Hindawi
Security and Communication Networks
Volume 2021, Article ID 9927325, 19 pages
https://doi.org/10.1155/2021/9927325

Research Article

Classification of Abnormal Traffic in Smart Grids Based on GACNN and Data Statistical Analysis

F. F. Hu,[1] S. T. Zhang,[1] X. B. Lin,[1] L. Wu,[1] and N. D. Liao[2]

[1] *CSG Power, Dispatching Control Center, Guangzhou 510663, China*
[2] *Hunan Provincial Engineering Research Center of Electric Transportation and Smart Distribution Network (Changsha University of Science and Technology), Changsha 410114, China*

Correspondence should be addressed to N. D. Liao; lndy97@csust.edu.cn

Received 11 March 2021; Accepted 7 June 2021; Published 17 June 2021

Academic Editor: Ricardo Chaves

With the continuous development of smart grids, communication networks carry more and more power services, and at the same time, they are also facing more and more security issues. For example, some malicious software usually uses encryption technology or tunnel technology to bypass firewalls, intrusion detection systems, etc., thereby posing a serious threat to the information security of smart grids. At present, the classification of network traffic mainly depends on the correct extraction of network protocol characteristics. However, the process of extracting network features by some traditional methods is time-consuming and overly dependent on experience. In order to solve the problem of accurate classification of power network traffic, this paper proposes a method of convolutional neural network based on genetic algorithm optimization (GACNN) and data statistical analysis. This method can simultaneously extract the time characteristics between different packet groups and the spatial characteristics in the same packet group. Therefore, it greatly saves manpower and gets rid of the dependence on experience value. The proposed method has been tested and verified on the UNSW-NB15 dataset and the real dataset collected by the power company. The results show that the proposed method can correctly classify abnormal network flows and is much better than traditional machine learning methods. In large-scale real network flow scenarios, the detection rate of the proposed method exceeds 97%, while the traditional method is generally less than 90%.

1. Introduction

1.1. Background. As a next-generation grid, smart grids have the characteristics of high controllability, high energy efficiency, and self-healing. Smart grids have been rapidly built in many countries and regions in the world, providing great convenience to people's lives [1]. Compared with traditional power grid, smart grids require more monitoring and control devices and are more widely distributed. In order to achieve comprehensive and real-time monitoring, low-cost wireless communication network and widely distributed public Internet are increasingly used in power grid. However, the massive access of the public network in the power system provides more access for malicious attacks. This will bring more harm to the power grid and users [2]. With some new energy access to smart grids, the security risk of smart grids increases greatly, such as violent attacks, denial of service attacks, and computer viruses [3].

Due to different business requirements in different areas of smart grids, the requirements for underlying network communication are also different. Table 1 shows the main network technology architecture of the current smart grids [4]. At present, the network communication of smart grids mainly depends on advanced metering infrastructure (AMI) [5]. The AMI network is mainly composed of home local area network (HAN), neighborhood network (NAN), and wide area network (WAN) [6].

It can be seen from Table 1 that different power users have different demands on the network. Timely monitoring and analysis of network traffic is of great significance to improving network performance and security defense. When the network traffic monitoring system is applied to the

Table 1: Network architecture of smart grids.

Network type	Network technique	Application scenarios
Wide area network (WAN)	IP, DWDM	Provide power data network (backbone network), Internet interconnection, and routing functions
	MPLS\MPLS VPN	Label switching is provided in backbone network to isolate traffic of different services
	ATM	Asynchronous Transfer Mode
	SDH	Provides physical access to MAN and WAN
Access network (AN)	MSTP	It can be used for LAN (Ethernet) access to MAN
	GPRS\3G\4G\5G	Access WAN through mobile communication network
	PON	It is a typical passive optical fiber network
Local area network (LAN)	IEEE 802.3\802.1q	LAN of electric power enterprise
	RS-485, PROFIBUS and other traditional field buses	Production control fields such as power plants and substations
	Industrial Ethernet	Interconnection of IED equipment in production control fields such as power plants and substations
Field area network (FAN)	N-PLC, B-PLC/BPL (narrowband, broadband power line carrier communication)	Used for data transmission such as metering and instrument data collection
	Wireless sensor networks	Data acquisition, monitoring and monitoring of power transmission, distribution and consumption side
	Internet of things, RFID	Collection of label data in equipment inspection
	PON/EPON/FTTH	Intelligent residential area provides optical fiber access for home users
Home area network (HAN)	N-PLC\B-PLC\BPL	Provide local network and home network access, remote meter reading and Internet access
	WLAN 802.11	Remote meter reading with local network or home network access
	Wireless sensor network	Used for control of smart home and home appliances in HAN

power enterprise, it can not only monitor the network traffic in real time but also analyze and warn the abnormal situation of the network [7].

Some early network traffic monitoring models obtain network characteristic information such as bandwidth and link utilization and judge whether network abnormalities occur according to thresholds. Although these models have certain traffic alarm functions, the performance of the models mainly depends on the accuracy of the characteristic values, and the model thresholds are generally set manually [8].

With the rapid development of next-generation networks, integrated services such as voice, video, and data are running in parallel, and the proportion of new traffic types P2P, streaming media, and games continues to increase. Today's network traffic recognition technology has some new problems:

(1) The detection efficiency of the existing methods is low, especially in the high-speed online traffic classification, and the detection method has a large storage overhead and a large amount of calculation.

(2) Some new network services mostly use variable ports, data encryption, and dedicated protocols. Existing methods can no longer accurately extract the characteristics of network traffics, which affects the recognition accuracy.

At present, as a large number of different AMI devices are connected to the network, it provides detailed basic data in many aspects for the power system, but at the same time, it also brings more network security risks. How to grasp the characteristics of different network behaviors in time and better understand the state of network traffic has become one of the key core issues in power communication research.

1.2. Our Contributions. This paper first establishes a standard power network flow metadata to eliminate the problem of multisource and heterogeneous equipment information service coordination. (1) Through the real-time or offline collection of the entire multisource network equipment flow, the unified analysis of different equipment, different network levels, and mass flow data is realized. (2) Through statistical analysis of standardized flow metadata, network congestion or abnormal conditions can be found as early as possible, which provides data support for power network security situation assessment.

Secondly, unlike traditional machine learning methods, this paper uses deep learning to analyze network traffic metadata and quickly find out the temporal and spatial characteristics of traffic. The main advantages of this method are as follows: (1) it is not necessary to judge the importance of features in advance but directly input the metadata to the model for training after preprocessing. (2) The model has self-learning ability, which eliminates the problem that the accuracy of traditional methods is excessively dependent on expert experience or threshold setting. (3) The model has the ability of dealing with encryption and multilevel business traffic, which overcomes the problem that traditional methods cannot analyze network traffic content.

1.3. Organization. The rest of this paper is organized as follows: Section 2 summarizes the methods of network traffic classification. Section 3 introduces the convolutional neural network (CNN), genetic algorithm (GA), and network traffic collection and preprocessing. Section 4 introduces the GACNN method proposed in this paper in detail. Section 5 details the process and results of two experiments conducted on the UNSW-NB15 and real power flow datasets. Finally, Section 6 summarizes the main work of the paper and the future research direction.

2. Related Work

With the construction of smart grids, power data networks and the business systems carried by them have developed rapidly, and a large amount of network traffic is generated in these systems every day.

Power critical information infrastructure such as AMI is the nerve center of economic and social operations, and it is also an important target that may be attacked through the network. A large number of experiments have found that the power network traffic under attack will be abnormal. These kinds of abnormal traffic are usually caused by malicious network attacks (such as worm propagation, DDOS attacks, botnets, and viruses) or network configuration errors and occasional line interruptions [9].

As one of the key technologies of network management and network security, network traffic classification can not only optimize network configuration and reduce network security risks but also provide better service quality. In the past two decades, domestic and foreign researchers have carried out a series of related studies and achieved some outstanding results.

Initially, the network traffic classification technology is relatively simple, because different network applications use different port numbers for communication, so the port number can be used as a function to identify network traffic [10]. Wang et al. [11] proposed a traffic detection algorithm based on port scanning behavior. The algorithm determines whether there is an abnormality in network traffic based on the ratio and similarity of the number of hosts and ports. However, some network protocols currently do not use fixed ports for communication, so this method cannot cope with the problem of sudden changes in ports.

In order to accurately identify different network services on dynamic ports, some researchers have proposed network traffic identification methods based on payload characteristics, such as deep packet inspection (DPI) [12, 13] and deep flow inspection (DFI) [14, 15].

The DPI technology mainly determines the type of each network service by judging whether the network characteristics match the fingerprint library characteristics. Sun et al. [12] first used DPI technology to quickly and accurately preclassify the traffic and then calculated the random characteristics of the packet load, so as to achieve effective identification of encrypted traffic.

DFI mainly uses application identification technology based on traffic behavior. This method believes that different applications will present different states on network session connections or data streams, but DFI does not pay attention to application load. In order to achieve the purpose of lightweight protection of the power Internet of Things, Wang and Wei [14] used DFI technology to analyze the collected network traffic of the Internet of Things and formed a feasible security prevention and control strategy model.

The DPI and DFI methods to process network traffic mainly rely on the application layer message characteristics of the service, and some of them need to analyze the message content, which may infringe user privacy. In addition, with the continuous emergence of new services, the feature database must be constantly updated so that the original method may be effective, which requires a lot of work.

A new generation of traffic detection methods based on statistical characteristics came into being. This classification method relies on statistical features or time series features and can handle encrypted and unencrypted traffic. These methods usually use classic machine learning algorithms to process analysis [16].

In recent years, machine learning methods have become very popular and have been widely used in fields such as image, sound, and text processing [17–25]. The traditional machine learning methods used in network traffic recognition mainly include Bayesian algorithms, support vector machine (SVM) algorithms, decision tree and integrated learning algorithms, etc. [26]. Distributed denial of service (DDoS) has always been a serious threat to the Internet. Hou et al. [27] proposed a scheme of feature selection and machine learning to identify DDoS traffic. The frequency of malicious activities such as botnets and port scanning is increasing. Although these attacks are simple, they may allow unauthorized network access. Flanagan et al. [28] proposed an MCODT anomaly detection system, which mainly uses the polynomial regression technique of clustering density to detect the abnormal behavior of NetFlow data.

In the power network anomaly detection method, some machine learning methods have also begun to be applied to this area of research.

Fei et al. proposed an improved CUSUM detection algorithm (BF-DT-CUSUM) that dynamically updates the threshold of address statistics, which is used to detect distributed denial of service attacks in power industrial control systems. Simulation experiments verify that the algorithm has good speed and accuracy in response to DDoS attacks [29]. However, the algorithm is difficult to detect other unknown attack types.

