD block, we use the MFM method to capture the most significant feature responses in the deep network structure. At the same time, we enhance the ability of the model to maintain deep features through residual connections. This method helps avoid information loss during training, especially when the network depth is large. Additionally, by combining the advantages of maximum pooling and average pooling, it can retain rich feature information while reduce the number of parameters.

**Algorithm 1:** Max-Feature-Map Residual with Dual-Channel Pooling (MRD)

Input: Feature map $X$
Output: Optimized feature map $D(X)$
1: Initialize feature map X, including channel number C, width W and height H;
2: For $h_i \in H$ do:
3: For $w_j \in W$ do:
4: For $c_k \in C$ do:
5: Calculate $MFM(X)_{h_i,w_j,c_k}$ by Eq.(1);
6: EndFor
7: EndFor
8: $X = MFM(X)_{h_i,w_j,c_k} + X$
9: EndFor
10:$D(X) = MaxPool(X) \oplus AvgPool(X)$
11: Return $D(X)$

During the training process, we choose Adam [32] as the optimizer to adjust and optimize the model weights with its effective adaptive learning rate. Training is performed in small batches (the batchsize is 32), which helps to increase the updating frequency of model and improve the accuracy of gradient estimation. In addition, we use the ReduceLROnPlateau learning rate [33] scheduler to dynamically adjust the learning rate. This strategy determines whether to reduce the learning rate based on the model's performance in terms of loss on the validation set, which ensures that the model

does not get stuck in a local minimum during the training process. Model training is scheduled for 50 epochs to ensure sufficient iterations to train the network. The learning rate starts at 0.001, and after the 46th epoch, the learning rate is reduced to $1 * 10^{-6}$. After each epoch, the validation set is employed to assess the model's performance by accuracy and loss metrics. And the classification result is calculated with the average of multiple experiments. Finally, after a series of training processes, our model showed excellent performance on the abnormal traffic detection, verifying the effectiveness of the MD and MRD blocks.

## 4 Experiments and Analyses

In this paper, the experimental environment are running on a server with RTX A5000 GPU and 24 GB RAM using Python3.8 + TensorFlow 2.10.0.

### *4.1 Datasets*

To verify the effectiveness of the MD-MRD-ResNeXt and its variants of the network proposed in this paper for abnormal traffic detection, this section conducts detailed performance comparison experiments. The experimental datasets used are USTC-TFC2016 [30] and ToN-IoT-Network [34]. The USTC-TFC2016 dataset is composed of two segments. The first segment includes a collection of ten varieties of malicious traffic, gathered by CTU (Czech Technical University) researchers in real-world network settings from 2011 to 2015. The second segment comprises a set of ten kinds of benign traffic, obtained through IXIA BPS (Ixia Breaking Point Systems). A total of 202,921 records from the USTC-TFC2016 dataset are used. The ToN-IoT-Network dataset was developed by the IoT lab at UNSW (The University of New South Wales) Canberra in collaboration with Cyber Range. It encompasses telemetry data from connected devices, logs from both Linux and Windows operating systems, as well as network traffic from IIoT systems. This heterogeneous data was collected from a medium-sized IoT network. A total of 260,462 records from the ToN-IoT-Network dataset are considered. The data used in USTC-TFC2016 and ToN-IoT-Network is detailed in Table 1. All these data are converted to images with grayscale format by method in [30].

