IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 18, 2022, accepted April 5, 2022, date of publication April 11, 2022, date of current version April 21, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3166150*

# Exploration on Automatic Management of GIS Using TL-CNN and IoT

**MIAO WANG[1], ZHENMING ZHANG[1], JIANGUO QIN[2], YANXI CHEN[1], AND BO ZHAI[1]**
[1]School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
[2]School of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot, Inner Mongolia 010051, China

Corresponding author: Jianguo Qin (wmiao@mail.nwpu.edu.cn)

**ABSTRACT** The research aims to effectively monitor the state of complex industrial equipment in real time, diagnose the internal Partial Discharge (PD) pattern of Gas Insulated Metal Enclosed Switchgear (GIS), implement effective health management, and change the traditional model optimization route of increasing training time in exchange for performance improvement. This paper studies the complex equipment-oriented Health Management System (HMS) based on Internet of Things (IoT) technology and Transfer Learning. Firstly, the principles of Transfer Learning and Deep Learning (DL) technology are introduced. Secondly, the requirements of GIS internal status recognition and management are studied. Furthermore, a GIS-oriented HMS based on Transfer Learning-optimized Convolutional Neural Networks (CNN) is proposed, and the training dataset is constructed. Finally, the proposed model is tested. The results show that the complex equipment-oriented HMS based on IoT technology, CNN, and Transfer Learning can detect the internal status of GIS in real time. Compared with the traditional DL algorithm and expert system, the proposed model has a shorter training time of only 16min, faster convergence speed, high testing recognition rate, and over 96% recognition rate. Compared with other mainstream algorithms, it has higher identification, the storage parameter volume is 408, and the storage space is 12.8MB. Moreover, the proposed Transfer Learning-optimized CNN model can accurately detect the status of GIS, identify abnormal statuses, and help prolong the service life of GIS. The proposed complex equipment-oriented HMS contributes to the intelligent manufacturing industry and provides a new direction for applying emerging Computer Technology in the intelligent industry.

**INDEX TERMS** Internet of Things, convolutional neural network, GIS, status recognition, equipment health management.

## I. INTRODUCTION

Chinese continuous industrialization has gradually entered the era of ''industry 4.0'', which helps to popularize industrial reform, thus affecting all walks of life through advanced technologies, such as Information Technology (IT). As a result, the manufacturing industry set off a new round of revolution towards intelligent manufacturing, and intelligent industrial equipment sees extensive application [1], [2]. In particular, Gas-Insulated Metal Closed Switchgear (GIS), [3] because of its unique advantages, such as small footprint, high reliability, and little electromagnetic contamination is widely used in modern power systems. However, due to the limited production, transmission, and installation conditions, GIS will have potential insulation defects after long-term operation and maintenance, resulting in various Partial Discharges (PDs) [4] that severely harm GIS. The Internet of Energy (IoE), aiming at comprehensive perception, interconnection, and intelligent decision-making, provides new GIS online monitoring and fault diagnosis opportunities. At the same time, it also poses new challenges for real-time and rapid diagnosis. As one of the critical links in IoE, online detection and fault diagnosis is the premise and foundation of comprehensive perception of equipment information, system interconnection, and multi-directional service application. Although the mechanism of PD pattern in GIS is complex, the IoE enables the preservation of equipment information to solve the representative and comprehensive problems of defect and fault samples. It is of great significance to realize the intelligent state perception and evaluation of GIS and ensure the safe and stable power

The associate editor coordinating the review of this manuscript and approving it for publication was Porfirio Tramontana.

system operation. This paper studies the complex industrial equipment-oriented Health Monitoring System (HMS) based on IoT technology and Transfer Learning.

Many domestic and international researchers have studied the automatic HMS for complex industrial equipment. For example, Zelyakovskiy *et al.* [5] proposed a technical diagnostic Intelligent Control System (ICS) for complex equipment and a flexible production system. The proposed system could automatically select input and output parameters using the Neural Network (NN) training process. Perhaps, the variation of active neuron numbers would not impact the diagnostic results because the selection and activation of redundant neurons could help avoid network failure. Qiang *et al.* [6] put forward an Inertial Neural Network (INN) of fuzzy memory reactor. The passivity criterion was given based on the eigenfunction, Linear Matrix Inequality (LMI), and the calculation of time scale. Two memristor and fuzzy correlation control protocols were designed against passivation. Ding *et al.* [7] invented a short-term load forecasting method in distributed cloud-edge environment based on the Long Short-Term Memory (LSTM) network to provide more accurate results for intelligent forecasting of industrial power load. The LSTM network could perform the prediction task, and the whole system was extended to the cloud edge platform to realize parallel neural computing. Narwariya *et al.* [8] suggested a method to capture and visualize the structure of complex equipment through graphs and used Graph Neural Network (GNN) to model multi-sensor Time Series Data (TSD). The proposed method was applied in residual life estimation to evaluate the advantages of merging graph structure on the publicly available turbofan engine benchmark dataset. The resulting network could focus on the possible failed module (node) through a simple Attention Mechanism (AM). Windau and Itti [9] constructed the Inertial Machine Monitoring System (IMMS) for non-intrusive upgrading of existing machines to detect and classify equipment faults or degraded status. They also provided strategies to optimize sensors' number, placement, and efficiency. Consequently, Support Vector Machines (SVM) and NN could detect and classify regular operation along with ten kinds of equipment abnormal behaviors (loose tape and machine component faults). Chen *et al.* [10] applied Deep Learning (DL) in complex equipment HMS and realized a Backpropagation Neural Network (BPNN)-based DL model. To sum up, the previous research on complex industrial equipment-oriented HMS mainly focuses on the simple application of Deep Neural Network (DNN) without much improvement. In simpler terms, they improved model performance mainly by increasing training time. Therefore, this paper innovatively optimizes the Convolutional Neural Network (CNN) model by Transfer Learning (TL). Transfer Learning has a higher starting point. The original performance of the untrained model is better before fine-tuning. Such that, in the training process, the model performance will be improved faster. As a result, the trained model will converge better. The proposed Transfer Learning-optimized CNN model realizes the intelligent monitoring and recognition of equipment PD, with a high recognition rate and strong reliability. The research content provides a basis for the refined management and automatic monitoring based on new Computer Technology and provides a new perspective for future research.

This paper studies intelligent health monitoring for complex industrial equipment. Firstly, the principle of Transfer Learning and CNN technology is introduced. Secondly, the GIS status recognition requirements in complex, intelligent equipment are studied. Thirdly, the Transfer Learning-optimized CNN model is implemented, based upon which the GIS-oriented HMS is implemented; finally, the proposed model is tested. The research results are of great significance to developing China's intelligent manufacturing industry and help ensure the safety of industrial equipment and relevant personnel.

## II. DESIGN OF GIS-ORIENTED HEALTH MANAGEMENT SYSTEM (HMS)

### A. GAS INSULATED METAL ENCLOSED SWITCHGEAR (GIS)

In GIS, at least part of the metal closed switchgear and control equipment are insulated using superior dielectric gas-Sulfur hexafluoride (SF6) compressed by a higher than atmospheric pressure [11] GIS comprises the circuit breaker, disconnector, grounding switch, transformer, lightning arrester, bus, connector, and outgoing terminal. These equipment or components are all enclosed in a metal grounded shell, in which there is a certain pressure of insulating gas, so they are also called fully enclosed combined circuit breakers [12], [13]. GIS uses gas higher than atmospheric pressure as an insulating medium for at least a part of metal-enclosed switchgear and control equipment. GIS comprises circuit breaker, disconnector, grounding switch, transformer, lightning arrester, bus, connector, and outgoing terminal. All these equipment or components are enclosed in a metal grounded shell and filled with insulating gas with a certain pressure, so it is also called a fully enclosed combined circuit breaker. GIS has been widely used worldwide since it was implemented in the 1960s. GIS has seen broad applications not only in high voltage but also in the field of Ultra-High Voltage (UHV). Compared with conventional open substation equipment, GIS has the advantages of compact structure, small footprint, high reliability, flexible configuration, convenient installation, strong safety, strong environmental adaptability, and minor maintenance workload, and the maintenance interval of its main components is not less than 20 years. Moreover, GIS is high-voltage electrical equipment with high operation reliability, less maintenance workload, and a long maintenance cycle. Its failure rate is only 20% $\sim$ 40% of conventional equipment, but GIS also has inherent disadvantages. Due to the influence of gas leakage, external moisture infiltration, conductive impurities, insulator aging, and other factors, GIS internal flashover fault may be caused. The fully sealed structure of GIS makes the fault location and maintenance more complicated. The average power outage maintenance time after the

accident is longer than that of conventional equipment, and its power outage range is more extensive, which often involves non-fault components.