In order to more effectively classify the increasing traffic of the power business system and to improve the business processing speed of the power system, Xu proposed a real-time traffic classification method for power business based on an improved random forest algorithm [30]. This method improves the real-time classification by pruning the random forest based on the classification interval weighting. Du et al. analyzed the flow data structure of the power network and verified that the normal flow data of the power has stable information entropy. On this basis, an algorithm based on five-tuple entropy of traffic and SVM is proposed to identify abnormal traffic [31]. Recent research on the use of machine learning algorithms for traffic classification is also mainly focused on the selection of optimized features [32].

Most of these machine learning-based traffic classification methods rely on feature selection, which limits their generalizability.

Recently, deep learning has been well applied in image recognition, natural language processing, and sentiment analysis. These methods can automatically select features through the training process and have strong versatility.

There are three main methods for detecting network traffic anomalies based on deep learning: deep Boltzmann machine [33, 34], stacked autoencoder [35, 36], and CNN [16, 37]. Ertam and Avcı developed GA-WKELM software. This method is based on the combination of genetic algorithm and extreme learning machine and mainly solves the problem of parameter optimization and selection of deep neural networks. However, this method can only complete one training and cannot dynamically update parameters and training samples [38]. Wang et al. proposed an end-to-end traffic classification model. This model processes the network traffic data into a specific file format and then classifies the network traffic through a one-dimensional CNN [39].

Due to business security considerations, the power data dispatching network uses all established channels exclusively and must not be reused. Gao and Yao [40] believe that the statistical characteristics of power communication network traffic have self-similarity, multifractal, periodicity, chaos, and other characteristics. The bandwidth occupied by data traffic of the power data dispatch network can only be a rigid superposition of the business traffic carried. Based on the above characteristics of power network flow, Lv and others used the autoregressive moving average (ARMA) model to predict the power network flow [41]. Lin et al. established a power grid operation situation awareness model based on the fuzzy analytic hierarchy process and the LSTM-attention mechanism [42].

Compared with traditional machine learning methods, deep Boltzmann machine-based methods can extract high-dimensional features of traffic data through learning. However, the robustness of this method is poor. When the input data contain noise, the extracted features will be inaccurate.

The anomaly detection method based on stacked autoencoders can learn traffic data layer by layer and extract traffic features with high accuracy. However, when the traffic data are destroyed, the detection accuracy of this method will be reduced.

The method based on CNN has strong robustness and high detection performance. However, the traditional CNN method generally uses a gradient descent algorithm for training, and if the initial weight of the network is incorrectly selected, it will also affect the learning performance and make the model fall into a local optimal state.

Through the analysis of the advantages and disadvantages of existing deep learning methods in network traffic classification, this paper mainly selects the CNN model to extract the characteristics of power network traffic by itself, eliminating human intervention or relying on expert knowledge. In addition, in order to improve the problem of CNN model parameter selection, considering that the genetic algorithm has an effective search ability for global and local optimal solutions, the genetic algorithm is used to find optimal solutions for CNN model parameters.

3. Background

3.1. Introduction to CNN. CNN was originally proposed by LeCun et al. in 1989 to solve the problem of digital image recognition [43]. CNN is also a neural network specially used to process data with a known grid-like topology. For example, time series data can be regarded as a one-dimensional grid sampled at a certain time interval, and image data can be regarded as a two-dimensional grid composed of pixels. In the calculation, the network mainly uses a mathematical operation called convolution. Convolution is a special linear operation, which can replace the general matrix calculation to achieve multiple operation effects [44]. With the development of CNN, many variants of convolution network structure appear, but their basic structures are mostly similar, mainly including input layer, convolution layer, pooling layer, full connection layer, and output layer. Figure 1 shows the basic structure of CNN network.

The input layer is used to input data or images. In order to facilitate the calculation of convolution layer, the input data needs to be preprocessed.

Convolution layer and convolution kernel of convolution layer are mainly used for feature extraction of input information, and the convolution function is shown in the following formula:

$$x_j^n = f\left(\sum_{i \in D_j} x_i^{n-1} K_{ij}^n + b_j^n\right), \tag{1}$$

where D_j is the input characteristic data, x_j^n is the characteristic value J of the nth layer, K_{ij}^n is convolution kernel function, $f\,()$ is the activation function, and b_j^n is the bias parameter. The activation functions used in this paper are sigmoid, ReLU, and softmax, and their formulas are, respectively,

$$S(x) = \frac{1}{1 + e^{-x}}, \tag{2}$$

$$\operatorname{ReLU}(x) = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0, \end{cases} \tag{3}$$

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}. \tag{4}$$

The convolutional layer and the pooling layer are calculated alternately. The calculation of the pooling layer is

$$x_j^{n+1} = f\left(\sum_j x_j^n \omega_j^{n+1} + b_j^{n+1}\right), \tag{5}$$

where ω_j^{n+1} is the weight constant of the feature map of the pooling layer.

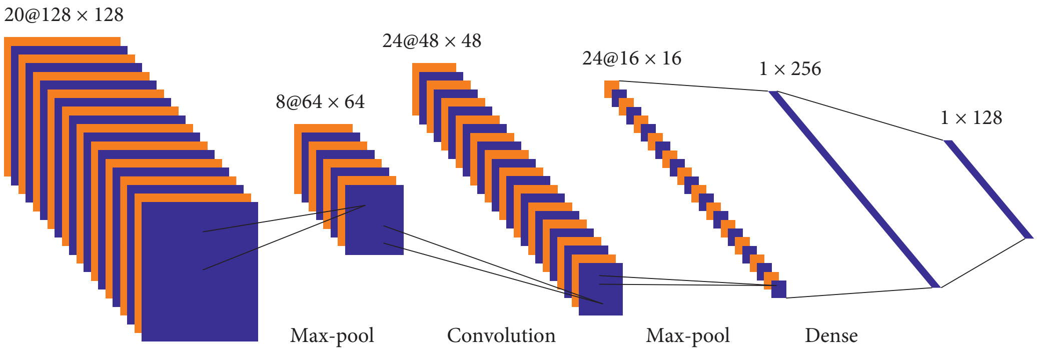


Figure 1: The basic structure of CNN.

Before the output layer gets the result, in the n-layer convolutional neural, as the input sample, f_n represents the activation function of each layer pooling, and ω^n represents the connection weight of each layer. The calculation of this process can be expressed as

$$y = f_n\left(\cdots\left(f_2\left(f_1\left(x \cdot \omega^1\right)\omega^2\right)\right)\cdots\right)\omega^n. \tag{6}$$

The result of forward propagation is calculated, and the error is compared with the known label value. The error function is expressed as

$$\mathrm{E} = \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(y_{ji}' - y_{ji}\right)^2, \tag{7}$$

where N is the number of training samples, M is the number of output neurons, y_{ji}' is the expected output value of the j-th output node of the i-th sample, and y_{ji} is the actual output value of the j-th output node of the i-th sample.

3.2. Genetic Algorithm (GA).

GA was first proposed by John Holland in 1962. The algorithm is designed and proposed according to the laws of biological evolution in nature. It is a computational model used to simulate the natural selection and genetic mechanism of Darwin's theory of biological evolution. As a metaheuristic search strategy, the GA is mainly used to find the best super parameter algorithm of machine learning [45]. Generally speaking, the genetic algorithm is divided into five stages [46]:

(1) Initial population: set the evolution algebra counter $t = 0$, set the maximum evolution algebra T, and randomly generate m individuals as the initial population P (0).

(2) Fitness function: the fitness of each individual in population P (T) was calculated.

(3) Selection: the selection operator is applied to the population. The purpose of selection is to inherit the optimized individual directly to the next generation or to generate new individuals through pairing and crossover and then inherit the next generation. The selection operation is based on the assessment of the fitness of the individuals in the population.

(4) Crossover: the crossover operator is applied to the population. Crossover operator plays a key role in the genetic algorithm.

(5) Mutation: the mutation operator is applied to the population. That is to change the gene value of some loci in the individual string of the population. The next-generation population P ($T + 1$) was obtained after selection, crossover, and mutation.

After the above five steps, the termination condition is judged. If $t = T$, the individual with the greatest fitness obtained in the evolution process is used as the optimal solution output, and the calculation is terminated.

3.3. Power Network Flow Collection and Metadata Generation.

In order to establish a scientific power network traffic monitoring model, it is necessary to collect the entire network traffic in real time and quickly judge and process abnormal network traffic, so as to reduce the difficulty and time of power network fault diagnosis and abnormal detection.

According to actual operation requirements, the traffic information is collected through network probes, as shown in Figure 2. According to the hierarchical characteristics of the power network, this model needs to deploy traffic collection probes on the links between the edge networks of different levels of companies and the IDC network, so as to achieve the purpose of obtaining all network traffic comprehensively and accurately.

Distributed hardware probes are deployed on different network nodes and are responsible for the traffic collection of the target network. The flow detection remote management center provides unified network monitoring and network analysis and management. The local visual management platform can perform real-time monitoring and retrospective analysis on the specified network link and can also perform playback analysis on local data packets.

In order to fully obtain the power network flow characteristic data, this paper extracts the flow characteristic metadata from the four dimensions of area, time, business, and link, so that a complete network flow analysis plan can be made. Some of the metadata definitions are shown in Table 2.

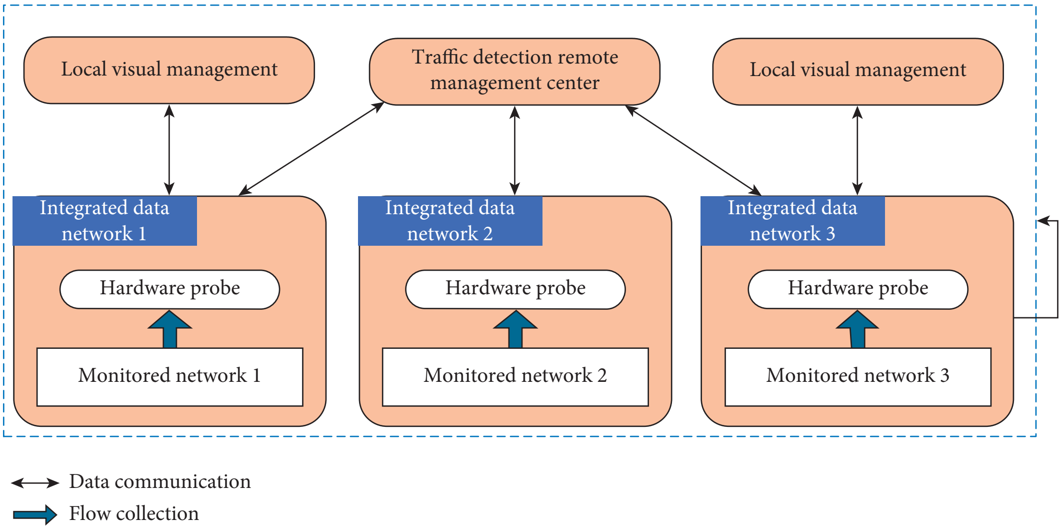


Figure 2: Network traffic collection architecture.