**Table 1:** Statistics of samples with different categories in two datasets

| USTC-TFC2016 | | | | | | ToN-IoT-Network | | |
|---|---|---|---|---|---|---|---|---|
| Category | Train | Test | Category | Train | Test | Category | Train | Test |
| BitTorrent | 6752 | 750 | Cridex | 14752 | 1639 | Normal | 54000 | 6000 |
| Facetime | 5400 | 600 | Geod | 11743 | 1305 | Password | 22513 | 2502 |
| FTP | 11184 | 1243 | Htbot | 9566 | 1063 | Dos | 25552 | 2839 |
| Gmail | 4945 | 550 | Miuref | 8017 | 891 | DDos | 17243 | 1916 |
| MySQL | 12571 | 1397 | Neris | 13466 | 1496 | Injection | 38782 | 4309 |
| Outlook | 6267 | 748 | Nsis-ay | 9805 | 1089 | MITM | 5909 | 657 |
| Skype | 5480 | 609 | Shifu | 12920 | 1436 | XSS | 54000 | 6000 |
| SMB | 5031 | 559 | Tinba | 13910 | 1546 | Scanning | 9560 | 1062 |
| Weibo | 4112 | 457 | Virut | 10110 | 1123 | Backdoor | 6692 | 744 |
| WorldOfWarcraft | 6841 | 760 | Zeus | 9709 | 1079 | Ransomware | 164 | 18 |

### *4.2 Evaluation Metrics*

In evaluating the efficacy of the MD-MRD-ResNeXt model, this paper employs four evaluation metrics: Accuracy (AC), Precision (PR), Recall (RC) and F1 score, which are computed by Eqs. (2)–(5). Among all these evaluation criteria, True Positive (TP) refers to the cases where abnormal network traffic is accurately identified by the MD-MRD-ResNeXt. False Positive (FP) refers to the instances where the MD-MRD-ResNeXt incorrectly labels normal network traffic as anomalous. False Negative (FN) refers to the cases where abnormal network traffic that the MD-MRD-ResNeXt fails to identify. True Negative (TN) refers to the cases where normal network traffic is accurately identified by the MD-MRD-ResNeXt.

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

The ratio of the proportion of correct predictions by the model to the total number of predictions.

$$PR = \frac{TP}{TP + FP} \tag{3}$$

The proportion of positive examples predicted by the model that are actually positive.

$$RC = \frac{TP}{TP + FN} \tag{4}$$

Among all actual positive examples, the proportion of positive examples correctly predicted by the model.

$$F1 = \frac{2 * PR * RC}{PR + RC} \tag{5}$$

The harmonic means of precision and recall.

### *4.3 Experimental Analyses*

#### *4.3.1 Ablation Experiments*

Differences in network architectures can lead to variations in performance. Table 2 presents the comparison results between MD-MRD-ResNeXt and MD-MRD-ResNet, both of which are based on the same strategy but utilize different backbone networks. As evident from Table 2, MD-MRD-ResNeXt outperforms MD-MRD-ResNet across most evaluation metrics. This can primarily be attributed to the structural differences between ResNeXt and ResNet. ResNeXt enhances the expressive capacity of the model by incorporating grouped convolutions and cardinality connections. This is achieved without significantly increasing the number of parameters. This gives it a superior ability to capture intricate features. In contrast, ResNet strengthens the training capabilities of the network through residual connections. While it might not capture the same features as ResNeXt in certain scenarios. However, within the MD-MRD architecture, the specific structure of ResNeXt provides additional advantages for certain tasks. The selections of ResNeXt as the backbone network in a specific MD-MRD architecture can provide effective performance. This further highlights the importance of considering subtle differences and potential impacts when choosing network structures for specific tasks. In the subsequent section, various ablation study strategies for the proposed MD-MRD-ResNeXt are designed.

**Table 2:** Verification results of different network structures on USTC-TFC2016 dataset

| Model | AC | PR | RC | F1 |
|---|---|---|---|---|
| **MD-MRD-ResNeXt** | **97.16%** | **97.24%** | **96.78%** | **96.94%** |
| MD-MRD-ResNet | 96.51% | 96.81% | 96.32% | 96.56% |

To validate the effectiveness of each module, the following ablation experiments were conducted, which are shown in Table 3. From Table 3, it can be observed that MD-MRD-ResNeXt demonstrates superior performance in Accuracy, Recall, and F1 score. The primary reason is the MD block enlargement of the receptive field of the convolutional kernel, enabling the model to capture richer information from the input and recognize various low-level abnormal traffic features. Meanwhile, the MRD block allows the network to focus more intently on crucial information, discarding irrelevant features, thereby enhancing feature quality. In contrast, when only employing the MD or MRD, there is a performance improvement, but it does not reach the optimum situation. This indicates that the MD and MRD are contributed to feature extraction and attention for essential information, especially in the MC-MRD-ResNeXt (MC represents multi-scale feature extraction without dilated convolution) experiment. The efficiency of the MD block was further validated too.