### B. INTERNET OF THINGS (IOT) TECHNOLOGY

The IoT [14] refers to the real-time collection of miscellaneous information with various devices and technologies, such as information sensors, Radio Frequency Identification (RFID) [15], Global Positioning System (GPS) [16], infrared sensors, and laser scanners. IoT uses network access to connect things and people and intelligently perceive, identify, and manage things. So far, IoT techniques cover numerous cutting-edge technologies, including RFID, sensors, Cloud Computing (CC), and network communication. IoT can be regarded as a kind of physical object-oriented intelligent manufacturing technology, the further expansion of Internet technology, and an emerging Information Service Architecture (ISA) based on Internet and RFID communication technology. IoT aims to empower IT infrastructures to provide safe and reliable ''goods'' information over the Internet and create an intelligent environment for smoother Information Exchange (IEX) within the Supply Chain (SC). Significantly, in recent years, the IoT has been matured and is reshaping people's lifestyles and benefiting all social aspects. Nonetheless, it has also brought a series of imminent technical ethics, security, and legal problems. IoT relies on user identification to be fully functional; this, however, has also become an inherent leak that is difficult to resolve merely through technical means; some simple expedients include informing users of specific privacy terms in advance. The IoT transmits intelligent services over the network from the users' standpoints. Thus, it can track users' health status over the long or short term, offer navigation information and road conditions, detect and analyze air or water quality, and digitally link residential areas and workplaces [17], [18]. This paper utilizes IoT sensors and cameras to collect the internal data of GIS. Fig. 1 depicts the basic architecture of the IoT.

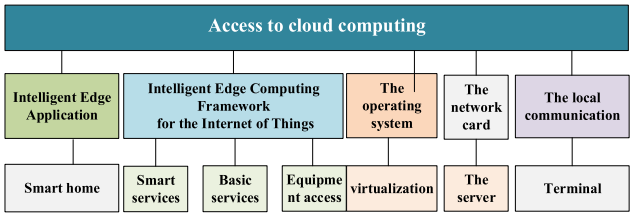


**FIGURE 1.** Basic architecture of IoT.

The following are the functional components of the IoT:

Equipment –It includes existing standard equipment, such as intelligent instruments or vehicles, integrated into the product design. It also includes new IoT-based devices, such as pet trackers. Such devices must have sensors, communication functions, and other elements (e.g., power supply).

Devices connected with sensors or actuators – Sensors can capture data (e.g., temperature) from the environment. The actuator responds to the command and changes the equipment status (for example, adjusting the thermostat's temperature). The instructions of the actuator can come from sensors on the same device or from other sources (for example, when the homeowner comes home, he can activate the thermostat through his mobile phone). The device can have two functions: sensor and actuator.

Communication hardware enables devices to connect to the network to send data from sensors to back-end systems. It includes hardware for wireless connection through Bluetooth, Wi-Fi, ZigBee, LoRa, and cellular network), various proprietary technologies, or wireless connection through the fixed network. Some devices will have hardware that connects to multiple types of networks.

The connection network (a cellular network, fixed network, or satellite network) can transfer the data from the sensor to the user's back-end system through the Internet or private network. Additionally, various applications can provide added value to end-users.

The server software includes a server for collecting and analyzing data from sensors and other sources (for example, weather forecast data). These server-side systems can be found in public or private cloud or local hardware. For a straightforward system, the server software can be a standard computer.

Software platforms such as device management, security, and data analysis can ensure the normal operation of IoT devices. Such a platform also includes data analysis software for analyzing data and improving business processes and a database for storing data.

The application software also includes services, such as billing and customer support.

The IoT value chain also includes system integrators or developers who design, build, and manage IoT services. Physical devices usually require installation and maintenance.

### C. TRANSFER LEARNING

Transfer Learning is a research topic in Machine Learning (ML), which refers to influencing a new learning process through a given related learning process, or the influence of acquired experience on the other activities to be completed. Generally, Transfer Learning learns new knowledge through existing knowledge. The core is to excavate the connection between existing knowledge and new knowledge to draw inferences from one instance [19]. Generally, it is too costly to learn from scratch directly on the target domain, so there is a need to resort to the existing relevant knowledge to help learn new knowledge faster. Transferring process is commonly found in humans learning various knowledge, skills, and social norms. Transfer Learning focuses on storing the solution model of existing problems and using it on other different but related problems. It transfers the pre-trained model parameters to retrain the new model. Considering that most data or tasks are relevant, new models can partly share the learned parameters (also known as the knowledge learned by the model) with the old ones through Transfer

Learning to speed up and optimize the learning efficiency while avoiding repetitive works [20], [21]. Transfer Learning mainly includes three dimensions: A. Study what knowledge can be used for Transfer Learning in different domains or tasks and what common knowledge can be transferred between different domains. B. Study what Transfer Learning algorithms can be used against specific transference targets. That is to design an appropriate algorithm to extract and transfer common knowledge. C. Study what circumstances are suitable for Transfer Learning and whether transference skills are suitable for specific applications, which involves the problem of negative transference. This paper uses the pre-trained data images of the Mstar database to construct the dataset. Fig. 2 outlines a rough classification of Transfer Learning.

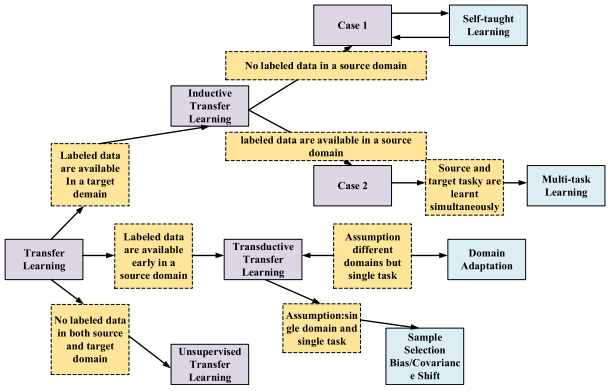


**FIGURE 2.** Rough classification diagram of transfer learning.

Mstar database adopts the measured ground still target image data published by the Plan supported by the United States (US) Defense Advanced Research Program. Most domestic and international research on image target recognition is based on the Mstar. The sensor collecting the data is a high-resolution spotlight Synthetic Aperture Radar (SAR) with a resolution of 0.3m $\times$ 0.3m. Working in X-band, the sensor employs the HH polarization mode. The collected images are preprocessed, containing various targets' slice images by 128 $\times$ 128. Most of the data are SAR slice images of stationary vehicles, including target images obtained by various vehicles in multiple azimuth angles. The dataset includes a training set and testing set recommended by the Plan. The training set is the target image data obtained when the radar working pitch angle is 17. By comparison, the testing set is the target image data obtained when the radar working pitch angle is. The data set also includes three categories. Targets of various categories also have different models. Targets of the same type but different models have some differences in equipment, but the overall scattering characteristics are not much different. Mstar mixed target data also contains slice images of a group of other military targets. These targets are the imaging pictures of each target when the radar works at various pitch angles. Therefore, the Mstar data set is applied to train the model in the present work. Transfer Learning can adjust the CNN model trained on a dataset to generalize to another dataset. Ideally, with the increase of the number of layers and complexity of the model, the error rate of the CNN model can continue to decrease. However, training a complex CNN requires a lot of labeled data and is extremely time-consuming. To this end, this section introduces Transfer Learning on the following principle. In the trained model, the output of the bottleneck layer can be well distinguished from 1,000 kinds of images through a single-layer fully connected layer neural network. Thus, the node vector output by the bottleneck layer can be used as a more expressive feature vector of any image. Therefore, the trained neural network can be directly used to extract the image features on the new dataset. The extracted feature vector is used to train a new single-layer fully connected neural network to address the new classification problem. Generally, given sufficient training samples, the effect of Transfer Learning is not as good as complete retraining. However, the training time and training samples required by Transfer Learning are far less than a learn-from-scratch process.