4. Method Principle and Process

This section first introduces the process of power network flow metadata preprocessing, then performs statistical analysis on the metadata, and finally focuses on the detailed description of how the GA algorithm optimizes the CNN parameters. The proposed GACNN power network traffic detection method is mainly divided into three stages: data preprocessing, model training, and model verification, and its overall framework is shown in Figure 3. In the preprocessing stage, a series of processing is performed on the collected traffic metadata, including data encoding, data normalization, data shaping, and data splitting. The preprocessed data will be converted into pictures for training and learning in the GACNN model.

4.1. Data Preprocessing. Network metadata are string data composed of multiple information elements extracted by the probe from the original network traffic. Each information element in the metadata occupies a fixed position in the character string, and the character strings are separated by a ˆ sign, and the final character string also ends with a ˆ. For the information element that does not have a value, the position does not need to be filled with any content; that is to say, the two ˆ are adjacent at this time.

For example, an extracted piece of network traffic metadata is shown in Figure 4.

For certain feature data of metadata that only contain a few types of data, tag coding can be used to solve it, such as protocol field: http code is 1, ftp code is 2, etc. For some very long strings contained in metadata, for example, the content of the DPI package is “Welcome to the Changsha city!” after byte encoding, it is “235619648990998464306777891175220 094081099454301910254853997472753078 5.”

Data standardization can improve the accuracy of the model. At present, there are many data standardization methods. In summary, they can be divided into linear methods (such as extreme value method and standard deviation method), broken line method (such as three-fold line method), and curve method (such as seminormal distribution). This article mainly adopts the min-max method, and its formula is as follows:

$$y_i = \frac{x_i - \min_{1\le j\le n}\{x_j\}}{\max_{1\le j\le n}\{x_j\} - \min_{1\le j\le n}\{x_j\}}, \quad (8)$$

where x_i is the input data sequence, y_i is the output sequence after the change, and the value is between [0, 1].

For the input of the CNN model, the format should generally be three-dimensional data (height, width, and channel). The data after data encoding and standardization are only some two-dimensional feature vector data. These data cannot be directly used for model training and testing. In this paper, each line of traffic metadata is reshaped according to the proportion of (50 * 50 * 3). When the length of metadata is less than 2500, the same value of metadata is used to fill. The purpose of this is to further increase the reliability of feature extraction. In addition, the channel value is filled according to the data type. For example, if the metadata are an abnormal type in the second category, it is filled with 1, and the normal type is filled with 0. In the multicategory, it is filled with the corresponding data. Figure 5 shows a data image obtained by reshape of a certain piece of traffic metadata.

4.2. Statistical Analysis of Traffic Metadata. In order to quickly and accurately understand the statistical characteristics of the traffic metadata collected by different probes, in addition, the metadata contain 217 explicit characteristics, but they lack some spatiotemporal implicit statistical characteristics, which is not convenient for deep understanding and analysis of abnormal traffic information. In this paper, the characteristics of traffic metadata in different time periods are statistically analyzed, so as to obtain some important supplementary implicit features.

Table 2: Some of the metadata definitions of network traffic.

ID	Name	Type	Length	Description	Dimension
1	EventID	String	64	Event unique ID	
2	OccurTime	Long	8	The time when the event first occurred	Time
3	RecentTime	Long	8	Time of the last occurrence of the event	Time
4	SrcGeographyLocationCountryOrRegion	String	128	Country or region of source IP	Area
5	DestGeographyLocationCountryOrRegion	String	128	Country or region of destination IP	Area
6	OriEventType	String	1024	Original event type	Business
7	EventSubType	String	128	Event subtype	Business
8	AbnTrfBaseline	String	20480	Baseline value of abnormal flow	Link
9	ESN	String	64	Link device serial number	Link

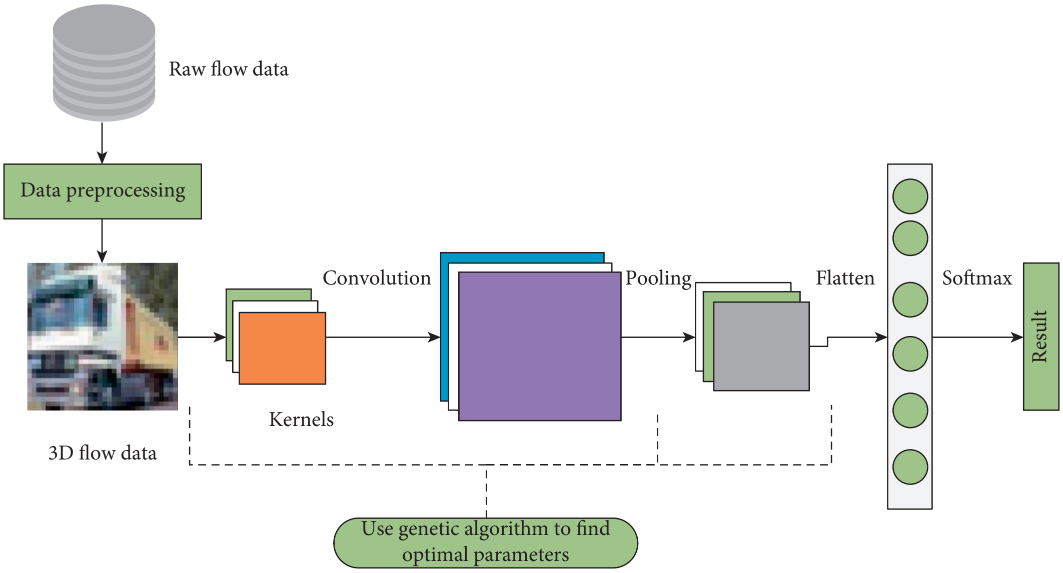
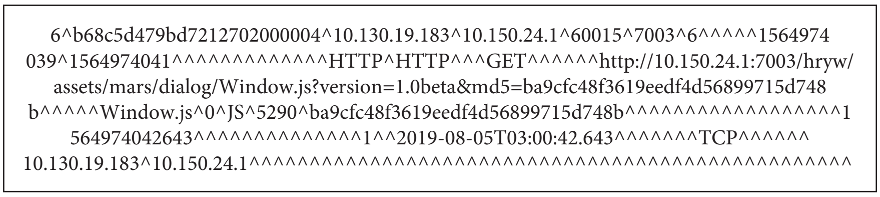


Figure 3: The overall framework of GACNN.



```
6^b68c5d479bd7212702000004^10.130.19.183^10.150.24.1^60015^7003^6^^^^^1564974
039^1564974041^^^^^^^^^^^^^HTTP^HTTP^^^GET^^^^^^http://10.150.24.1:7003/hryw/
assets/mars/dialog/Window.js?version=1.0beta&md5=ba9cfc48f3619eedf4d56899715d748
b^^^^^Window.js^0^JS^5290^ba9cfc48f3619eedf4d56899715d748b^^^^^^^^^^^^^^^^^^1
564974042643^^^^^^^^^^^^^^1^^2019-08-05T03:00:42.643^^^^^^^TCP^^^^^^
10.130.19.183^10.150.24.1^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
```

Figure 4: A piece of element probe traffic metadata.

Figure 6(a) shows the number of uplink bytes of a stream (octetDeltaCount) in the 2019 year, and Figure 6(b) shows the relationship between the number of uplink bytes (octetDeltaCount) and TCP and UDP source ports (sourceTransportPort).

Figure 7 shows the change of traffic ports from July 26, 2019, to August 5, 2019. It can be observed that the flow port is relatively stable, but the fluctuation is relatively large on August 2, 2019.

Figure 8 shows the main distribution of network traffic source ports in 2019 (in Figure 8(a)) and the distribution of traffic source ports in July and August of 2019 (in Figure 8(b)). It can be seen from the right figure that the source ports are still quite different in July and August of 2019.

Figure 9 shows the statistical distribution of source ports of network traffic, mainly including raw data, rolling average, and rolling standard deviation. It can be seen from the figure that these values are mainly concentrated between 10000 and 25000.

4.3. GA Optimizes the Parameters of CNN. GA is an algorithm that simulates biological evolution for individual selection, crossover, and mutation. Its main core is parameter coding, initial group setting, and fitness function

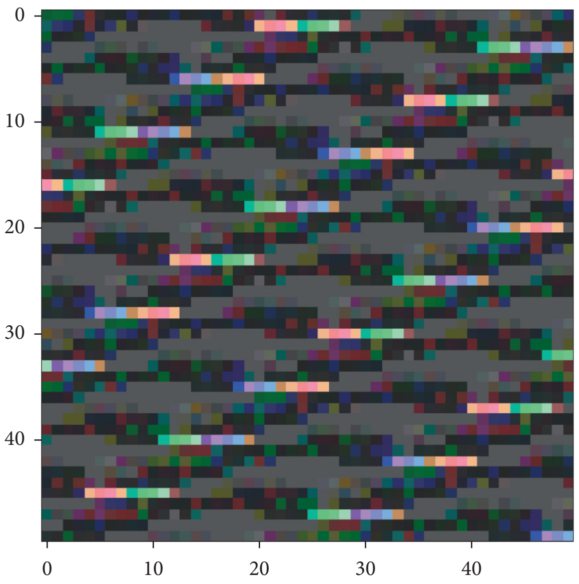


Figure 5: A certain flow metadata image after reshape.

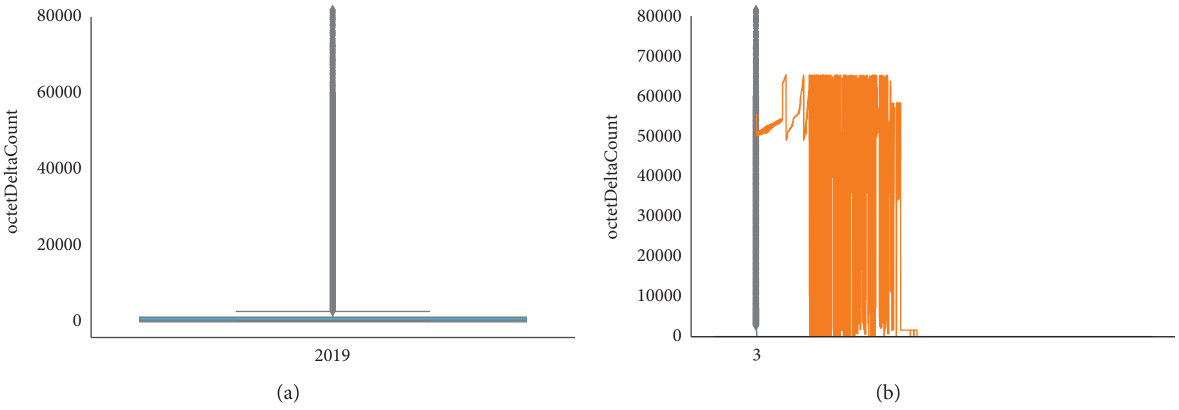


Figure 6: The number of uplink bytes of a stream (octetDeltaCount) in 2019. (a) Yearly load. (b) sourceTransportPort.

