




SA1D-CNN: A Separable and Attention Based Lightweight Sensor Fault Diagnosis Method for Solar Insecticidal Lamp Internet of Things

XING YANG ^{1,3} (Graduate Student Member, IEEE), LEI SHU ^{2,3} (Senior Member, IEEE), KAILIANG LI², ZHIQIANG HUO ⁴, AND YU ZHANG ⁵ (Member, IEEE)

¹College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

²College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China

³School of Engineering, University of Lincoln, LN6 7TS Lincoln, U.K.

⁴Institute of Health Informatics, University College London, WC1E 6BT London, U.K.

⁵Department of Aeronautical and Automotive Engineering, Loughborough University, LE11 3TU Loughborough, U.K.

CORRESPONDING AUTHOR: LEI SHU (E-mail: lei.shu@njau.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62072248, in part by the Jiangsu Agriculture Science and Technology Innovation Fund under Grant CX(21)3060, in part by the High-End Foreign Experts Recruitment Plan of the Ministry of Science and Technology China under Grant G2021145009L, in part by the International Cooperation Capacity Enhancement Plan of Nanjing Agricultural University under Grant 2021-PY-10, and in part by the China Scholarship Council under Grant 202106850023.

ABSTRACT Sensor faults can produce abnormal and spurious observations in the solar insecticidal lamp Internet of Things (SIL-IoTs) system. Early detection and identification of the sensor node's abnormality are critical to ensure the SIL-IoTs system's reliability. In this study, we propose a lightweight separable 1D convolution neural network that can be implemented in SIL-IoTs nodes to identify sensor faults, reduce detecting delay, and decrease data transmission. However, the reliability of data acquired by sensors is decreased because a SIL-IoTs node releases high voltage pulse discharge (a kind of electromagnetic interference) when pests collide with its metal mesh. This kind of data fluctuation impacts fault diagnosis accuracy. Consequently, fault-related feature maps and temporal signals are characterized via a novel time and channel attention module (TCAM) method, which contributes to separating electromagnetic interference noise from sensor faults of SIL-IoTs nodes. A real-world testbed is applied to validate the effectiveness of the proposed method on sensor fault diagnosis in the SIL-IoTs system. Experimental results demonstrate that the proposed method can detect four typical sensor faults with the best trade-off between accuracy (99.9% average accuracy and 97.6% average F1-score) and efficiency (351 KB inference model size and 4.33 W average energy consumption on Raspberry Pi).

INDEX TERMS Sensor fault diagnosis, solar insecticidal lamp internet of things, depthwise separable convolution, attention mechanism.

I. INTRODUCTION

In the agricultural environment, IoT-based devices are usually deployed in harsh settings and are inevitably vulnerable to unexpected failures and malfunctions, introducing adverse influences to not only the sensor system but also the agricultural operation system [1]. Solar insecticidal lamp Internet of Things (SIL-IoTs) is a novel agricultural IoT concept that aims to destroy pests using solar insecticidal lamps (SILs). An example of a SIL-IoTs device is shown in Fig. 1, where energy is harvested by the solar panel and stored in the battery

during the day. The lure lamp and metal mesh are turned on at night to attract phototactic pests and then kill them by releasing high voltage pulse discharge. IoT devices are adopted to 1) count the number of killed pests to predict the regularity of pest infestation in different regions; 2) and monitor the residual energy of the battery for maintaining the battery. The on-off time of the SIL can be determined by the predicted number of killed pests and residual energy of the battery to turn off the SIL when there is no gathering of pests. These data can be processed locally or

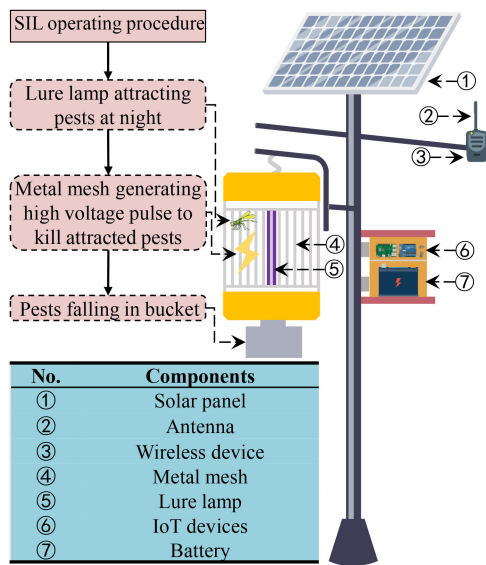


FIG. 1. The operating procedure and components of an SIL-IoTs node.

transmitted to the cloud center by a wireless communication device (e.g., ZigBee).