### D. DEEP NEURAL NETWORKS

#### 1) DEEP LEARNING

2006 has witnessed a new research upsurge in ML, namely DL; since then, academia has studied DL and gradually applied it in different industries. In 2012, Stanford University takes the lead in using 16,000 Central Processing Unit (CPU) parallel computing platforms to build a training model: Deep Neural Networks (DNN) [22]. This technology has made a great breakthrough in Speech Recognition (SR) and Image Recognition (IR). In 2016, Alpha Dog, an artificial go software developed based on DL, defeats Li Shishi, the world's top go master. After that, well-known high-tech companies worldwide have launched tremendous investment and research institutes for DL and efforts to cultivate Research and Development (R & D) personnel [23], [24]. ML technology studies how computers simulate or realize the learning behavior of animals to learn new knowledge or skills, rewrite the existing data structure, and then improve the program performance. From a statistical perspective, DL predicts the data distribution, specifically, creates a model through data learning and then predicts new data through this model, requiring a similar distribution between the test and training data [25]. Its basic feature is to imitate the mode of information transmitting and processing between brain neurons, which are well illustrated in applications, such as computer vision and Natural Language Processing (NLP). Therefore, DL has a strong relation with NN in ML, and NN structures are also the main algorithm or means to realize DL; more concretely, DL can be regarded as an enhanced NN algorithm [26].

#### 2) ARTIFICIAL NEURON

As a mathematical function, artificial neurons imitate biological neurons' basic operational mechanism; thus, they have

specific characteristics of biological neurons [27]. Its structure is shown in Fig. 3.

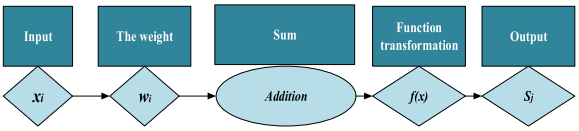


**FIGURE 3.** Artificial neuron structure.

The artificial neuron receives the front neuron's weighted signal. This neuron will be activated (activation status) under all weight statuses' joint action, expressed by Eq. (1).

$$f(x) = \sum_{i=1}^{n} x_i w_i \tag{1}$$

In (1), $f(x)$, $x_i$, and $w_i$ represent the ultimate output status, the input signal, and the weight of the input. There are $n$ groups in total.

Given an input, a neuron output a specific value; each neuron corresponds to a threshold. If the sum of the inputs exceeds the threshold, the neuron is activated; otherwise, it will be inhibited. The Transfer Functions (TF) of artificial neurons are as follows [28], [29].

a. The expression of the Linear Function (LF) is shown in Eq. (2):

$$f(x) = kx \tag{2}$$

b. The expression of slope function reads:

$$f(x) = \alpha(x \geq \theta) \tag{3}$$
$$f(x) = kx(-\theta < x < \theta) \tag{4}$$
$$f(x) = -\alpha(x \leq \theta) \tag{5}$$

The calculation of the transition function reads:

$$f(x) = \alpha(x \geq \theta) \tag{6}$$
$$f(x) = \beta(x \leq \theta) \tag{7}$$

The Sigmoid function $f(x)$ reads:

$$f(x) = a + \frac{b}{1 + e^{-dx}} \tag{8}$$

$k$, $a$, $d$, $\alpha$, and $\beta$ are used to set the required parameters. The TF needs to be selected in combination with the specific application range. The LF will amplify the output signal, the nonlinear slope function can minimize the impact of network performance degradation, and the S-type function is used in the hidden layer.

### 3) CONVOLUTIONAL NEURAL NETWORKS

DL structure is divided into CNN and Deep Belief Networks (DBN) [30], which are constructed through simulation of human neurons; each neuron receives information, processes, and transmits it to all adjacent neurons. Its structure is sketched in Fig. 4.

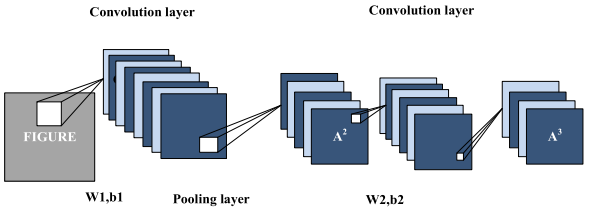


**FIGURE 4.** Forward propagation processing method of CNN.

In Fig. 4, the leftmost side is the data input layer. It processes the data, such as de-averaging (centralizing all dimensions of the input data to 0 to prevent data-excessive deviations from affecting the training effect), normalization (normalizing all data to the same range), and Principal Component Analysis (PCA)/whitening. CNN only does the "de-averaging" step for the training set. In the middle is CONV: convolution calculation layer, linear product summation. RELU: excitation layer. Rectified Linear Unit (ReLU) is a kind of Activation Function (AF). POOL: pooling layer takes the regional average or maximum. On the far right is FC: fully connected layer.

The operation of CNN is divided into two stages:

a. Forward propagation – The dataset sample $i$ is represented by $(x^i, y^i)$, $x^i$ denotes the output vector, and $y^i$ stands for the sample label. The $l$-th layer is set as the convolution layer. $x^{l-1}$ and $x^l$ refer to the input and output of the $l$-th layer, respectively [31], [32]. The calculation of the output of the $l$-th layer is expressed as in Eq. (9).

$$x^l = f(x^{l-1} * w^l + b^l) \tag{9}$$

In Eq. (9), $w^l$ represents the convolution kernel of the $l-1$-th layer and the $l$-th feature maps, $*$ denotes the convolution operation, $b^l$ indicates the bias, and $f$ stands for the nonlinear AF. The lower sampling layer uses the principle of correlation between local images to reduce calculation.

b. Backward propagation – it is commonly used for adjusting network structure parameters. It updates the parameters by backward propagation of the residuals of network nodes [33], [34]. Fig. 5 demonstrates the process of backward propagation.

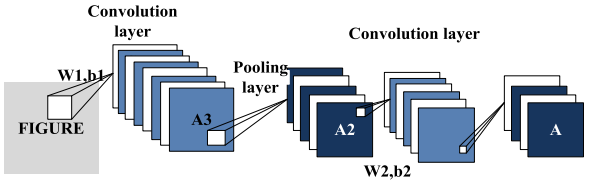


**FIGURE 5.** Backward propagation process.

The specific steps are as follows.

a. Parameter initialization – each layer's number of input ($W$) and output feature maps are initialized by Eq. (10).

$$\Delta W = 0, \Delta b = 0 \quad W \in [-\sqrt{\frac{6}{fan_{in}}} + \sqrt{\frac{6}{fan_{in} + fan_{out}}}], \quad b = 0(10) \tag{10}$$

In Eq. (10), $fan_{in}$ and $fan_{out}$ are the number of output and input feature maps of each layer.

b. The output of each layer is calculated by Eqs. (11) and (12).

$$W \in [-\sqrt{\frac{6}{fan_{in}}} + \sqrt{\frac{6}{fan_{in} + fan_{out}}}], \quad b = 0 \tag{11}$$

$$z^l = x^{l-1} * w^l + b^l \tag{12}$$

In Eq. (12), $l$ represents the number of layers; $x$ denotes the feature map; $w$ refers to the convolution kernel; $f$ stands for the front of the current layer, and $f$ is the AF.

c. Suppose the output results are divided into $m$ categories, that is, the final output is $m$ dimensional data; In that case, the Mean Square Error (MSE) between the output and the sample label can be used to calculate the cost function $J(w, b, x, y)$ of the $n$-th sample, as manifested in Eq. (13).

$$a^l = f(z^l) \tag{13}$$

$t_i^n$ and $y_i^n$ represent the expected and actual output of the $i$-th neuron in the $n$-th sample, respectively. Then, the derivative of the single-sample cost function to the parameter is obtained by the $l$-th layer sensitivity, as counted by Eq. (14).

$$J(w, b; x, y) = \frac{1}{2}\sum_{i=1}^{m}(t_i^n - y_i^n)^2 = \frac{1}{2}\left\|t^n - y^n\right\|_2^2 \tag{14}$$

$n_i$ represents the output layer, and the sensitivity of the output layer can be calculated by Eqs. (15) and (16).