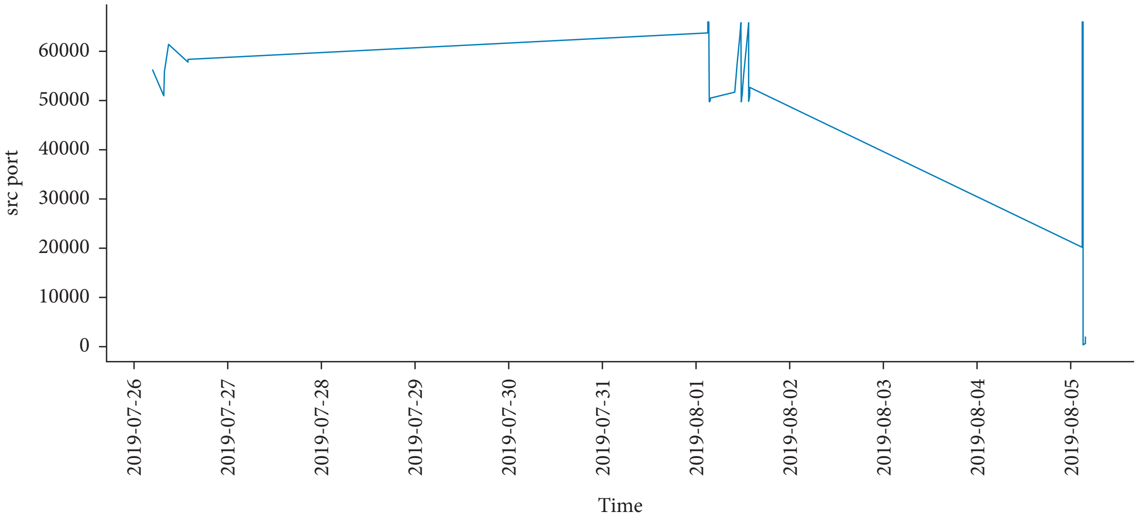


Figure 7: The change of traffic ports from July 26, 2019, to August 5, 2019.

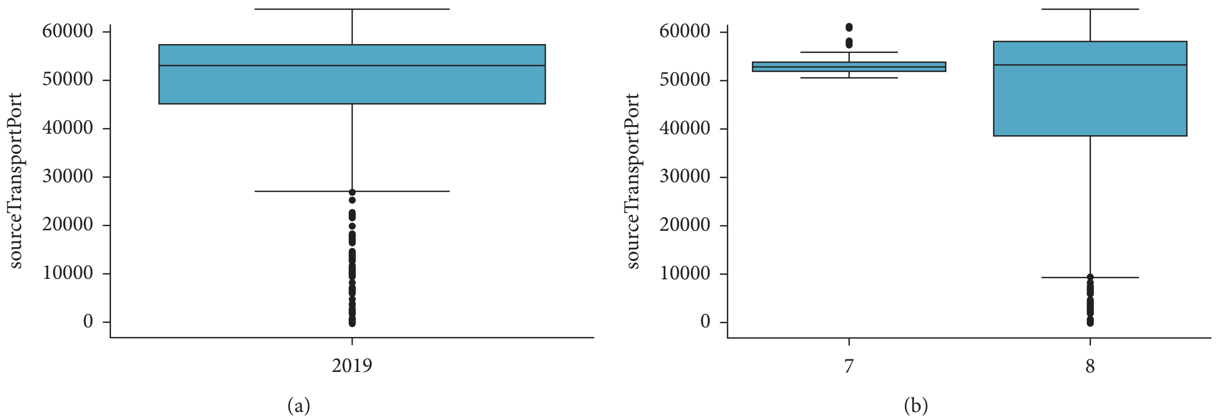


Figure 8: The distribution of network traffic source ports in 2019. (a) Yearly load. (b) Month load.

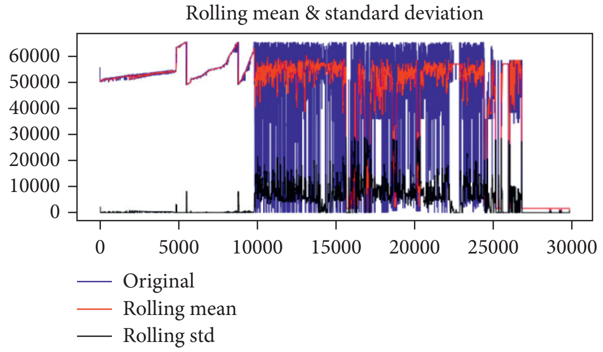


Figure 9: The statistical distribution of source ports of network traffic.

determination, and the optimal solution is obtained through the final search. In this paper, the GA is used to optimize the parameters of CNN such as weight, bias, and optimizer selection, and the optimization process is shown in Figure 10.

In the initial population stage, some important parameters of CNN are randomly set, mainly including the filters, the kernel_size, the activation parameters of the Conv2D layer, the loss rate of the dropout layer, and the unit parameters of the dense layer.

(i) Filters: the number of CNN filters (16, 32, 48, 64).

(ii) Kernel_size: the shape and size of the filter in CNN. The initial random setting is as follows: (2 ∗ 2), (3 ∗ 3) or (5 ∗ 5).

(iii) Activation: the activation function used in the CNN model is randomly selected from {relu, selu, elu, sigmoid, tanh}.

(iv) Loss rate: randomly generated values between [0.1, 0.5].

(v) Optimizer: select randomly from [“adamax,” “adadelta,” “adam,” “adagrad,” “nadam”].

(vi) Pooling: pooling (none, maximum pooling, average pooling).

(vii) Loss: the loss is used to match the function of the network, select randomly from [“categorical_crossentropy,” “mse,” “ focal_loss”].

In the selection stage, an individual is randomly selected from the generated population for model training.

In the crossover stage, two crossover points are randomly set in the individual coding string, and then partial gene exchange is performed.

In the mutation stage, the mutation operation is performed on the individual coding string with mutation probability and the value of a random bit.

5. Experiment and Results

In order to evaluate the proposed abnormal traffic detection scheme, this article uses Python, Scikit-learn, NumPy, Pandas, TensorFlow, and Keras to conduct training and testing on a 64 bit Windows computer, which is configured as Intel(R) Core(TM) i3- 4005U 1.7G CPU, 8 GB RAM, 250G solid state drive.

The labeled dataset is a key factor to ensure the performance of deep learning. In order to verify the performance of the GACNN method, the internationally public network

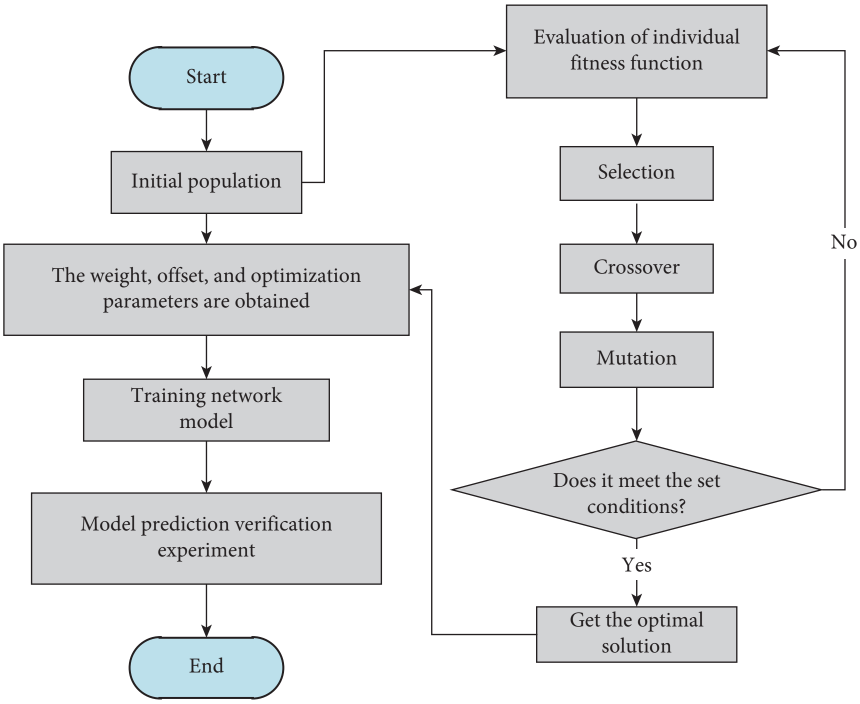

Start
Initial population
The weight, offset, and optimization parameters are obtained
Training network model
Model prediction verification experiment
End
Evaluation of individual fitness function
Selection
Crossover
Mutation
Does it meet the set conditions?
No
Yes
Get the optimal solution

Figure 10: The process of optimizing CNN parameters by GA.

intrusion dataset UNSW_NB15 [47] and a part of the traffic dataset are collected by a probe of a Chinese power company.

This article uses accuracy, precision, recall, and F1-scores to evaluate model detection performance. In order to measure the values of these 4 indicators, it is generally calculated by using true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Accuracy is the proportion of correctly classified samples to the total number of samples, which is defined as

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}. \tag{9}$$

Precision: its meaning is the ratio of samples that are actually positive samples among all the samples predicted to be positive, and its expression is

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{10}$$

Recall rate: its meaning is the ratio of positive cases predicted to be positive in a sample that is actually positive, and its expression is

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{11}$$

F1-score: the F1-score considers both the precision rate and the recall rate, so that the two can reach the highest at the same time and strike a balance. The F1-score expression is

$$\mathrm{F1} = \frac{2 \times P \times R}{P + R}. \tag{12}$$

In these formulas, TP is the number of successful detections of the current network traffic category, TN is the number of other network traffic types successfully detected, FP is the number of other network traffic categories identified as the current network traffic category, and FN is the number of current network traffic categories identified as other network traffic categories.