**Table 3:** Ablation verification results on USTC-TFC2016 dataset

| Model | Block | | | | Results | | | |
|---|---|---|---|---|---|---|---|---|
| | MC | MD | MRD | ResNeXt | AC | PR | RC | F1 |
| ResNeXt | × | × | × | √ | 95.24% | 95.07% | 94.76% | 95.09% |
| MC-ResNeXt | √ | × | × | √ | 95.89% | 96.07% | 95.90% | 96.21% |
| MD-ResNeXt | × | √ | × | √ | 96.52% | 96.74% | 96.40% | 96.72% |
| MRD-ResNeXt | × | × | √ | √ | 95.94% | 96.08% | 95.89% | 95.98% |
| MC-MRD-ResNeXt | √ | × | √ | √ | 96.28% | 96.59% | 96.05% | 96.32% |
| **MD-MRD-ResNeXt** | × | √ | √ | √ | **97.16%** | **97.24%** | **96.78%** | **96.94%** |

### *4.3.2 Comparison Experiment*

In this study, eight existing abnormal traffic detection methods: 2D-CNN [30], BiDLSTM [35], CNN-BiLSTM [36], PCNN [37], RESNETCNN [38], ResNet-GRU [39], MTC-BYOL [40] and DC-AAE [41] are compared with our MD-MRD-ResNeXt. These methods were chosen because they are not only theoretically well-studied but also empirically validated across various datasets and scenarios, providing a solid benchmark for comparison. These detection methods were evaluated alongside our model based on AC, PR, RC, and F1 score. To provide a comprehensive assessment of the performance of each method, experiments were conducted on the USTC-TFC2016 and ToN-IoT-Network datasets.

Table 4 compares various methods' performance on the USTC-TFC2016 dataset. Experimental results show that BiDLSTM has the worst effect. Although BiDLSTM is effective for time-series data, it struggles to capture multi-dimensional spatial features. 2D-CNN falls short in capturing sufficient

contextual information from complex IoT traffic data, which also results in poorer performance. The comparisons of CNN-BiLSTM, PCNN, and DC-AAE with an encoder-decoder can capture more stable and smooth features to enhance extraction performance. Models RESNETCNN and ResNet-GRU can obtain better performance by a residual network, while they always need more computational resources and meticulous parameter tuning for less optimal performance. The model MD-MRD-ResNeXt in this article performed the best on all evaluation indicators, which is attributed to the unique structure of the model. Because the MD block and MRD block are designed to deeply mine the spatial characteristics of network traffic data. The MD block enables the model to capture richer contextual information and fine-grained anomaly indicators. The use of maximum feature mapping and residual blocks further enhances the weight of these features in model decision-making. At the same time, our model adopts an innovative training strategy and dynamically adjusts the learning rate through adaptive learning rate to accelerate the convergence speed of the model. In addition, we introduce early stopping to prevent overfitting and ensure that the model can achieve the best generalization ability on different data sets during training. The results on the ToN-IoT-Network dataset are similar to those of Table 4, which further prove the effectiveness of our model as shown in Table 5.