$$\delta^l = \frac{\partial J}{\partial z^l} \tag{15}$$

$l = n_i - 1, n_i - 2, \cdots, 2)$; the sensitivity of each layer is expressed in Eq. (16).

$$\delta^{(ni)} = -(y - a^{(ni)}) \cdot f'(z^{(ni)}) \tag{16}$$

d. After the backpropagation of the cost function, the Gradient Descent Method (GDM) can obtain the partial derivative of the cost function to the parameters in the network, update the parameters, and finally, obtain the minimum of the cost function. The partial derivative of the cost function $J(w, b, x, y)$ to $w^{(t)}$ and $b^{(t)}$ is calculated by the Eqs. (17)-(20).

$$\nabla w^{(t)} J(w, b, x, y) = \frac{\partial J(w, b, x, y)}{\partial z_i^{l+1}} * \frac{\partial z_i^{l+1}}{\partial w_{ij}^l} = \delta^{(l+1)}(a^{(j)})^T \tag{17}$$

e. The weight parameters $w^{(l)}$ and $b^{(l)}$ are updated by Eqs. (19) and (20).

$$w^{(l)} = w^{(l)} - \alpha[(\frac{1}{m}\Delta w^{(l)}] \tag{18}$$

$$\nabla b^{(t)} J(w, b, x, y) = \frac{\partial J(w, b, x, y)}{\partial z_i^{l+1}} * \frac{\partial z_i^{l+1}}{\partial b_{ij}^l} = \delta^{(l+1)}(a^{(j)})^T \tag{19}$$

$$b^{(l)} = b^{(l)} - \alpha[(\frac{1}{m}b^{(l)}] \tag{20}$$

In Eqs. (19)-(20) $\alpha$ represents the Learning Rate (LR).

### E. CONVOLUTIONAL NEURAL NETWORKS PRETRAINING AND EXPERIMENTAL TEST DESIGN

This section employs a five-layer CNN model. The first convolution layer contains eight $5 * 5$ filters, the convolution kernel of the second sampling layer is $2 \times 2$, the third convolution layer contains sixteen $4 * 4$ filters, and the last layer is the fully connected layer. CNN inputs large-scale samples using ReLU as AF and adjusts each layer's weight and error through forward-propagation and backpropagation based on Eq. (19). The training parameters are set with 50,000 iterations, four data batch, initial LR = 0.0001, Adam optimizer, and a 0.1/20,000 iterations LR attenuation.

Specifically, the SAR data images pre-trained on the Mstar database are used for the experiment, with a resolution of $0.3\text{m} * 0.3\text{m}$ and a target image size of $128 * 128$ pixels. BRDM2, BTR60, D7, 2S1, T62, T72, ZIL131, and ZSU1234 are used as the experimental data, and each image is manually labeled. The experimental data are divided into three parts. The first four types of datasets contain 1,250 training samples and 1,120 testing samples; the last four types of datasets include 1,270 training samples and 1,190 testing samples; the actual status map set of GIS consists of 300 training samples and 100 testing samples. The PD pattern recognition of GIS is tested, and the recognition results of four kinds of discharge patterns are compared.

Then, the Coefficient of Determination (CD) (between 0-1) is used to determine the Goodness Of Fit (GOF) of the model. The larger the CD is, the higher the model GOF is, and the lower the CD is, the worse the GOF is. In particular, the CD, also known as the determinable coefficient, is often used to represent the relationship between multiple random variables. It is a statistical index used to measure the reliability of the regression model to explain the change of dependent variables. Eqs. (21) and (22) are the functional relationship expressions.

$$\bar{\bar{y}} - \bar{y} = a + b(x - \bar{x}) \tag{21}$$

$$R^2 = \frac{\sum(\bar{\bar{y}} - \bar{y})^2}{\sum(y - \bar{y})^2} \tag{22}$$

In Eqs. (21) and (22), $\bar{\bar{y}}$ represents the regression fitting value of the independent variable $y$. $\bar{y}$ is the average of the independent variable $y$. $a$ and $b$ are the calculated constant.

The transfer Learning process is presented in Figure 6. First, it transfers the models' knowledge that has achieved excellent results under other target tasks and then uses the GIS PD signal for the model's cascaded fine-tuning. Afterward, the model effect is judged. The PD time domain map is transformed into a single channel through binary processing in the data preprocessing. The input image is compressed into the size required by the corresponding CNN model. The data enhancement method randomly selects 20% of the training data to improve the model generalization. Subsequently, input data are normalized by the Z-score method to compare data indexes from different sources. After that, the model parameters trained under the known task are transferred to solve the

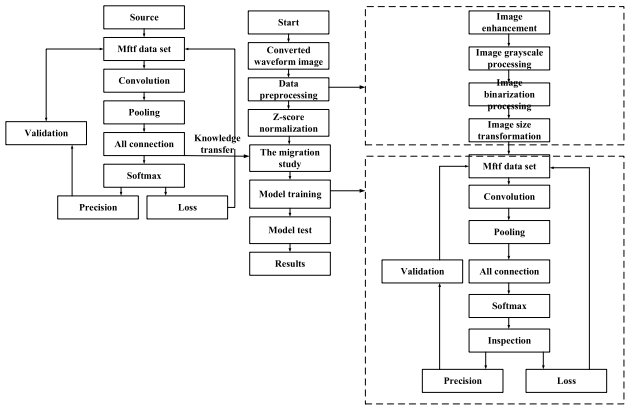


**FIGURE 6.** GIS PD pattern recognition and classification process.

problem that some DL models cannot be trained due to the gradient disappearance when they are trained from scratch. This paper transfers the Mftf dataset. The backpropagation algorithm and Stochastic Gradient Descent (SDG) algorithm are used to fine-tune the model. The normalization technology is introduced into the pooling layer, and the dropout method is introduced into the fully connected layer. Finally, the Transfer Learning-optimized CNN model is tested with the remaining map to verify its generalization, fault recognition accuracy, and testing time.

Insulation defects will occur in the manufacturing, transportation, installation, and operation of GIS, including metal particle defects, metal tip defects, suspended electrode defects, insulator air gap defects, and insulator surface dirt. This section designs an experiment to simulate four types of typical defects by changing the position of the Ultra High Frequency (UHF) PD sensor to replace the transformation of defect position. The test voltage is set between $-5*10^{-5}$V-$5*10^{-5}$V. Altogether, 1,000 groups of data are obtained for each type of defect, and then the GIS PD pattern recognition and classification dataset is constructed. Finally, the diameter of the central conductor is 120mm, the diameter of the cylinder is 400mm, the thickness of the cylinder wall is 10mm, and the cylinder length is 2.2m. The metal tip defect is explicitly simulated: a metal needle with a length of 30 mm is installed on the high-voltage conductor, and a source with 3GHz and 30dB attenuation is applied. Then, to ensure consistency with the actual situation, both ends of the model are truncated by the Finite Difference Time Domain Method (FDTD). Seven Perfect Matched Layers (PML) are set to match the wave impedance of its adjacent medium. Additionally, both the high-voltage conductor and the cylinder are ideal conductors, ignoring the surface loss. The relative dielectric constant filled in GIS is 1.00205, the density under the pressure of 0.4MPa is 23.7273 kg/ $m^3$, the relative permeability is 1, and the conductivity is 1.1015 × 10-5 S/ m. calculation.

The maximum frequency is 3GHz, and the unit size is set to 10 mm×10 mm×10 mm, the simulation time is 300ns, and the time step is set to 9.616 74 × $10^{-6}\mu$s. The simulation conditions of the other three defects are the same.