5.1. Tested on UNSW-NB15 Dataset.

5.1. Tested on UNSW-NB15 Dataset. The original network packets (Pcap files) of the UNSW-NB15 dataset were created by the IXIA PerfectStorm tool in the network-wide laboratory of the UNSW Canberra Network Center, which is used to generate network traffic test of real normal activity and synthetic attack activity. The source files of the dataset are divided based on the simulation dates of January 22, 2015, and February 17, 2015, respectively [47].

The total training dataset selected in the experiment contains 175341 records, and the test set contains 82332 records. These records come from different types of attacks and normal conditions. Furthermore, there are 9 types of the attacks including Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms. All the features in the dataset can be divided into 6 subtypes, namely, flow features, basic features, content features, time features, additional generated features, and labelled features. Table 3 lists the 12 features included in the dataset.

5.1.1. Binary-Classification Test. In order to improve the understanding and analysis of datasets, this paper carries out a series of preprocessing operations before binary

Table 3: The 12 features included in the UNSW-NB15 dataset.

No.	Feature name	Type	Feature description	Feature classification
1	srcip	Nominal	Source IP address	Flow features
2	sport	Integer	Source port number	Flow features
3	proto	Nominal	Transaction protocol	Flow features
4	sbytes	Integer	Source to destination bytes	Basic features
5	dbytes	Integer	Destination to source bytes	Basic features
6	swin	Integer	Source TCP window advertisement	Content features
7	smeansz	Integer	Mean of the flow packet size transmitted by the src	Content features
8	sjit	Float	Source jitter (mSec)	Time features
9	djit	Float	Destination jitter (mSec)	Time features
10	ct_flw_http_mthd	Integer	No. of flows that have methods such as get and post in http service.	Additional generated features
11	ct_ftp_cmd	Integer	No of flows that have a command in ftp session	Additional generated features
12	attack_cat	Nominal	The name of each attack category	Labelled features

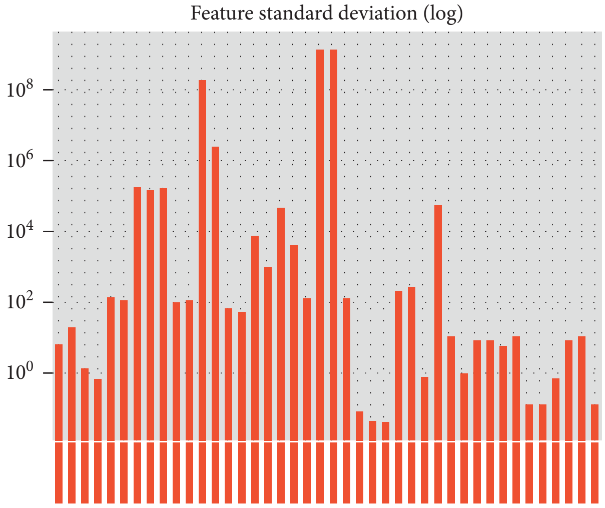


Figure 11: The standard deviation of features included in the UNSW-NB15 dataset.

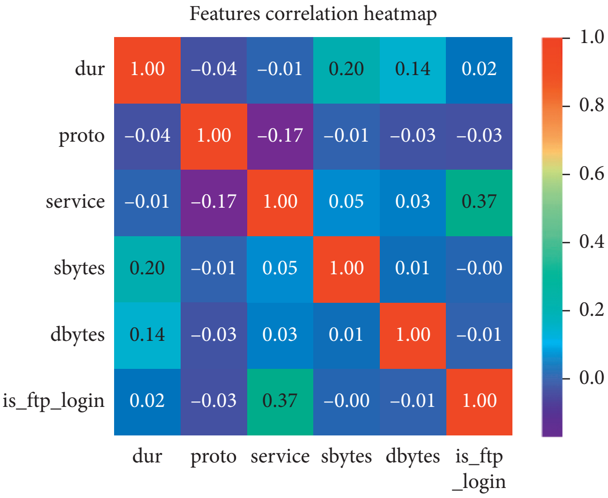
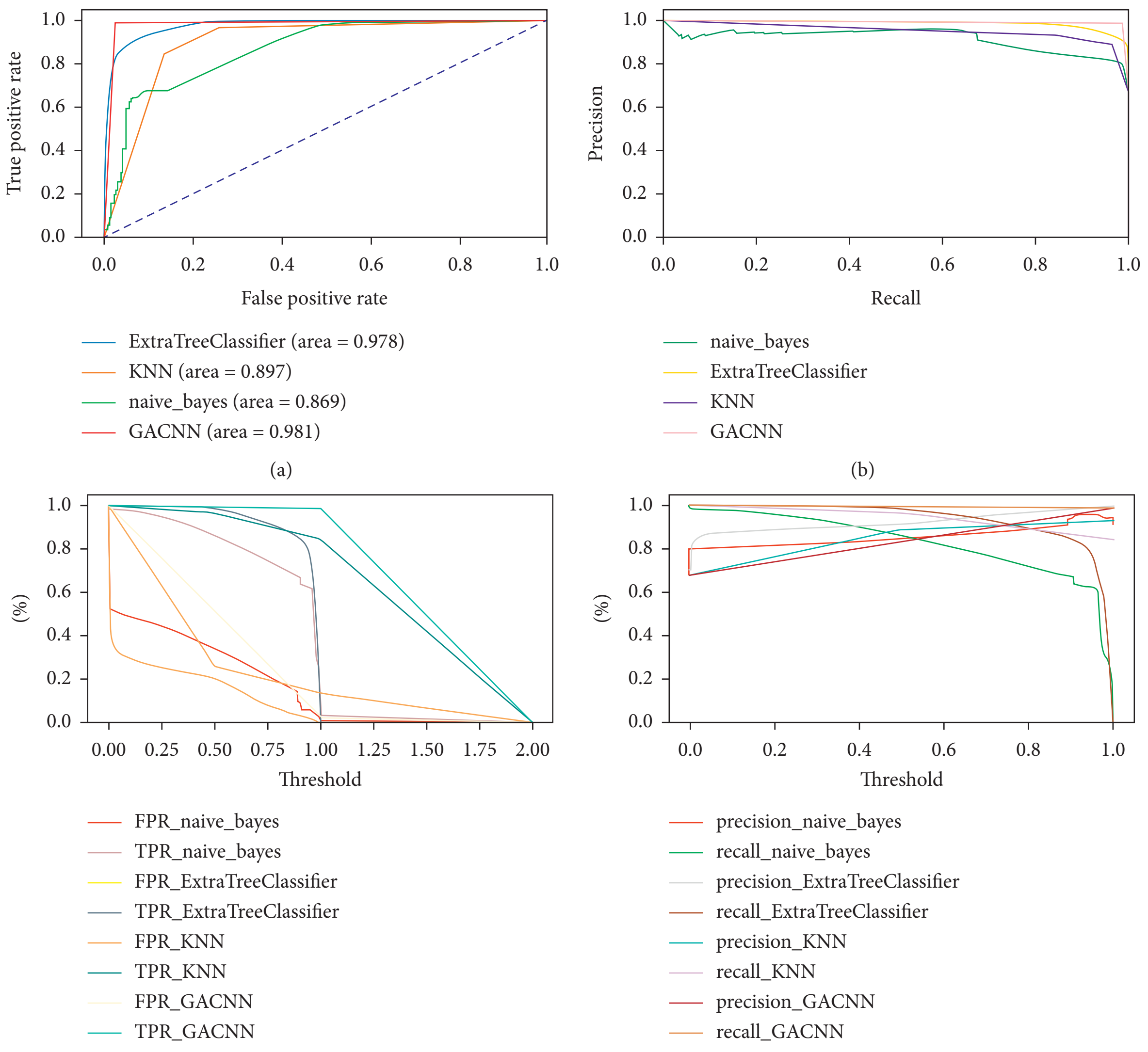


Figure 12: Heat map of some features in the UNSW-NB15 dataset.

classification, mainly including standardization, deduplication, dimension reduction, and feature correlation analysis. Among them, the results of standardization and correlation analysis are shown in Figures 11 and 12. As can be seen from Figure 11, the values of sload, dload, stcpd, dtcpd, and other characteristics after the standard are more



(a)

(b)

(c)

(d)

Figure 13: Binary-classification results in the UNSW-NB15 dataset.

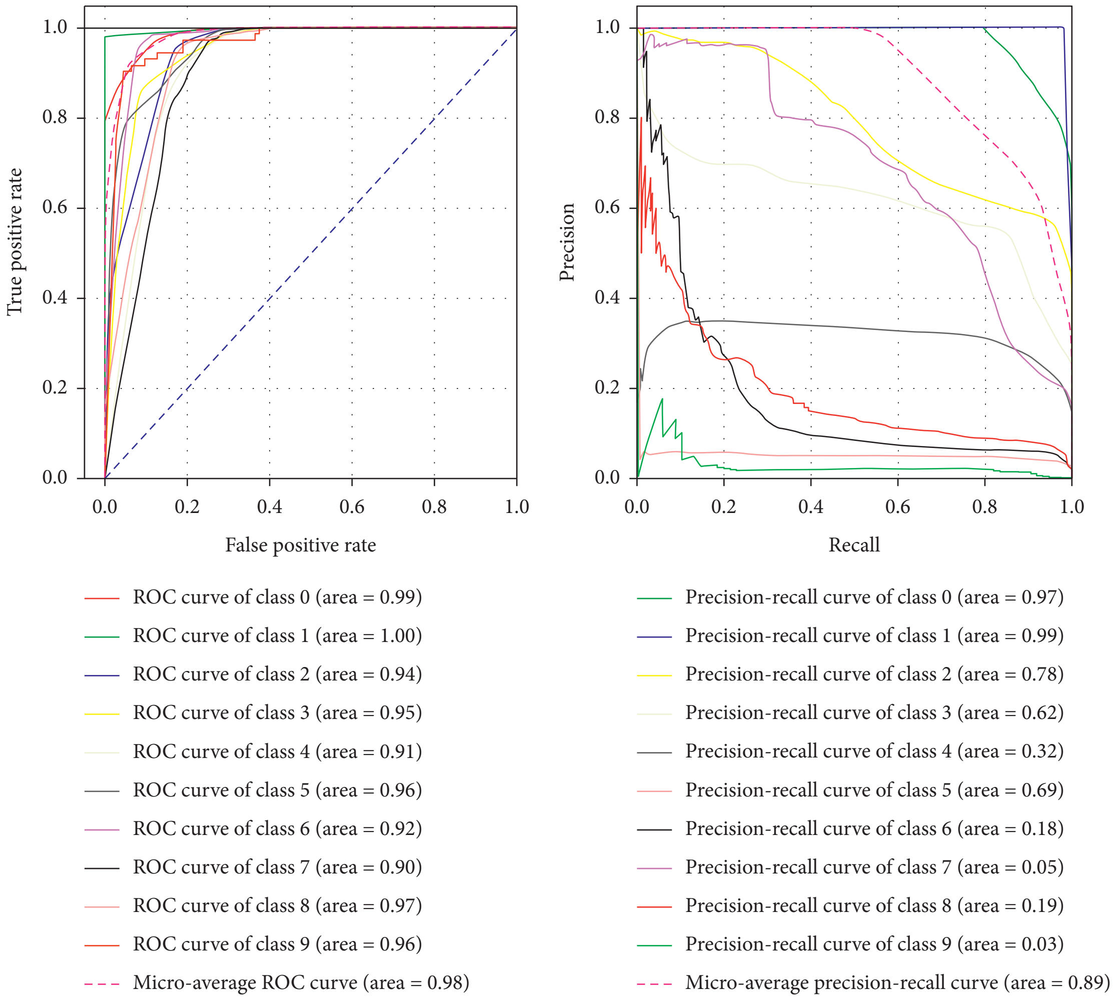
prominent, reflecting the importance of these characteristics. It can be seen from Figure 12 that the selected eigenvalues have very little correlation. For example, the correlation between proto and service is -0.17, indicating that the selected eigenvalues are very suitable.