**Table 4:** Verification results of all comparative abnormal traffic detection in USTC-TFC2016

| Model | AC | PR | RC | F1 |
|---|---|---|---|---|
| 2D-CNN [30] | 92.93% | 93.26% | 90.18% | 91.69% |
| BiDLSTM [35] | 90.11% | 91.69% | 88.64% | 90.16% |
| CNN-BiLSTM [36] | 94.67% | 95.39% | 95.07% | 95.20% |
| PCNN [37] | 94.36% | 94.82% | 93.71% | 94.26% |
| RESNESTCNN [38] | 95.13% | 95.47% | 94.95% | 95.21% |
| ResNet-GRU [39] | 95.30% | 95.58% | 95.22% | 95.22% |
| MTC-BYOL [40] | 96.24% | 96.67% | 96.03% | 96.34% |
| DC-AAE [41] | 95.30% | 95.64% | 94.67% | 95.15% |
| **MD-MRD-ResNeXt** | **97.16%** | **97.24%** | **96.78%** | **97.17%** |

**Table 5:** Verification results of all comparative abnormal traffic detection in ToN-IoT-Network

| Model | AC | PR | RC | F1 |
|---|---|---|---|---|
| 2D-CNN [30] | 88.83% | 89.87% | 88.25% | 89.05% |
| BiDLSTM [35] | 86.23% | 91.71% | 81.63% | 86.37% |
| CNN-BiLSTM [36] | 88.43% | 89.95% | 87.33% | 88.61% |
| PCNN [37] | 91.64% | 92.59% | 91.12% | 91.85% |
| RESNESTCNN [38] | 92.77% | 94.07% | 91.84% | 92.95% |
| ResNet-GRU [39] | 90.29% | 92.92% | 88.61% | 90.68% |
| MTC-BYOL [40] | 91.87% | 92.25% | 90.28% | 91.25% |

(Continued)

**Table 5 (continued)**

| Model | AC | PR | RC | F1 |
|---|---|---|---|---|
| DC-AAE [41] | 90.26% | 92.45% | 88.97% | 90.67% |
| **MD-MRD-ResNeXt** | **93.12%** | **94.41%** | **92.36%** | **93.38%** |

Figs. 2 and 3 demonstrate the distribution of results for all the compared methods of abnormal traffic detection on the USTC-TFC2016 and ToN-IoT-Network datasets, respectively. On the USTC-TFC2016 dataset, our model exhibits superior performance compared to the results on the ToN-IoT-Network dataset. The difference can be partly attributed to the more consistent and rule-conforming feature distribution of the USTC-TFC2016 dataset. This dataset exhibits some anomalous traffic characteristics, which enables the model to capture key information more effectively. Furthermore, this dataset might offer more balanced sample diversity and category distribution, reducing the risk of overfitting during training. Conversely, the ToN-IoT-Network dataset, with its more varied and complex IoT device traffic patterns, demands a higher level of generalization from the model. This is because each type of device might generate distinct traffic features. And the diversity and complexity of attack traffic in the dataset pose greater challenges. In summary,our proposed model has demonstrated a notable performance across different datasets compared to other popular methods, reaffirming its effectiveness and robustness.