SIL-IoTs nodes are deployed in difficult outdoor environments; thus, they are vulnerable to faults. In addition, SIL-IoTs nodes hung on metal poles with heights from 2 m to 3 m results in repair and maintenance difficulty. When a fault occurs, the SIL-IoTs node may be abnormal or unavailable for a long time. For instance, when the metal mesh fails, the SIL-IoTs node continuously attracts pests but cannot kill them, leading to pests' aggregation and harm to crops. Hence, sensors are applied to monitor the working conditions of SIL-IoTs nodes, e.g., voltage and current sensors. However, the performance of sensors in SIL-IoTs nodes is affected by burn-in, damage, low quality, and improper maintenance [2]–[4]. When sensors become faulty, data unreliability can be a major drawback for the functions of SIL-IoTs (e.g., energy management, counting of killed pests, and monitoring components of SIL-IoTs) [5].

Recent advances in SIL-IoTs mainly focus on node deployment [6], [7], energy management [8], and interference in data transmission and acquisition due to high voltage pulse discharge [9]–[11]. Only one survey paper [12] studied the fault characteristics of SIL-IoTs without a specific diagnostic model. A sensor fault method is therefore needed for SIL-IoTs. Specifically, the influence of equipment operation on sensor data acquisition in specific scenarios is an important factor to consider in the design of fault diagnosis methods. For SIL-IoTs, the SIL releases high voltage pulse discharge (from 2150 V to 6000 V) when pests collide with the metal mesh, which leads to fluctuations in voltage and current measurements [9]–[11]. This fluctuation can be treated as noise and does not represent a sensor fault. Anti-noise control contributes to the robustness of fault diagnosis model [13], [14]. Therefore, it is critical to separate this condition from actual sensor fault states in SIL-IoTs.

A. RELATED WORKS

In recent decades, a variety of studies have been conducted on diagnosing sensor nodes or sensor systems in a wide range of condition monitoring systems. Broadly, the applied techniques can be categorized into model-based methods, signal-based methods, and data-driven methods. Model-based methods construct sophisticated white-box models by using the underlying physical principles. Filters [15], [16] and state observers [17] are the most effective approaches that diagnose sensor faults by monitoring and evaluating the system residuals. Model-based methods have low diagnostic latency; however, it is difficult to set the appropriate threshold and their performance highly relies on parameter estimation and expert knowledge.

Signal-based methods do not rely on system parameters. For instance, Deng *et al.* [18] adopted the envelope demodulation analysis method to extract faulty features. The results indicate that faulty features of rolling bearings can be separated and extracted effectively. In addition, Cui *et al.* [19] applied variational mode decomposition to decompose vibration signals into a series of intrinsic mode functions, which can be used to enhance periodic fault features of reconstructing signals.

Recently, data-driven methods using deep learning techniques have been widely applied to analyze 1D signals for sensor fault diagnosis, such as convolutional neural networks (CNNs). In previous studies, 1D signals were usually converted to 2D spectrum images, and then 2D-CNN was applied to classify faults, resulting in a more complicated model design and consuming additional computing process resources.

Recent researches focused on processing 1D signals directly without any transformation [20]–[25]. Typical works include wide first-layer kernels to diagnose 1D vibration signals [20], conditional adversarial diagnosis for the model collapse in domain adversarial training [21], a residual learning-based 1D-CNN model for a strong noise environment [22] and automatically determining the thresholds [23], deep convolutional transfer learning for the difficulty of massive labeled data [24], and a multiattention mechanism for understanding and learning features [25]. According to the specific scenario requirements, the above approaches have investigated optimized 1D-CNN models against noise and few labeled data.

To evaluate developed models on IoT devices, there are growing appeals for resource-limited application of fault diagnosis models [26]. Due to the advantages of small resource consumption, short response time, and bandwidth saving, lightweight CNN models (CNN models with fewer parameters and computational costs) are beneficial for monitoring IoT devices.

SqueezeNet [27], Xception [28], MobileNet [29], and ShuffleNet [30] are the most typical lightweight CNN architectures. The strategies can be summarized as follows:

- Using optimal convolution instead of regular convolution. For instance, ShuffleNet divides input channels into several groups and performs convolution independently for each group of channels. SqueezeNet uses 1 × 1 filters

instead of 3×3 filters and decreases the number of input channel. MobileNet and Xception adopt depthwise separable convolution based on existing CNN architectures.

- Maximizing the models' performance with existing architecture. For instance, ShuffleNet changes the order of the channels to enhance the interaction between different grouped channels. Linear bottlenecks are added in MobileNet V2 [31] to avoid manifold collapse. Then, the squeeze and excite (SE) module is adopted in MobileNet V3 [32] to improve the channel interaction by evaluating the importance of different filters.
- Using global average pooling instead of fully connected layer to significantly decrease the number of parameter and improve the training speed.