In order to better verify the classification effect of the model, a simple CNN model and some traditional machine learning methods such as ExtraTreeClassifier [48], KNN [49], and naïve_bayes [50] are selected for comparison. The experiment uses a dichotomy method from the training dataset to divide the data into two, and both the training and test sets contain (87670, 43) feature data. This experiment is only for the identification of normal and abnormal network flows, and the experimental results are shown in Figure 13.

This experiment is only for the identification of normal and abnormal network flows. This paper selects ROC and AUC curves commonly selected in machine learning to describe the classification accuracy of the model. The full name of ROC is "Receiver Operating Characteristic." The area of the ROC curve is AUC (Area Under the Curve). AUC is used to measure the performance (generalization ability) of machine learning algorithms for "two classification problems." The most ideal classifier is to classify the sample completely correctly, that is, FP = 0 and FN = 0. So, the ideal classifier TPR = 1 and FPR = 0.

The experimental results are shown in Figure 13. It can be seen from the ROC graph in Figure 13 that the AUC value of ExtraTree is 0.978, KNN is 0.897, naive_bayes is 0.869, and GACNN is 0.981. The test results show that the GACNN method is better than the previous three methods.

The same situation also appeared in the test results of the other three indicators. In the graph composed of precision and recall, GACNN is relatively stable when the recall value changes from 0.0 to 1.0, and most of them are fixed at 1.0. However, the other three algorithms fluctuate greatly. In the graph composed



(a)

(b)

Figure 14: The classification results of GACNN.

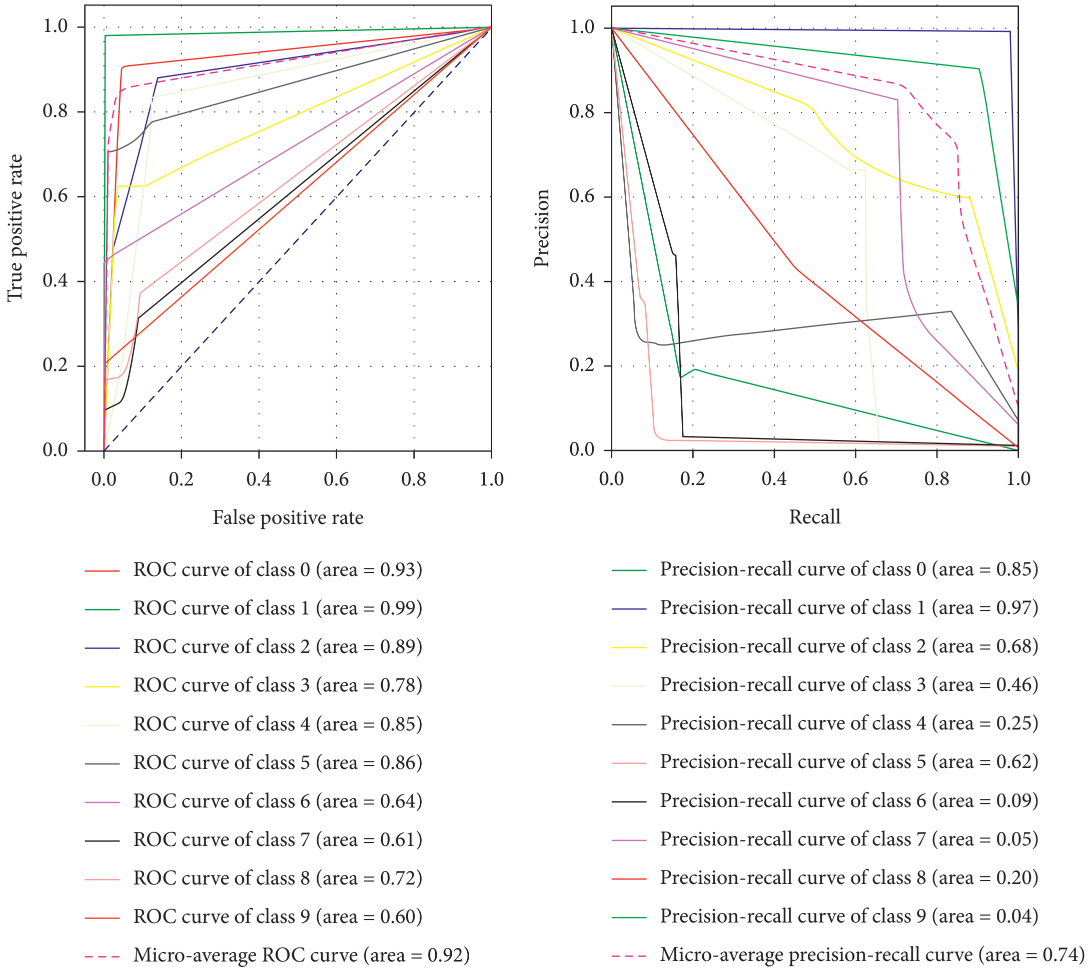
of threshold and %, when the training threshold is from 0.0 to 1.0, the TPR value of GACNN is basically stable at 1.0. The other three methods are generally lower than 1.0.

5.1.2. Multiclassification Test. In order to accurately classify different network traffic, this paper further tests this dataset and compares the classification effects of the previous four methods in multiclassification. The experiment first performed a series of preprocessing on the metadata and divided the dataset into two subdatasets for training and testing. In addition, both data subsets contain (87670, 43) dimensional feature values and 10 types of attack data. The coding of the 10 attack types is as follows: {0: "Normal," 1: "Generic," 2: "Exploits," 3: "Fuzzers," 4: "DoS," 5: "Reconnaissance," 6: "Analysis," 7: "Backdoor," 8: "Shellcode," and 9: "Worms"}.

The GACNN model has been trained many times to find the optimal parameters of the model and then use the optimal parameter training and test data. The experimental results of the four methods are shown in Figures 14–17.

In Figure 14, the classification AUC value of the GACNN method for each category exceeds 0.91, and the average AUC value of all types reaches 0.98. In addition, from the precision-recall diagram in Figure 14, it can be seen that AUC values of model classification for 0, 1, 2, 3, and 5 types of network traffic are all more than 0.6, and the average AUC value is 0.89, but the AUC value of 4, 6, 7, 8, and 9 types of network traffic is lower.

In the left subgraph of Figure 15, the classification AUC value of the ExtraTree method for each category is relatively good, basically exceeding 0.6, and the average AUC value of all types also reaches 0.92. In addition, in the right subgraph of Figure 15, the AUC values of the model classification for 0, 1, 2, and 5 types of network traffic are all greater than 0.6, and the average AUC value is 0.74, but the AUC value of 3, 4, 6, 7, 8, and 9 types of network traffic is also lower.



(a)

(b)

Figure 15: The classification results of ExtraTreeClassifier.

In the left subgraph of Figure 16, the classification AUC value of the KNeighbors method for each category is relatively good, basically exceeding 0.6, and the average AUC value of all types also reaches 0.96. In addition, in the right subgraph of Figure 16, the AUC values of the model classification for 0, 1, 2, and 5 types of network traffic are all greater than 0.6, and the average AUC value is 0.86, but the AUC value of 3, 4, 6, 7, 8, and 9 types of network traffic is also lower.

In the left subgraph of Figure 17, the classification AUC value of the naïve_bayes method for each category is relatively balanced, basically exceeding 0.80, but the average AUC value of all types is only 0.82. In addition, in the right subgraph of Figure 15, only the AUC value of the model classification of type 0 network traffic is greater than 0.6, and the AUC values of the other types of network traffic are very low.

From the above experimental results, it can be known that in the multiclassification test of the UNSW-NB15 dataset, the GACNN method is basically better than the other three methods.

5.2. Test on Real Power Dataset. In order to further verify the anomaly classification effect of the proposed method in real network scenarios, the continuous flow data grouping of a power company information network in July and August 2019 was obtained through flow probes and processed and analyzed in the experimental platform. The preprocessed data are used as the input of the experimental platform, and the performance of the algorithm is evaluated through the processing of different algorithms. Some preprocessing results are shown in Figures 5–9.

Due to the different parameters of CNN, the detection performance of the network will be different. In order to evaluate the classification performance of the GACNN method, some initial parameter settings are shown in this experiment in Tables 4 and 5.