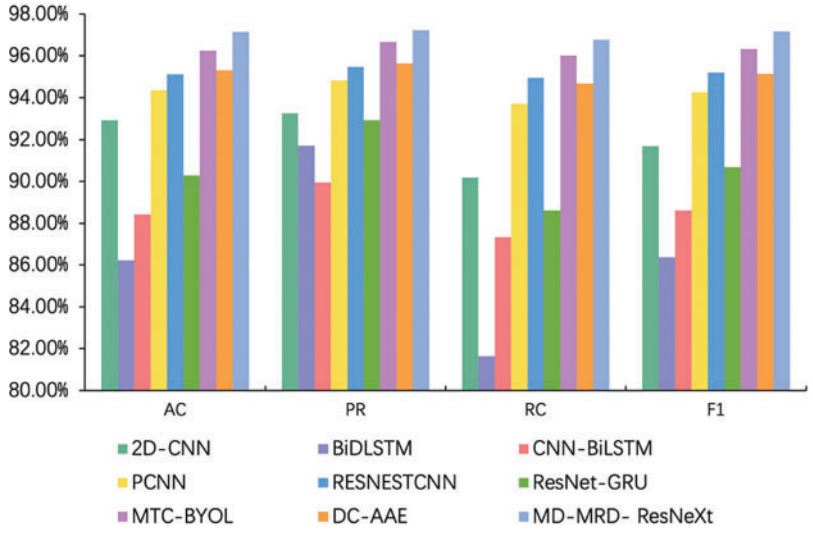


**Figure 2:** Verification results of all comparative abnormal traffic detections in USTC-TFC2016

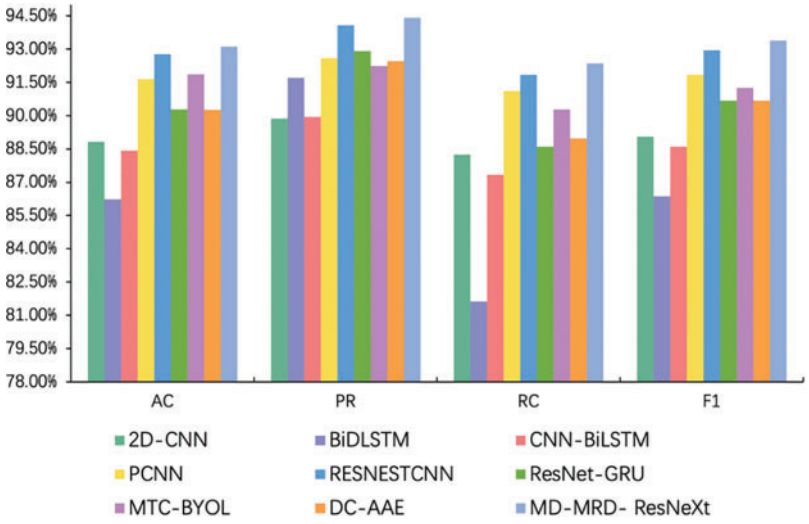


**Figure 3:** Verification results of all comparative abnormal traffic detections in ToN-IoT-Network

## 5 Conclusion

To address the issues of multidimensionality and complexity in feature extraction, this paper proposes the MD-MRD-ResNeXt model for abnormal network traffic detection. By introducing a multi-scale dilated feature extraction module, the model expands its receptive field to capture and integrate features of various scales in parallel. MD block can fully consider all relevant data information in the decision-making process. The proposed MRD block further ensures that the model can accurately extract key features for the diverse features and complex relationships in high-dimensional data. The MD-MRD-ResNeXt model not only optimizes the model's computational efficiency but also significantly improves its generalization ability in abnormal network traffic detection tasks. Experimental results show that the MD-MRD-ResNeXt model performs well in terms of AC, PR, RC, and F1, highlighting its efficiency and practicality in identifying abnormal network traffic. With the development of emerging attack patterns (including unknown zero-day attacks), our model needs to be further improved to address these new attacks in broader network environments. Meanwhile, the imbalanced data always have a significant influence on abnormal detection, which is also our future work.

**Acknowledgement:** The authors thank all research members who provided support and assistance in this study.

**Funding Statement:** This work is supported by the Key Research and Development Program of Xinjiang Uygur Autonomous Region (No. 2022B01008), the National Natural Science Foundation of China (No. 62363032), the Natural Science Foundation of Xinjiang Uygur Autonomous Region (No. 2023D01C20), the Scientific Research Foundation of Higher Education (No. XJEDU2022P011), National Science and Technology Major Project (No. 2022ZD0115803), Tianshan Innovation Team Program of Xinjiang Uygur Autonomous Region (No. 2023D14012) and the "Heaven Lake Doctor" Project (No. 202104120018).

**Author Contributions:** Research conception and design: Tingting Su; Data collection: Gaoqiang Dong; Result analysis and interpretation: Tingting Su, Jia Wang; Manuscript preparation: Tingting Su, Jia Wang, Wei Hu and Jeon Gwanggil.