In the fault diagnosis domain, several researchers [26], [33], [34] considered the lightweight fault diagnosis model and discussed the performance of the model on IoT devices. For instance, [33] concatenated a rectified linear unit lightweight convolution method for extracting features by a deeper network and fast learning mechanism. Interestingly, depthwise separable convolution was adopted and achieved great performance in reducing parameters according to the results of these works.

However, in these works, a signal processing operation is required with long-term signals (e.g., 1024×1 [33], 5000×1 [26], and $12,800 \times 1 \times 2$ [34]) before setting them as input data. Long-term signals lead to extra resource consumption and are not suitable for agriculture IoT devices with the low acquisition frequency (e.g., 0.2 Hz for environmental conditions).

B. CHALLENGES

1D-CNN fault diagnosis methods have high diagnostic accuracy and can be improved to adjust to specific scenarios, e.g., the SIL-IoTs in this paper. 1D-CNN fault diagnosis schemes in the SIL-IoTs scenario encounter the following challenges:

- Due to the lack of supervision for SIL-IoTs nodes, it is difficult to detect faults promptly. In addition, detecting faults in the cloud center via collecting bulk data from IoT nodes leads to data redundancy, bandwidth occupation, and privacy leakage. Thus, embedding the fault diagnosis scheme in SIL-IoTs nodes is suggested.
- Due to limited energy, storage and computing capacity, it is difficult to store and analyze a large quantity of data on the node. Large input data (e.g., $12,800 \times 1$) are also not suitable for real-time diagnosis. In addition, the SIL-IoTs is a type of wireless sensor network, and the misdiagnosis of one node has an impact on neighboring nodes (e.g., a diagnostic accuracy difference of 2% leads to a misdiagnosis of 20 nodes, which then affects dozens to hundreds of neighboring nodes). Hence, a rapid response to faults is critical for SIL-IoTs.
- High-quality and labelled data are critical for the initial stage of the SIL-IoTs. However, it is difficult to collect and label the noise data caused by high voltage discharge

because the moment when pests collide with the metal mesh is unpredictable.

C. CONTRIBUTIONS

To address the mentioned challenges, a lightweight 1D-CNN-based sensor fault diagnosis scheme for SIL-IoTs nodes is proposed, namely, separable and attention 1D-CNN (SAID-CNN). By diagnosing faults in SIL-IoTs nodes, fault states are transmitted to the cloud rather than all measurements. In this work, we focus on a lightweight fault diagnosis scheme rather than an exhaustive survey of fault types in SIL-IoTs nodes.

The main contributions of this article are summarized as follows:

- 1) Because data fluctuation cases are caused by high voltage pulse discharge of the SIL (no-fault state and may lead to the misdiagnosis of faults), we propose the time-channel attention module (TCAM) to enhance fault-related features and suppress irrelevant features.
- 2) To reduce diagnostic delay and data transmission, a lightweight and easy implementing fault diagnosis model (namely SAID-CNN) is embedded and performed in the SIL-IoTs node for the first time. SAID-CNN is composed of depthwise separable convolution and TCAM.
- 3) The diagnostic accuracy of the proposed method is verified in SIL-IoTs nodes. The results indicate that the proposed method has the best trade-off performance between diagnostic accuracy and computationally efficiency than the referenced algorithms.

The rest of the paper is organized as follows: Section II describes the system model and four different types of sensor fault. Section III presents the design of our method. Section IV introduces the experimental data collection and Section V discusses the performance of the proposed model. Finally, Section VI concludes the paper.

II. SYSTEM MODEL AND PROBLEM STATEMENT

This section describes and explains the system model, specific sensor faults, and abnormal data types caused by high voltage discharge.

A. SYSTEM MODEL

To help readers understand how sensors monitor the SIL-IoTs, a simple SIL-IoTs monitoring system is illustrated, as shown in Fig. 2. Under normal conditions, voltage and current sensors are used to monitor the status of SIL. However, high voltage discharge is released while killing insects, resulting in unexpected data fluctuation [9].

In practice, the time and frequency of high voltage discharge are hard to determine; therefore, a discharge simulation module is designed to stably control the discharge frequency, which controls the discharge frequency by setting up the on-off frequency of 8 electromagnetic relays. Subsequently, the microcontroller (Arduino) transmits sensed data to the Raspberry Pi. Ultimately, data are analyzed by Raspberry Pi, and

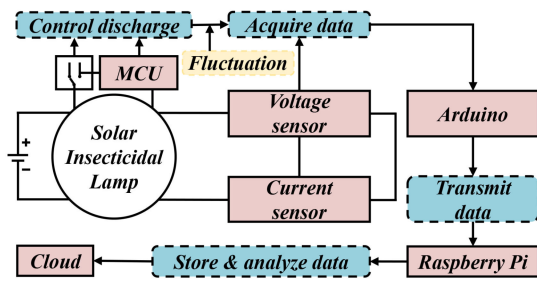


FIG. 2. System diagram. Data are acquired from voltage and current sensors to Arduino. Then Raspberry Pi analyzes data transmitted from Arduino and sends fault conditions to the cloud center.

abnormal conditions are reported and transmitted to the cloud center by the wireless device.