The experiment collected normal network flows and three abnormal network flows (C&C attack, SSH brute force attack, and webshell attack). Among them, the normal stream is 563 MB, the webshell stream is 8.8 MB, and the

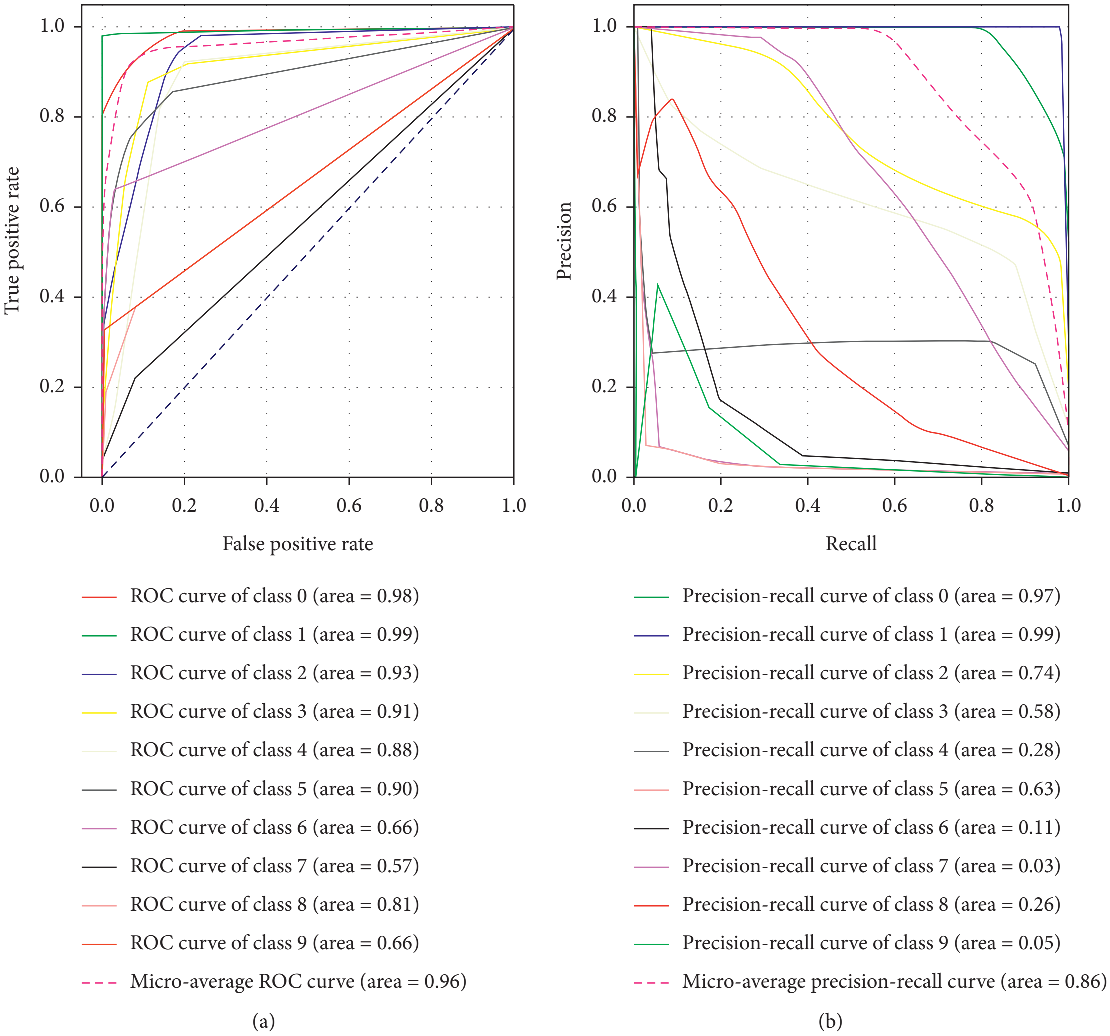


Figure 16: The classification results of KNeighborsClassifier.

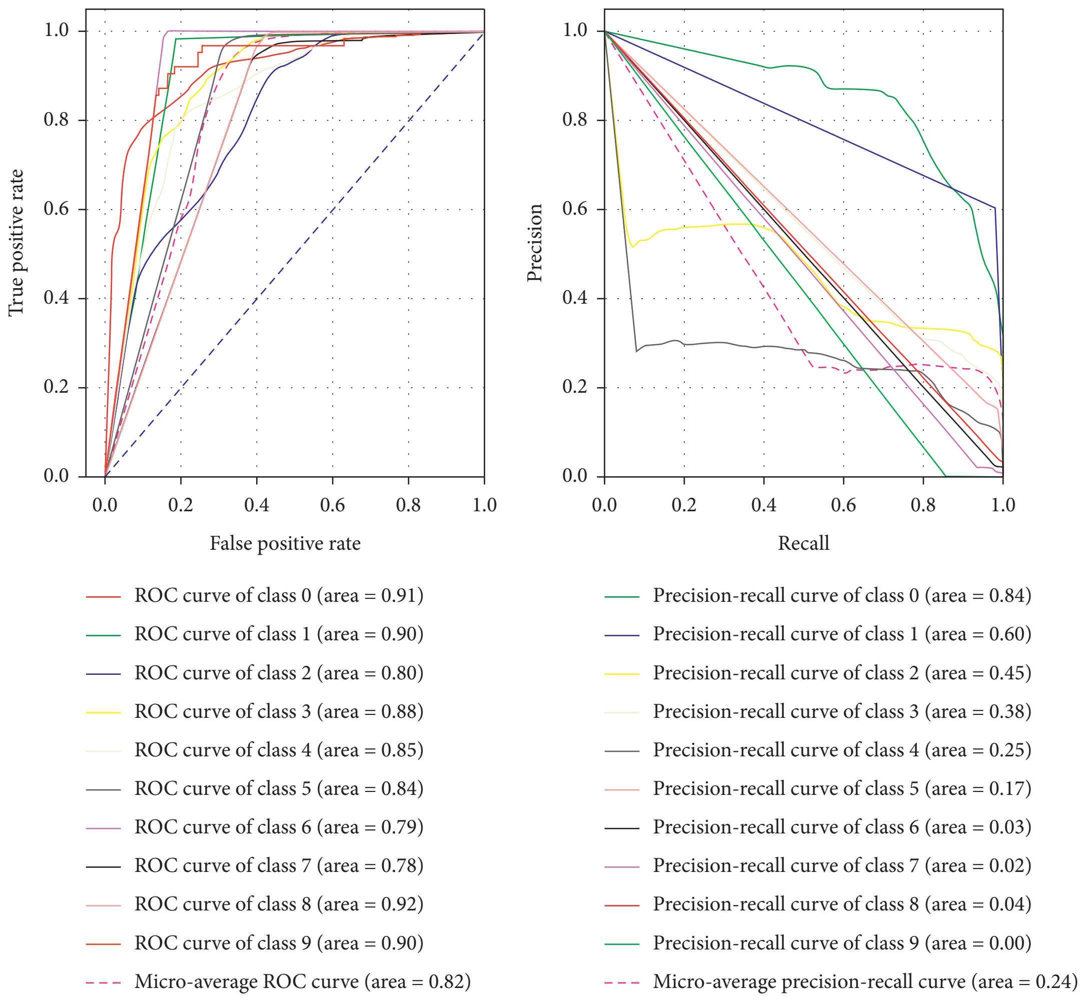
SSH stream is 5.74 MB. However, C&C is relatively small, only 24 kB. These probe stream metadata are converted into CSV files after preprocessing. In the training and detection stage of the CNN model, the original CSV format data cannot be directly used. Then, merge the CSV files of different attack types and convert them into an NPY file.

In GACNN model training, the dataset is divided into training set and test set, and the training set contains (216066, 50, 50, and 3) records, and the test set contains (72022, 50, 50, and 3) records. The GACNN model contains 5 two-dimensional convolutional layers Conv2D, 5 batch_normalization layers, 5 max_pooling2d layers, a flatten layer, 2 dense layers, and 2 dropout layers.

Like the previous experiment, this experiment also selects the former three machine learning methods for comparison. Table 6 lists the detection results of these algorithms for three attack types (ExtraTree-E, KNeighbors-K, naïve_bayes-N, and GACNN-G). As can be seen from the results in Table 6, compared with the other three detection methods, the convolution neural network optimization method based on the genetic algorithm has greatly improved the detection rate and unknown attack detection rate, reduced the false alarm rate, and achieved good results.

For example, in binary detection, the precision, recall, and F1-score of GACNN are all higher than 0.96, while other methods are lower than 0.95.

In the multiclassification detection, the precision values of the other three methods are lower than 0.93, recall values are lower than 0.96, and the F1-score values are basically lower than 0.94. On the contrary, the three detection values obtained by GACNN method are relatively good, and all of them reach the values above 0.96.



(a)

(b)

Figure 17: The classification results of naïve_bayes.

Table 4: The initial parameters of CNN.

Parameters	Value
Filters	32, 64
Kernel_size	(3, 3), (5, 5),
Activation	Relu, selu, elu
Input_shape	(50, 50, 3)
Output_shape	2, 10
Dropout_rate	0.1, 0.2, or random number between 0.1 and 0.5
Optimizer	“adamax,” “adadelta,” “adam,” “adagrad”

Table 5: The initial parameters of GA.

Parameters	Value
Mutation probability	0.01 or the random number is greater than the threshold value
Crossover probability	0.5
Reproduction algebra	5
Population size	10

Table 6: The results of four methods in power network flows.

Class	Type	Precision				Recall				F1-score			
		E	K	N	G	E	K	N	G	E	K	N	G
Two	Normal	0.928	0.921	0.892	1.00	0.962	0.976	0.949	1.00	0.945	0.947	0.920	1.00
	Abnormal	0.917	0.928	0.873	1.00	0.885	0.875	0.837	0.962	0.901	0.901	0.854	0.975
Multi	C&C	0.905	0.805	0.776	0.973	0.911	0.940	0.932	0.993	0.908	0.867	0.846	0.983
	SSH	0.863	0.922	0.892	0.963	0.917	0.905	0.852	0.982	0.889	0.913	0.872	0.972
	Webshell	0.799	0.852	0.925	0.974	0.957	0.950	0.961	0.989	0.870	0.898	0.943	0.981
	Whole	0.847	0.907	0.853	0.978	0.956	0.959	0.950	0.986	0.898	0.932	0.899	0.982

6. Conclusion

With the continuous development of smart grids, the current power information network is constantly expanding, and the possibility of failures in the network is also increasing. Aiming at the high false alarm rate and low detection efficiency of current network traffic anomaly detection, this paper proposes a genetic algorithm-optimized convolutional neural network method to deal with power network traffic anomaly detection.

Compared with the traditional machine learning method, this paper mainly innovates and improves from the following aspects:

(1) According to the business characteristics and requirements of the power system, this paper establishes a network flow metadata collection model for the power system. This model mainly monitors network traffic and equipment indicator status from the four dimensions of time, area, event, and link, thereby more effectively improving the quality of network services.

(2) Aiming at the problem that the classification accuracy of CNN depends on parameter settings, this paper proposes to use the genetic algorithm to find the best CNN parameters, which can quickly improve the training accuracy of the CNN method. From the experimental results, this method is superior to other detection methods in the anomaly detection of network flow.

(3) The classification accuracy of traditional machine learning largely depends on the reasonable selection of network features. The method proposed in this paper uses raw traffic or metadata directly as model input and trains features through self-learning without manual intervention. In addition, the hidden spatiotemporal features and package content analysis can be completed through multiple convolutional learning and spatiotemporal analysis of the model, thereby reducing the complexity of the task and improving the accuracy of the model classification.