**Availability of Data and Materials:** The datasets used in this article are public data sets: The first dataset is the USTC-TFC2016 dataset, and the access method is as follows: https://github.com/yungshenglu/USTC-TFC2016. The dataset was further processed using the tools in https://github.com/yungshenglu/USTC-TK2016 to adapt to the needs of this study. The second dataset is the TON-IoT-Network dataset, which can be accessed as follows: https://research.unsw.edu.au/projects/toniot-datasets.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References


[1] K. L. Dias, M. A. Pongelupe, W. M. Caminhas, and L. Errico, "An innovative approach for real-time network traffic classification," *Comput. Netw.*, vol. 158, pp. 143–157, Jul. 2019. doi: 10.1016/j.comnet.2019.04.004.

[2] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Comput.*, vol. 50, no. 7, pp. 80–84, Jul. 2017. doi: 10.1109/MC.2017.201.

[3] C. Ieracitano, A. Adeel, F. C. Morabito, and A. Hussain, "A novel statistical analysis and autoencoder driven intelligent intrusion detection approach," *Neurocomputing*, vol. 387, pp. 51–62, Apr. 2020. doi: 10.1016/j.neucom.2019.11.016.

[4] G. Andresini, A. Appice, and D. Malerba, "Autoencoder-based deep metric learning for network intrusion detection," *Inf. Sci.*, vol. 569, pp. 706–727, Aug. 2021. doi: 10.1016/j.ins.2021.05.016.

[5] V. Hnamte, H. Nhung-Nguyen, J. Hussain, and Y. H. Kim, "A novel two-stage deep learning model for network intrusion detection: LSTM-AE," *IEEE Access*, vol. 11, pp. 37131–37148, Apr. 2023. doi: 10.1109/ACCESS.2023.3266979.

[6] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset," *J. Big Data*, vol. 7, pp. 1–20, Nov. 2020. doi: 10.1186/s40537-020-00379-6.

[7] H. Jiang, Z. He, G. Ye, and H. Zhang, "Network intrusion detection based on PSO-Xgboost model," *IEEE Access*, vol. 8, pp. 58392–58401, Mar. 2020. doi: 10.1109/ACCESS.2020.2982418.

[8] N. Saini, V. Bhat Kasaragod, K. Prakasha, and A. K. Das, "A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection," *Concurr. Comput.: Pract. Exp.*, vol. 35, no. 28, pp. e7865, Jul. 2023. doi: 10.1002/cpe.7865.

[9] S. Bhattacharya *et al.*, "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU," *Electronics*, vol. 9, no. 2, pp. 219, Jan. 2020. doi: 10.3390/electronics9020219.

[10] L. Pan and X. Xie, "Network intrusion detection model based on PCA + ADASYN and XGBoost," in *Proc. 3rd Int. Conf. on EBIMCS. Association for Computing Machinery*, New York, USA, Dec. 2020, pp. 44–48. doi: 10.1145/3453187.3453311.

[11] J. Chen, Y. Chen, S. Cai, S. Yin, L. Zhao and Z. Zhang, "An optimized feature extraction algorithm for abnormal network traffic detection," *Future Gener. Comput. Syst.*, vol. 149, pp. 330–342, Dec. 2023. doi: 10.1016/j.future.2023.07.039.

[12] T. D. Diwan *et al.*, "Feature entropy estimation (FEE) for malicious IoT traffic and detection using machine learning," *Mob. Inf. Syst.*, vol. 2021, pp. 1–13, Dec. 2021. doi: 10.1155/2021/8091363.

[13] K. S. Jhansi, P. Varma, and S. Chakravarty, "Swarm optimization and machine learning for android malware detection," *Comput. Mater. Contin.*, vol. 73, no. 3, pp. 6327–6345, 2022. doi: 10.32604/cmc.2022.030878.