### F. NETWORKING MODULE

This section selects the Alibaba cloud server to connect the equipment data acquisition system to realize the cloud platform's software installation and service deployment. Five servers are chosen to build the cloud-based data collection system and deploy the message queue service and web service to the same server based on the cost and performance consideration. Because these two services have high network transmission and bandwidth requirements, selecting a high-bandwidth supportive server is needed. The server provides Message Queuing Telemetry Transport (MQTT) service and inputs data to the cloud platform. At the same time, the server provides background services and can visualize system functions to customers through web pages. Finally, this paper selects three servers to form a cluster to store all historical data.

i The local data transmission network can be divided into local hardware and gateway. In implementing the local hardware transmission network, there is a need to simultaneously interpret the transmission mode according to the output of different sensors, including hall ring, magnetic suction temperature sensor, laser displacement sensor, and proximity sensor, which all output analog signals through Serial Port (SPORT). Therefore, the output of these four sensors first needs to be connected to the data acquisition card to complete Digital to Analog Conversion (DAC). Then the data acquisition card uses the Remote Terminal Unit (RTU) protocol to output SPORT signals to the SPORT server. Then, the analog signal is converted digitally and output over the Transmission Control Protocol (TCP) through network cable and transmitted to the gateway. In particular, the triaxial accelerometer comes with a data analysis function itself and output TCP-compatible data; it can be directly connected to the SPORT server and output uniformly through the SPORT server. To sum up, the SPORT server outputs five types of sensor data, distinguishable by the gateway through a network address and port number. The programmable Logic Controller (PLC) register stores the signal information representing the device's start, standby, operation, and stop. The device can read the data in the corresponding address through the User Datagram Protocol (UDP), and its output can also be directly connected to the gateway. Further, data processed by the local gateway module will be transmitted to the cloud through the Fourth Generation (4G) wireless network card over the Message Queue Telemetry Transport (MQTT) protocol. The implementation of the local gateway network mainly involves the interaction between the front-end page, background system, and database. This paper uses a single-page web display application in the front-end page design, which will be introduced in detail in the subsequent software design section. Firstly, the single-page web application directly subscribes to the sensor data through the MQTT technology in real time. Secondly, the initialization configuration

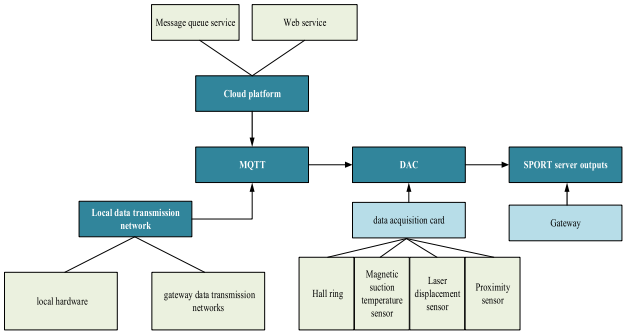


**FIGURE 7.** Network module flow chart.

information function, query function, and report printing function interact with the background system through Hypertext Transfer Protocol (HTTP). Finally, The background system uses a database driver and Mongo database for data transmission. The cloud data transmission network can be realized according to different data types. According to the output of the local gateway, the cloud will receive two types of data: sensor data and PLC data. There are three transmission paths available for sensor data. (1) Offline storage through the cluster. These historical data are used to train the fault diagnosis algorithm model. The final real-time diagnosis results are saved in the database. The terminal regularly sends requests to obtain the diagnosis results through Hypertext Transfer Protocol (HTTP). (2) Timestamped TSD in the spatiotemporal database. Timestamped TSD facilitates the calculation of statistical indexes. The terminal also regularly requests to calculate statistical values through HTTP protocol and then obtain the returned results. (3) Terminal-subscribed data through the MQTT technology. These data are displayed on the front-end page. By comparison, PLC data can only have one transmission path. After the cloud platform subscribes to PLC data, the script will analyze the PLC data according to the specific rules and save various status information to the database. Then, the terminal regularly sends requests to obtain PLC data through HTTP to display the device status (Start or Standby) on the front page.

### G. PARTIAL DISCHARGE DATA ACQUISITION

The online detection methods of GIS PD can be divided into the acoustic method, chemical method, pulse current method, and UHF method. UHF method has become the primary method in GIS PD detection technology because of its strong anti-interference ability, high sensitivity, good real-time performance, and fault location. Because GIS's installation and operation place are very complex, interference signals are relatively strong and miscellaneous. According to the actual production experience, most interferences are concentrated in the lower frequency. When PD occurs in GIS, the signal spectrum is vast, and the discharge process can stimulate UHF electromagnetic wave signals from several Hertz to thousands of MHz. Therefore, the UHF sensor is adopted to collect PD signals in the frequency band above 0.5GHz to avoid substantial amounts of electromagnetic interference. The system installs multi-channel UHF sensors at the GIS utility hole or on the basin insulator to obtain the PD signal and acquire the PD signals.

Since UHF sensor-collected signals are high frequency, high-speed Analog-to-Digital Converter (ADC) and signal processor must be used, which is exceptionally costly and inefficient. Therefore, the signal must go through Radio Frequency (RF) filters. After pre-amplification, mixing, and detection, the signal can be output with 1∼10 MHz bandwidth before entering the data processing module. The data processing module mainly includes the ADC and Field Programmable Gate Array (FPGA). Remarkably, the designed system utilizes a 14 bit and 50 Msps ADC module to realize the synchronous sampling of PD signals.

Through digital processing technologies, such as wavelet analysis of FPGA, the local discharge signal is extracted under intense noise and uniformly transmitted to Digital Signal Processing (DSP) module. Then, the signals are sent to the terminal analysis unit after centralized processing. Finally, the terminal analysis unit completes the analysis, identification, and judgment of the discharge characteristic spectrum of PD information.

The ADC and DSP module cannot directly process UHF sensor-collected signals. Thus, the signal must be preprocessed. First, the signal passes through the first-order Low Pass Filter (LPF), the wave below 1.5 GHz is retained for RF amplification, and a 0.5∼1.5GHz frequency sweep is performed on the signal. Then, the three-stage mixing circuit can effectively shield the frequency band with large interference and improve the system detection sensitivity. Afterward, the down-conversion technology is adopted to adjust the Intermediate Frequency (IF) output bandwidth to 10∼20MHz and convert the UHF signal into an ADC and DSP addressable frequency band. The IF signal is converted into a digital signal through ADC, enters the data processing module, and separates the PD signal characteristics.

Analyzing the original waveform of GIS finds that there may be substantial electromagnetic interference in the site, such as mobile communication or subway interference. These interference and PD signals enter the data processing module through the mixing circuit, inconveniencing the PD diagnosis. Meanwhile, these interferences are generally concentrated in a narrow band. They can be significantly weakened by classifying and summarizing the interference signals and then performing digital band-pass filtering on their respective narrow bands.

After the concentrated narrow-band interference is weakened, the white noise can be eliminated by wavelet processing, and the local discharge waveform can be extracted. The distribution law of PD signal and white noise in the depth of wavelet decomposition is not consistent; the low-scale high-frequency coefficient contains most of the white noise, while the high-scale coefficient concentrates the PD signal. Therefore, the PD signal should be concentrated in the high-scale coefficients as much as possible after wavelet

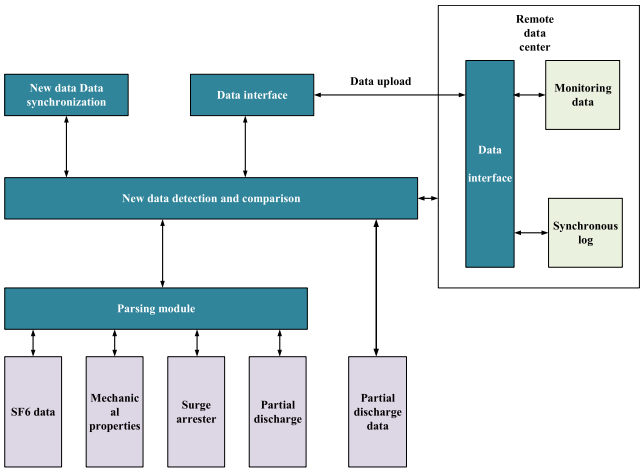


**FIGURE 8.** Data synchronization function structure.

decomposition, and the distribution of white noise cannot be changed. Figure 8 describes the structure of PD data collection and data synchronization function.

Table 1 lists the specific data classification.