In order to further improve the accuracy and efficiency of the method, the next step is to continue to work on the following aspects:

(1) Improved CNN structure: the network in this algorithm is based on a simple 2-dimensional CNN model. In future work, you can try to use ResNet, VGGNet, and GoogLeNet models to build deeper mixed networks to further improve classification accuracy.

(2) Continue to increase the testing of the method in large-scale power network traffic to realize practical application in engineering.

Data Availability

Two datasets are used in the paper, among which UNSW-NB15 can be downloaded directly from the Internet and the other is the enterprise internal test dataset. The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the key scientific and technological project of “Research and Application of Key Technologies for Network Security Situational Awareness of Electric Power Monitoring System of China Southern Power Grid Corporation, (no. ZDKJXM20170002)” the teaching reform and scientific research project of “OBE-Based Engineering Education Professional Certification and Evaluation Information System” of Changsha University of Science and Technology, and Open fund project of the Hunan Provincial Engineering Research Center of Electric Transportation and Smart Distribution Network (Changsha University of Science and Technology) (no. 3040102-1105008).

References

[1] J. Liu and J. Weng, “Review of smart grid security,” *Information Network Security*, vol. 5, pp. 78–84, 2016.

[2] X. Wang and X. Cheng, “Overview of network security technology in smart grid,” *Computer CD Software and Application*, vol. 20, pp. 159–161, 2013.

[3] W. Wang and Z. Lu, “Cyber security in the smart grid: survey and challenges,” *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, 2013.

[4] L. Xu, “Network communication architecture and key technologies of smart grid,” *Electrical Technology*, vol. 8, pp. 16–20, 2010.

[5] Z. Xia, J. Tan, K. Gu, and W. Jia, “Detection resource allocation scheme for two-layer cooperative IDSs in smart grids,” *Journal of Parallel and Distributed Computing*, vol. 147, pp. 236–247, 2021.

[6] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Securing advanced metering infrastructure using intrusion detection system with data stream mining," *Lecture Notes in Computer Science*, Springer, vol. 7299Berlin, Germany, , 2012.

[7] H. Kasai, W. Kellerer, and M. Kleinsteuber, "Network volume Anomaly detection and identification in large-scale networks based on online time-structured traffic tensor tracking," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 636–650, 2016.

[8] L. Wu, S. Zhang, T. Lei, and Y. Cai, "A traffic detection model for power system data network," *Information and Communication*, vol. 7, pp. 45-46, 2013.

[9] Y. Zhang, X. Li, D. Li et al., "Abnormal flow detection of industrial control network based on convolutional neural network," *Computer Application*, vol. 39, no. 5, pp. 1512–1517, 2019.

[10] D. Li, X. Wang, B. Yu, and T. Huang, "Network traffic classification method based on one-dimensional convolution neural network," *Computer Engineering and Applications*, vol. 56, no. 3, pp. 94–99, 2020.

[11] P. Wang, Q. Zheng, G. Niu et al., "Port scan detection algorithm based on traffic statistics," *Acta Communication Sinica*, vol. 28, no. 12, pp. 14–18, 2007.

[12] Z. Sun, J. Zhai, and Y. Dai, "An encryption flow identification method based on DPI and load randomness," *Journal of Applied Sciences*, vol. 37, no. 5, pp. 711–720, 2019.

[13] J. Garcia, "A clustering-based analysis of DPI-labeled video flow characteristics in cellular networks," in *Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pp. 991–994, Lisbon, Portugal, May 2017.

[14] Y. Wang and X. Wei, "A security model of ubiquitous power internet of things based on SDN and DFI," in *Proceedings of the 2020 Information Communication Technologies Conference (ICTC)*, pp. 55–58, Nanjing, China, May 2020.

[15] H. Chen, Z. Hu, Z. Ye, and W. Liu, "A new model for P2P traffic identification based on DPI and DFI," in *Proceedings of the 2009 International Conference on Information Engineering and Computer Science*, pp. 1–3, Wuhan, China, December 2009.

[16] X. Liu, Z. Tang, and B. Yang, "Predicting network attacks with CNN by constructing images from NetFlow data," in *Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pp. 61–66, Washington, DC, USA, May 2019.

[17] C. Xu, Z. Su, Q. Jia, D. Zhang, Y. Xie, and A. Yang, "Neural dialogue model with retrieval attention for personalized response generation," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 113–122, 2020.

[18] W. Jiang, Y. Wang, Y. Jiang et al., "Mobile internet mobile agent system dynamic trust model for cloud computing," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 123–136, 2020.

[19] Y. Xu, X. Meng, Y. Li, and X. Xu, "Research on privacy disclosure detection method in social networks based on multi-dimensional deep learning," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 137–155, 2020.

[20] S. Weng, Y. Liu, Y. Shi, B. Ou, C. Zhang, and C. Wang, "A general framework of reversible data hiding with controlled contrast enhancement," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 157–177, 2020.

[21] K. Gu, X. Dong, and L. Wang, "Efficient traceable ring signature scheme without pairings," *Advances in Mathematics of Communications*, vol. 14, no. 2, pp. 207–232, 2020.

[22] T. Wang, D. Zhao, and Y. Feng, "Two-stage multiple kernel learning with multiclass kernel polarization," *Knowledge-Based Systems*, vol. 48, pp. 10–16, 2013.

[23] T. Wang, L. Zhang, and W. Hu, "Bridging deep and multiple kernel learning: a review," *Information Fusion*, vol. 67, pp. 3–13, 2021.

[24] F. Yu, S. Qian, X. Chen et al., "Chaos-based engineering applications with a 6D memristive multistable hyperchaotic system and a 2D SF-simm hyperchaotic map," *Complexity*, vol. 2021, Article ID 6683284, 21 pages, 2021.

[25] N. Liao, Y. Song, S. Su, X. Huang, and H. Ma, "Detection of probe flow anomalies using information entropy and random forest method," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 1, pp. 433–447, 2020.

[26] Z. Tang, X. Zeng, J. Chen, and Z. Guo, "Survey of network traffic analysis based on machine learning," *Network New Media Technology*, vol. 9, no. 5, pp. 1–8, 2020.

[27] J. Hou, P. Fu, Z. Cao, and A. Xu, "Machine learning based DDos detection through NetFlow analysis," in *Proceedings of the MILCOM 2018—2018 IEEE Military Communications Conference (MILCOM)*, pp. 1–6, Los Angeles, CA, USA, October 2018.

[28] K. Flanagan, E. Fallon, A. Awad, and P. Connolly, "Self-configuring NetFlow anomaly detection using cluster density analysis," in *Proceedings of the 2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 421–427, PyeongChang, South Korea, February 2017.

[29] J. Fei, T. Zhang, Y. Ma, and C. Zhou, "A DDoS attack detection method for power grid industrial control system based on BF-DT-CUSUM algorithm," *Telecom Science*, vol. 31, no. s1, pp. 106–112, 2015.

[30] Y. Xu, "Real time traffic classification method of power business based on improved random forest algorithm," *Power System Protection and Control*, vol. 44, no. 24, pp. 82–89, 2016.

[31] J. Du, W. Su, and Q. Peng, "Anomaly detection of power integrated data network based on traffic structure," *Electronic Technology and Software Engineering*, vol. 18, no. 49, pp. 1–12, 2014.

[32] Z. Wu and Y. Dong, "Research on feature selection method of network video traffic classification," *Computer Engineering and Application*, vol. 54, no. 6, pp. 7–13, 2018.

[33] A. Tamer, P. Dilina, and A. Mark, "An evaluation of the performance of restricted Boltzmann machines as a model for anomaly network intrusion detection," *Computer Networks*, vol. 144, pp. 111–119, 2018.

[34] X. Sun, S. Ma, Y. Li et al., "Enhanced echo-state restricted Boltzmann machines for network traffic prediction," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1287–1297, 2020.

[35] Q. P. Nguyen, K. W. Lim, D. M. Divakaran, K. H. Low, and M. C. Chan, "GEE: a gradient-based explainable variational autoencoder for network anomaly detection," in *Proceedings of the 2019 IEEE Conference on Communications and Network Security (CNS)*, pp. 91–99, Washington, DC, USA, June 2019.

[36] R. Dargenio, S. Srikant, E. Hemberg, and U. O'Reilly, "Exploring the use of autoencoders for botnets traffic representation," in *Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW)*, pp. 57–62, San Francisco, CA, USA, May 2018.

[37] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatio-temporal analysis and

CNN deep learning," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1688–1711, 2019.

[38] F. Ertam and E. Avcı, "A new approach for internet traffic classification: GA-WK-ELM," *Measurement*, vol. 95, pp. 135–142, 2017.

[39] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," in *Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics*, Beijing, China, July 2017.

[40] D. Gao and S. Yao, "Analysis and prediction of business flow of power backbone communication transmission network," *Automation and Instrumentation*, vol. 12, pp. 214–217, 2017, in Chinese.

[41] H. Lv, J. Fan, and X. Ma, "Information flow monitoring and prediction analysis platform for electric power communication network," *Science and Technology Innovation*, vol. 16, pp. 79-80, 2020, in Chinese.

[42] J. Lin, S. Fan, Z. Xu et al., "Power grid operation situation awareness evaluation model based on fuzzy analytic hierarchy process and LSTM-attention mechanism," *Electric Power Information and Communication Technology*, vol. 18, no. 4, pp. 58–66, 2020, in Chinese.

[43] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of the 2010 IEEE International Symposium on Circuits and Systems*, pp. 253–256, Paris, France, May 2010.

[44] C. Wang, Z. Wang, Q. Duan et al., "Photovoltaic power prediction based on convolution long short memory hybrid neural network optimized by genetic algorithm," *Acta Physiologica Sinica*, vol. 69, no. 10, pp. 143–149, 2020.

[45] F. Amini and G. Hu, "A two-layer feature selection method using genetic algorithm and elastic net," *Expert Systems with Applications*, vol. 166, Article ID 114072, 2021.

[46] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Systems With Applications*, vol. 164, Article ID 113981, 2021.

[47] M. Nour and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proceedings of the Military Communications and Information Systems Conference (MilCIS)*, November 2015.

[48] J. Sharma, C. Giri, O. C. Granmo, and M. Goodwin, "Multilayer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation," *EURASIP Journal on Information Security*, vol. 15, 2019.

[49] N. Sameera and M. Shashi, "Encoding approach for intrusion detection using PCA and KNN classifier," *Advances in Intelligent Systems and Computing*, vol. 1090, pp. 187–199, 2020.

[50] G. Piraisoody, C. Huang, B. Nandy, and N. Seddigh, "Classification of applications in HTTP tunnels," in *Proceedings of the 2013 IEEE 2nd International Conference on Cloud Networking (CloudNet)*, pp. 67–74, San Francisco, CA, USA, November 2013.