[14] Y. Li *et al.*, "Robust detection for network intrusion of industrial IoT based on multi-CNN fusion," *Measurement*, vol. 154, pp. 107450, Mar. 2020. doi: 10.1016/j.measurement.2019.107450.
[15] G. Shi, X. Shen, F. Xiao, and Y. He, "DANTD: A deep abnormal network traffic detection model for security of industrial internet of things using high-order features," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21143–21153, Mar. 2023. doi: 10.1109/JIOT.2023.3253777.
[16] B. Xia, D. Han, X. Yin, and G. Na, "RICNN: A ResNet & inception convolutional neural network for intrusion detection of abnormal traffic," *Comput. Sci. Inf. Syst.*, vol. 19, no. 1, pp. 309–326, 2022. doi: 10.2298/CSIS210617055X.
[17] K. Li, W. Ma, H. Duan, H. Xie, J. Zhu and R. Liu, "Unbalanced network attack traffic detection based on feature extraction and GFDA-WGAN," *Comput. Netw.*, vol. 216, pp. 109283, Oct. 2022. doi: 10.1016/j.comnet.2022.109283.
[18] S. Cai, H. Xu, M. Liu, Z. Chen, and G. Zhang, "A malicious network traffic detection model based on bidirectional temporal convolutional network with multi-head self-attention mechanism," *Comput. Secur.*, vol. 136, pp. 103580, Jan. 2024. doi: 10.1016/j.cose.2023.103580.
[19] P. R. Kanna and P. Santhi, "Unified deep learning approach for efficient intrusion detection system using integrated spatial-temporal features," *Knowl.-Based Syst.*, vol. 226, pp. 107132, Aug. 2021. doi: 10.1016/j.knosys.2021.107132.
[20] T. Anitha, S. Aanjankumar, S. Poonkuntran, and A. Nayyar, "A novel methodology for malicious traffic detection in smart devices using BI-LSTM-CNN-dependent deep learning methodology," *Neural Comput. Appl.*, vol. 35, no. 27, pp. 20319–20338, Jul. 2023. doi: 10.1007/s00521-023-08818-0.
[21] S. Zhu, X. Xu, H. Gao, and F. Xiao, "CMTSNN: A deep learning model for multiclassification of abnormal and encrypted traffic of internet of things," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11773–11791, Feb. 2023. doi: 10.1109/JIOT.2023.3244544.
[22] H. Wang, X. Di, Y. Wang, B. Ren, G. Gao and J. Deng, "An intelligent digital twin method based on spatio-temporal feature fusion for IoT attack behavior identification," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 11, pp. 3561–3572, Aug. 2023. doi: 10.1109/JSAC.2023.3310091.
[23] M. M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, and G. Fortino, "A hybrid deep learning model for efficient intrusion detection in big data environment," *Inf. Sci.*, vol. 513, pp. 386–396, Mar. 2020. doi: 10.1016/j.ins.2019.10.069.
[24] X. Duan, Y. Fu, and K. Wang, "Network traffic anomaly detection method based on multi-scale residual classifier," *Comput. Commun.*, vol. 198, pp. 206–216, Jan. 2023. doi: 10.1016/j.comcom.2022.10.024.
[25] J. Yu, X. Ye, and H. Li, "A high precision intrusion detection system for network security communication based on multi-scale convolutional neural network," *Future Gener. Comput. Syst.*, vol. 129, pp. 399–406, Apr. 2022. doi: 10.1016/j.future.2021.10.018.
[26] J. He, X. Wang, Y. Song, and Q. Xiang, "A multiscale intrusion detection system based on pyramid depthwise separable convolution neural network," *Neurocomputing*, vol. 530, pp. 48–59, Apr. 2023. doi: 10.1016/j.neucom.2023.01.072.
[27] J. Zhang, Y. Ling, X. Fu, X. Yang, G. Xiong and R. Zhang, "Model of the intrusion detection system based on the integration of spatial-temporal features," *Comput. Secur.*, vol. 89, pp. 101681, Feb. 2020. doi: 10.1016/j.cose.2019.101681.
[28] M. Sathya *et al.*, "A novel, efficient, and secure anomaly detection technique using DWU-ODBN for IoT-enabled multimedia communication systems," *Wirel. Commun. Mob. Comput.*, vol. 2021, pp. 1–12, Dec. 2021. doi: 10.1155/2021/4989410.
[29] P. Ravi Kiran Varma, R. R. Sathiya, and M. Vanitha, "Enhanced Elman spike neural network based intrusion attack detection in software defined internet of things network," *Concurr. Comput.*, vol. 35, no. 2, pp. e7503, 2023. doi: 10.1002/cpe.7503.
[30] W. Wang, M. Zhu, X. W. Zeng, X. Z. Ye, and Y. Q. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in *Proc. Int. Conf. on Information Networking*, Da Nang, Vietnam, Apr. 2017, pp. 712–717. doi: 10.1109/ICOIN.2017.7899588.