**TABLE 1.** PD data collection category.

| Serial number | Data type | Explanation |
|---|---|---|
| 1 | User data management | User and authorization data |
| 2 | Parameters | Number, address, and basic parameters |
| 3 | Real-time monitoring data | Real-time data, such as mechanical characteristics, lightning arrester, PD, and contact video monitoring |
| 4 | Historical monitoring data | Historical data of various monitoring is used to generate monitoring curves |
| 5 | Algorithm analysis data | Data storage of intermediate result of algorithm analysis |
| 6 | Analysis result data | Analysis results of monitoring data |
| 7 | Data synchronization log | Synchronization log of each monitoring data |
| 8 | Expert database feature data | Characteristic data of various monitoring data and judgment of user analysis results |
| 9 | Maintenance method database | Provide equipment maintenance suggestions according to the analysis results |

The model fits the training data well but poorly for other data in the actual training. Such a fitting degree can evaluate the model generalization in model selection. Notably, k-fold cross-validation can be utilized for model correction: part of the original data is used as the training set and the remainings as the verification set. Firstly, the training set is used to train the model, and then the verification set is used to test the model generalization. Additionally, labeled data are less available in reality, and k-fold cross-validation can reuse some original data for a classification or regression problem. K-fold cross-validation takes 1/k of the training set as the testing set, trains each model k times, tests k times, and averages the error rate of k times. Finally, it selects the model $M_i$ with the lowest average rate.

a. Divide all training sets S into k disjoint subsets. Suppose the number of training samples in S is m, then each subset has m/k training samples.

b. Take one $M_i$ from the model set M each time and select k-1 training subsets. After training $M_i$ with this k-1 subsets, the hypothesis function $h_i$ is obtained. Finally, the remaining $S_j$ is used for testing, and empirical errors are obtained.

c. Since one $S_j$ is retained every time, k empirical errors will be obtained. The empirical error of $M_{\hat{i}}$ is the average of k empirical errors.

d. Select the $M_i$ with the lowest average empirical error rate, and then use all S to perform another training to get the final result.

Data are classified as follows:

(1) True Positive (*TP*) – The model recognizes the positive samples as positive.

(2) False Negative (*FN*) – The model recognizes the positive samples as negative.

(3) False Positive (*FP*) – The model recognizes the negative samples as positive.

(4) True Negative (*TN*) – The model recognizes the negative samples as negative.

The matrix can be used for easily understandable binary classification problems. At the same time, the confusion matrix can be easily applied to problems with three or more class values by adding rows and columns. The classification index is obtained from the confusion matrix, from which more advanced classification indexes can be obtained. The indexes include *Accuracy*, *Precision*, *Recall*, *Specificity*, and sensitivity. Then, the binary classification can segment the samples according to their real categories and the predicted categories by ML. Thus, total samples $= TP+FN+FP+TN$. *Accuracy* is the most commonly used classification performance index to express the model accuracy. It is defined as the number of correct model recognition/the total number of samples, as calculated by Eq. (23). Generally, the higher the *Accuracy* is, the better the model performance is.

$$Accuracy = (TP + TN)\big/(TP + FN + FP + TN) \quad (23)$$

*Precision* or accuracy rate indicates the proportion of truly positive samples in the positively predicted samples. Generally, the higher the *Precision* is, the better the model effect is, as calculated by Eq. (24).

$$Precision = TP/(TP + FP) \quad (24)$$

*Recall* – also known as recall rate, shows the classifier's predictability against the positive samples. True positive rate (*TPR*) indicates the ratio of correctly identified positive samples to the total true positive samples. Generally, higher

*Recall* means that the more positive samples are predicted correctly, and, thus, the better the model effect is, as counted by Eq. (25):

$$Recall = TP/(TP + FN) \tag{25}$$

*Precision* and *Recall* a pair of contradictory indexes. A high *Precision* usually indicates a low*Recall*. Likewise, the higher *Recall* is, the low *Precision* will be.

*Specificity* refers to the ratio of the negatively identified samples to the total number of negative samples. The calculation of False Positive Rate (*FPR*) reads:

$$FPR = FP/(TN + FP) \tag{26}$$

Generally, the lower the proportion of negative samples identified positively in all negative samples is, the better the model performance is, as calculated by Eq. (27).

$$Specificity = 1 - FPR \tag{27}$$

F-Score – its physical meaning is a weighted average of the *Recall* and *Precision*, during which the weight of the former is F times of the latter. Then, the F-Score is defined as the harmonic average of *Recall* and *Precision*. The value of F-Score = 1 varies from 0 to 1, 1 being the optimal and 0 being the worst.

The above 1,000 groups of PD data are used for pattern recognition and classification. 80% of the data are randomly selected for training, 30% are used as the training set, 50% are used as the cross-validation set, and the remaining 20% are used for testing, and the model is tested multiple times.

## III. EXPERIMENTAL RESULTS OF GIS INSULATED SWITCHGEAR

### A. CLASSIFICATION AND RECOGNITION RESULTS OF THE MODEL TRAINING SET

Fig. 9 illustrates the results of the model's classification and recognition against the first four categories in the Mstar dataset.

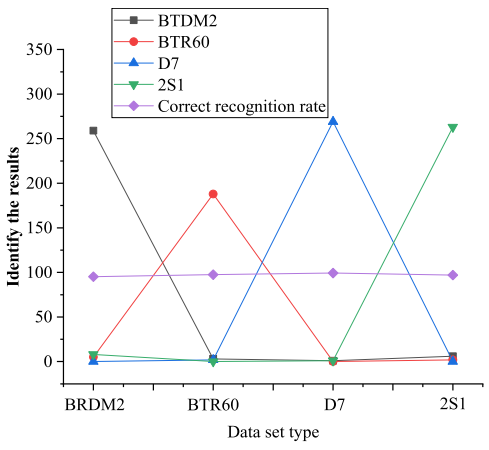


**FIGURE 9.** Classification and recognition results of the first four types of Mstar.

In Fig. 9, when the first four types of Mstar train the proposed Transfer Learning-optimized CNN, the recognition rate is generally more than 95%, the recognition rate for the BRDM2 dataset is the lowest, reaching 95.22%, and the recognition rate for 2S1 data set is slightly higher, reaching 97.05%. The recognition rate of the BTR60 data set is higher, reaching 99.26%, and the recognition rate for the D7 data set is the highest, 99.26%. Thus, after training, the overall model recognition rate reaches 97.23%.

### B. MODEL TRAINING SET EFFECT TEST

Table 2 compares the effects of the Transfer Learning-optimized CNN, CNN, and Deep CNN (DCNN) approaches.

**TABLE 2.** Comparison of testing results of mainstream network structures.

| Network structure | Recognition rate | Training time |
|---|---|---|
| CNN | 94.75% | 101.141s |
| DCNN | 95.55% | 104.235s |
| Transfer Learning-optimized CNN | 98.34% | 7.078s |

As shown in Table 2, compared with the mainstream network structures CNN and DCNN, the recognition rate of the proposed Transfer Learning-optimized CNN reaches 98.34%, slightly higher than 94.75% of CNN and 95.55% of DCNN. Compared with other CNN and DCNN structures requiring hundreds of seconds of training time, the Transfer Learning-optimized CNN structure takes about 7.078 seconds to train, and the effect is remarkable. Next, the first four types in Mstar dataset is used to verify the classification and recognition rate of the proposed Transfer Learning-optimized CNN for small datasets, as detailed in Table 3.

**TABLE 3.** Verification of the proposed method on small sample datasets.

| Number of samples | Recognition rate | Number of samples | Recognition rate |
|---|---|---|---|
| 1,000 | 70% | 600 | 96.4% |
| 2,000 | 88.6% | 700 | 96.6% |
| 3,000 | 95.2% | 800 | 96.8% |
| 4,000 | 95.4% | 900 | 96.9% |
| 5,000 | 96.2% | 1,000 | 96.9% |

Table 3 suggests that when the data volume reaches 600, the verification results stabilize at about 96.5%, close to 97.23% (the training results on the large dataset), which means that the training results of medium-sized training sets are good. When data volume increases from 100 to 300, the recognition rate increases sharply, from 70% to 95.2%, obviously higher than 95%, and beats the traditional CNN. Therefore, the proposed Transfer Learning-optimized CNN is also excellent for training and recognizing small samples sets with at least a data volume of 300.

### C. TEST RESULTS OF ACTUAL GAS INSULATED SWITCHGEAR DATASET

Further, the proposed Transfer Learning-optimized CNN model is trained and tested and is compared with different models, as in Fig. 10.

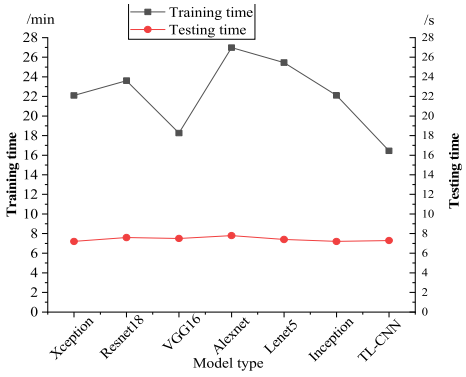


**FIGURE 10.** Comparison of training and testing time of different models.

Fig. 10 reveals the relationship between the six mainstream DL models and the proposed Transfer Learning-optimized CNN. Excessive training duration will make it impossible to monitor and manage the health status of complex equipment in real time, so it is necessary to shorten the testing time. Meanwhile, lengthy training time leads to poor model updating, hinders model training, and, thus, fails to improve the model accuracy. The training time of the proposed Transfer Learning-optimized CNN model is 16min, the shortest by comparison with other DL models. In terms of the testing time, the proposed Transfer Learning-optimized CNN model spends almost as little as the model with the shortest training time, with only about 0.1s difference. Therefore, the proposed Transfer Learning-optimized CNN model performs best in training and testing time.

### D. TEST RESULTS OF DISCHARGE DEFECT TYPES OF ACTUAL GAS INSULATED SWITCHGEAR

The trained Transfer Learning-optimized CNN and other mainstream NN models are used to classify the pictures, and various GIS PD pattern recognition results are obtained, as in Fig. 11.

In Fig. 11, M represents free metal defect discharge, N denotes metal tip defect discharge, P means air gap defect discharge, and O stands for suspended electrode defect discharge. Fig. 8 displays that the LeNet5 model has a low PD pattern recognition rate due to the obvious lack of learning of feature representation, which is related to the excessive loss of image features or the insufficient depth of the model the inability to extract feature information fully. The proposed Transfer Learning-optimized CNN model has a recognition rate of 99.6% for M-type defects, 99.2% for N-type defects, 89.6% for O-type defects, 97.6% for O-type defects, and the

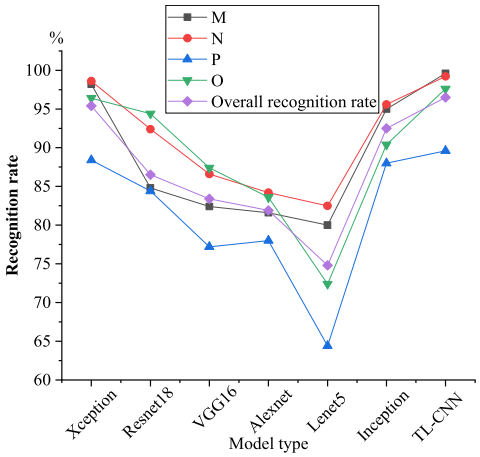


**FIGURE 11.** PD pattern recognition results of various GIS.

overall recognition rate is 96.5%. Compared with the other algorithm models, it has the highest recognition rate for various PD patterns. The overall recognition rate is above 96%, which shows that the proposed Transfer Learning-optimized CNN model has the best recognition effect in practical application. In all pattern recognition methods, the recognition rate for insulator air gap fault is relatively low, mainly caused by the small air gap in epoxy resin or the instability of PD in the air gap between insulating materials and metal inserts. By comparison, the algorithm models that employ start-from-scratch training, such as VGG16, AlexNet, Inception, cannot be trained due to the gradient disappearance. The Transfer Learning training method improves the model recognition rate significantly.

For GIS PD pattern recognition and classification, the model trained with the Mstar dataset will be transferred, and then the parameter update training will be carried out. The proposed Transfer Learning-optimized CNN models' parameter volume is 408, and the storage cost space is 12.8MB. Obviously, the MobileNet model has the best performance, indicating that it is more suitable to be embedded into the intelligent terminal under the ubiquitous IoE. MobileNet is less bulky and has less computation and higher accuracy. It has significant advantages in lightweight NNs. Under Google Pixel-1 mobile phone test, all versions of MobileNet can keep a running time below 120ms. The following is the structure of MobileNet applied in this paper: (a). The whole network does not calculate the average pooling and SoftMax layers, with altogether 28 layers. (b). In the whole network structure, the convolution with a step size of 2 is more characteristic, and the convolution acts as the downsampling simultaneously. (c). The 27 layers after the first layer are the repeated convolution operation of depthwise separable convolution. (d). Each convolution layer (including conventional, depthwise, and pointwise convolution) is followed by batch normalization and ReLU AF. (e). The last fully-connected layer does not use the AF.

### *E. CONVOLUTIONAL NEURAL NETWORK ALGORITHM CORRELATION*

Figs. 12 and 13 illuminate the simulation diagrams of the relationship between the training and testing sets' measured values and predicted values. The simulation value represents the model predicted value, and the measured value is the actual value of PD.

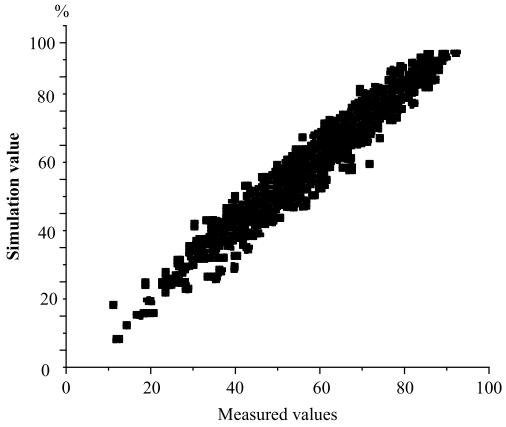


**FIGURE 12.** Simulation of the relationship between measured and predicted values of the training set.

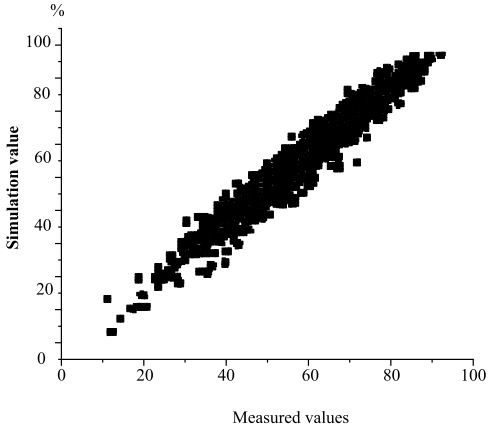


**FIGURE 13.** Simulation of the relationship between measured and predicted values of the testing set.

$R^2$ of the training set and the testing set are 0.95 and 0.91 after the CNN model training. Then, the correlation is evaluated between the simulated and measured values. The test results show that both the training and testing sets show a high correlation between simulated and measured values. The fast-fitting is a positive correlation line, and its GOF is strong. When the error ratio of training and testing is close to 1.02, the gap between the testing and training errors is tiny. Therefore, the training is more successful, and the training effect is better than other similar algorithms.

## IV. CONCLUSION

The IoT technology and Transfer Learning-optimized CNN are used to study the complex equipment-oriented HMS to optimize the GIS state management and reduce the possible risks to personnel and equipment in the production process. Firstly, it expounds on the principle of Transfer Learning and DL technology. Secondly, the internal state identification and management requirements of GIS are studied. On this basis, a GIS-oriented HMS is proposed using the Transfer Learning-optimized CNN, and the training dataset is constructed. Finally, the proposed Transfer Learning-optimized model is verified. The experimental findings imply that the proposed Transfer Learning-optimized CNN model has good robustness. Meanwhile, given the strong training ability of CNN, the experimental results after training show that the overall recognition rate reaches 97.23%, which further shortens the iterative evolution speed of network training. The overall recognition rate of the Transfer Learning-optimized CNN model is 96.5%. Compared with the other six algorithm models, the proposed model has the highest recognition rate for various defects-induced PD, with an overall recognition rate of 96%. The model training and testing time are short, and the convergence speed is fast. The training and testing sets' $R^2$ is 0.95 and 0.91, respectively, which have a high correlation. Besides, the fast fitting is a positive correlation line, so the GOF is very high. Thus the proposed Transfer Learning-optimized CNN can effectively manage equipment health and further improve the recognition rate and model reliability of poor working state caused by GIS's PD.