[31] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. on CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 5987–5995. doi: 10.1109/CVPR.2017.634.

[32] H. Xia, J. Ma, J. Ou, X. Lv, and C. Bai, "Pedestrian detection algorithm based on multi-scale feature extraction and attention feature fusion," *Digit. Signal Process.*, vol. 121, pp. 103311, Mar. 2022. doi: 10.1016/j.dsp.2021.103311.

[33] Q. Chen *et al.*, "Neighborhood rough residual network-based outlier detection method in IoT-Enabled maritime transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 11, pp. 11800–11811, Nov. 2023. doi: 10.1109/TITS.2023.3285615.

[34] A. R. Gad, A. A. Nashat, and T. M. Barkat, "Intrusion detection system using machine learning for vehicular Ad Hoc networks based on ToN-IoT dataset," *IEEE Access*, vol. 9, pp. 142206–142217, Oct. 2021. doi: 10.1109/ACCESS.2021.3120626.

[35] Y. Imrana, Y. Xiang, L. Ali, and Z. Abdul-Rauf, "A bidirectional LSTM deep learning approach for intrusion detection," *Expert. Syst. Appl.*, vol. 185, pp. 115524, Dec. 2021. doi: 10.1016/j.eswa.2021.115524.

[36] J. Sinha and M. Manollas, "Efficient deep CNN-BiLSTM model for network intrusion detection," in *Proc. 3rd Int. Conf. on AIPR, Association for Computing Machinery*, New York, USA, Jun. 2020, pp. 223–231. doi: 10.1145/3430199.3430224.

[37] Y. Zhang, X. Chen, D. Guo, M. Song, Y. Teng and X. Wang, "PCCN: Parallel cross convolutional neural network for abnormal network traffic flows detection in multi-class imbalanced network traffic flows," *IEEE Access*, vol. 7, pp. 119904–119916, Aug. 2019. doi: 10.1109/ACCESS.2019.2933165.

[38] Y. Li, D. Han, M. Cui, F. Yuan, and Y. Zhou, "RESNETCNN: An abnormal network traffic flows detection model," *Comput. Sci. Inf. Syst.*, vol. 20, no. 3, pp. 997–1014, 2023. doi: 10.2298/CSIS221124004L.

[39] G. Zhao, C. Ren, J. Wang, Y. Huang, and H. Chen, "IoT intrusion detection model based on gated recurrent unit and residual network," *Peer Peer Netw. Appl.*, vol. 16, pp. 1887–1899, Jun. 2023. doi: 10.1007/s12083-023-01510-z.

[40] M. S. Towhid and N. Shahriar, "Encrypted network traffic classification using self-supervised learning," in *Proc. IEEE 8th Int. Conf. on NetSoft*, Milan, Italy, Aug. 2022, pp. 366–374. doi: 10.1109/NetSoft54395.2022.9844044.

[41] L. Zhang, J. Yin, J. Ning, Y. Wang, B. Adebisi and J. Yang, "A novel unsupervised malware detection method based on adversarial auto-encoder and deep clustering," in *Proc. 9th Int. Conf. on DSA*, Urumqi, China, Oct. 2022, pp. 224–229. doi: 10.1109/DSA56465.2022.00038.