However, although this paper has achieved good research results, there are still has some deficiencies. The research on the GIS model is not enough. The proposed method has not been verified under extreme conditions, so future work will focus on improving the detection and management effectiveness of the model under extreme conditions against GIS PD.

## REFERENCES

[1] M. Elsisi, M. Q. Tran, K. Mahmoud, and D. E. A. Mansour, "Towards secured online monitoring for digitalized GIS against cyber-attacks based on IoT and machine learning," *IEEE Access*, vol. 9, pp. 78415–78427, 2020.

[2] M. Elsisi, M. Q. Tran, K. Mahmoud, and D. E. A. Mansour, "Effective IoT-based deep learning platform for online fault diagnosis of power transformers against cyberattack and data uncertainties," *Measurement*, vol. 190, Feb. 2022, Art. no. 110686.

[3] M. Elsisi, M.-Q. Tran, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Deep learning-based industry 4.0 and Internet of Things towards effective energy management for smart buildings," *Sensors*, vol. 21, no. 4, p. 1038, Feb. 2021.

[4] M. Elsisi, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Reliable industry 4.0 based on machine learning and IoT for analyzing, monitoring, and securing smart meters," *Sensors*, vol. 21, no. 2, p. 487, Apr. 2021.

[5] D. V. Zelyakovskiy, G. N. Sineva, S. V. Volobuev, P. V. Prokofiev, and K. E. Tokarev, "The intelligent system for monitoring and technical diagnostics of complex technological equipment in the agro-industrial complex," in *Proc. J. Phys., Conf.*, vol. 1801, no. 1, Feb. 2021, Art. no. 012032.

[6] Q. Xiao, T. Huang, and Z. Zeng, "Passivity and passification of fuzzy memristive inertial neural networks on time scales," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3342–3355, Dec. 2018.

[7] Z. Ding, J. Wang, Y. Cheng, and C. He, "Alice: A LSTM neural network based short-term power load forecasting approach in distributed cloud-edge environment," in *Proc. J. Phys., Conf.*, vol. 1624, no. 5, Oct. 2020, Art. no. 052017.

[8] J. Narwariya, "Graph neural networks for leveraging industrial equipment structure: An application to remaining useful life estimation," *Int. Vis. Comput.*, vol. 6, no. 3, pp. 120–128, Jun. 2020.

[9] J. Windau and L. Itti, "Inertial machine monitoring system for automated failure detection," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 93–98.
[10] Y. Chen, S. Hu, H. Mao, W. Deng, and X. Gao, "Application of the best evacuation model of deep learning in the design of public structures," *Image Vis. Comput.*, vol. 102, Oct. 2020, Art. no. 103975.
[11] Y. Cao, X. Zhu, and Y. Tan, "Design of heat dissipation structure for high current medium voltage AC gas-insulated metal-enclosed switchgear," *High Voltage Appl.*, vol. 54, no. 1, pp. 200–206, Sep. 2018.
[12] D. Scerpella, N. G. Bouranis, and M. J. Webster, "Using geographic information systems (GIS) for targeted national recruitment of community-dwelling caregivers managing dementia-related behavioral and psychological symptoms: A recruitment approach for a randomized clinical trial," *Geographic Inf. Syst.*, vol. 13, no. 3, p. 16, Mar. 2021.
[13] J. Xue, H. Wang, Y. Liu, K. Li, X. Liu, and X. Fan, "Surface charge distribution patterns of a truncated cone-type spacer for high-voltage direct current gas-insulated metal-enclosed transmission line/gas-insulated metal-enclosed switchgear," *IET Sci., Meas. Technol.*, vol. 12, no. 4, pp. 436–442, May 2018.
[14] S. Sawyer, "The Internet of Things," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 70, no. 6, pp. 638–639, Nov. 2019.
[15] A. Hany, W. Robert, and W. Gary, "Fog computing and the Internet of Things: A review," *Big Data Cogn. Comput.*, vol. 2, no. 2, pp. 10–11, March,2018.
[16] R. E. McRoberts, Q. Chen, B. F. Walters, and D. J. Kaisershot, "The effects of global positioning system receiver accuracy on airborne laser scanning-assisted estimates of aboveground biomass," *Remote Sens. Environ.*, vol. 207, no. 2, pp. 42–49, Mar. 2018.
[17] H. Saadi, R. Touhami, and M. Yagoub, "TDMA-SDMA-based RFID algorithm for fast detection and efficient collision avoidance," *Int. J. Commun. Syst.*, vol. 31, no. 1, p. e3392, Aug. 2018.
[18] F. Terroso-Saenz, A. González-Vidal, A. P. Ramallo-González, and A. F. Skarmeta, "An open IoT platform for the management and analysis of energy data," *Future Gener. Comput. Syst.*, vol. 92, pp. 1066–1079, Mar. 2019.
[19] A. Karbalay, X. Qian, and E. R. Dougherty, "Optimal Bayesian transfer learning," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3724–3739, Jul. 2018.
[20] D. George, H. Shen, and E. A. Huerta, "Classification and unsupervised clustering of LIGO data with deep transfer learning," *Phys. Rev. D, Fields*, vol. 97, no. 10, May 2018, Art. no. 101501.
[21] G. Montavon, W. Samek, and K. R. Müller, "Methods for interpreting and understanding deep neural networks," *Digit. Signal Process.*, vol. 73, no. 1, pp. 1–15, Jun. 2018.
[22] S. Zhang, F. Sun, N. Wang, C. Zhang, Q. Yu, M. Zhang, P. Babyn, and H. Zhong, "Computer-aided diagnosis (CAD) of pulmonary nodule of thoracic CT image using transfer learning," *J. Digit. Imag.*, vol. 32, no. 6, pp. 995–1007, Dec. 2019.
[23] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, and E. Ferrero, "Opportunities and obstacles for deep learning in biology and medicine," *J. Roy. Soc. Interface*, vol. 15, no. 141, pp. 45–49. May 2018.
[24] D. S. Kermany, M. Goldbaum, W. Cai, and C. C. S. Valentim, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
[25] A. M. Abdelhakim and M. Abdelhakim, "A time-efficient optimization for robust image watermarking using machine learning," *Expert Syst. Appl.*, vol. 100, no. 7, pp. 197–210, Jun. 2018.
[26] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, and A. Farag, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Med. Image Anal.*, vol. 2, no. 3, pp. 94–96, Aug. 2018.
[27] S. Vekkot and D. Gupta, "Speaker-independent expressive voice synthesis using learning-based hybrid network model," *Int. J. Speech Technol.*, vol. 23, no. 5, pp. 35–38, Jan. 2020.
[28] A. Bassam, B. Ayyoub, and A. RitahaniIsmail, "A proposed mathematical model for tripartite synapse to enhance artificial model for artificial neuron-astrocyte networks," *Int. J. Adv. Comput. Sci. Eng.*, vol. 9, no. 5, pp. 8342–8346, Nov. 2020.
[29] V. K. Rai and R. Sakthivel, "Design of artificial neuron network with synapse utilizing hybrid CMOS transistors with memristor for low power applications," *J. Circuits, Syst. Comput.*, vol. 1, no. 6, pp. 167–170, Oct. 2020.
[30] L. Diener, M. Janke, and T. Schultz, "Deep neural networks," *IEEE Access*, vol. 30, no. 5, pp. 1233–1243, 2022.
[31] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, no. 10, pp. 321–331, Dec. 2018.
[32] J. Lee, J. Park, K. L. Kim, and J. Nam, "SampleCNN: End-to-end deep convolutional neural networks using very small filters for music classification," *Appl. Sci.*, vol. 8, no. 1, p. 150, Jan. 2018.
[33] G. Scarpa, M. Gargiulo, A. Mazza, and R. Gaetano, "A CNN-based fusion method for feature extraction from sentinel data," *Remote Sens.*, vol. 10, no. 2, p. 236, Feb. 2018.
[34] M. Jaritz, R. De Charette, and E. Wirbel, "Sparse and dense data with cnns: Depth completion and semantic segmentation," *Sensor*, vol. 9, no. 5, pp. 52–60, Jan. 2018.