B. PROBLEM STATEMENT

For SIL-IoTs, Abnormal sensor conditions (data) are defined and derived for three reasons: 1) electromagnetic interference, 2) current increase and voltage decline when the SIL releases high voltage pulse discharge, and 3) sensor faults of SIL-IoTs nodes. Under normal working conditions, high voltage discharge is usual when the SIL delivers killing insect tasks. Therefore, the corresponding data should not be diagnosed as a fault in this case. Distinguishing abnormal sensor conditions or data caused by high voltage discharge from actual sensor faults is essential for monitoring the state of SIL-IoTs. In this study, four typical sensor faults are considered: outlier fault, gain fault, stuck fault and offset fault [35], which are defined below:

$$x' = \begin{cases} x + \theta, & \text{Outlier fault} \\ \beta x + \eta, & \text{Gain fault} \\ \omega, & \text{Stuck fault} \\ x + \omega + \eta, & \text{Offset fault} \end{cases} \quad (1)$$

where x and x' are measurements of the fault-free sensor and fault sensor, respectively. $x + \theta$ is an outlier fault (caused by short-term vibration and electromagnetic inference, deviates significantly from x , θ is the deviation value). $\beta x + \eta$ is a gain fault (caused by the excessive signal noise inside the sensors, the change rate of x does not match the expectation, β is the gain coefficient). ω is a stuck fault (caused by the power supply interruption and sensors vandalism, x becomes a constant). $x + \omega + \eta$ is an offset fault (caused by the improper calibration of sensors, addition of deterministic bias from x , η is the noise in data which is much smaller than ω).

The fault states of the SIL-IoTs sensors are labeled differently as numeric numbers, shown in Table 1. To diagnose the system's ten health states, a novel deep learning based fault diagnosis method is proposed. The inputs are integrated voltage and current data, and the output is the predicted label that indicates the specific sensor fault type.

TABLE 1 Sensor Conditions Labelling

Fault types	Discharge	No discharge	Label
No fault		✓	0
No fault	✓		9
Voltage sensor outlier fault	✓		1
Voltage sensor gain fault	✓	✓	2
Voltage sensor stuck fault	✓	✓	3
Voltage sensor offset fault	✓	✓	4
Current sensor outlier fault	✓	✓	5
Current sensor gain fault	✓	✓	6
Current sensor stuck fault	✓	✓	7
Current sensor offset fault	✓	✓	8

III. PROPOSED METHODOLOGY

In this paper, the depthwise separable convolution layer [29] is used as a basic tool. The TCAM is proposed for extracting important features of input to improve diagnostic accuracy. Benefiting from parameter reduction in the depthwise separable convolution layer and feature recalibration of the TCAM, a lightweight fault diagnosis scheme, named SA1D-CNN, is proposed.

A. FRAMEWORK DESIGN

Running data of various modes are measured and labeled in the first step of the proposed fault diagnosis framework as shown in Fig. 3. The appropriate model structure is critical for improving accuracy and reducing the complexity of the model. According to [29], the depthwise separable convolution layer and the global average pooling layer significantly contribute to the reducing of model's parameters under a slight decrease in diagnostic accuracy. Moreover, an attention mechanism can be added to networks, which helps improve diagnostic accuracy based on adding a few parameters. To improve diagnostic speed, the fault diagnosis model is deployed on IoT devices, which is conducive to eliminating the influence of data transmission delay and reducing bandwidth usage. In this manner, the online lightweight model can be applied to diagnose sensor faults periodically.

B. DEPTHWISE SEPARABLE CONVOLUTION

Depthwise separable convolution decomposes the regular convolution into a depthwise convolution and a 1×1 convolution, which is also called pointwise convolution [29]. The difference between regular convolution and depthwise separable convolution is shown in Fig. 4. Assume that input is represented as $X \in \mathbb{R}^W \times C$, each filter size is $(k \times 1)$, each stride size is 1, and the output is represented as $X' \in \mathbb{R}^{W'} \times C'$, where W and W' denote the length of input features and output features, with C and C' denoting the number of input channels and output channels, respectively. For regular convolution, the inputs are convolved with C filters to obtain C' feature maps. For depthwise separable convolution, an input channel is convolved by only one depthwise convolution. Thus, the number of feature map is equal to the number of input channel. To generate C' feature maps, C' filters with a size of $1 \times 1 \times C$

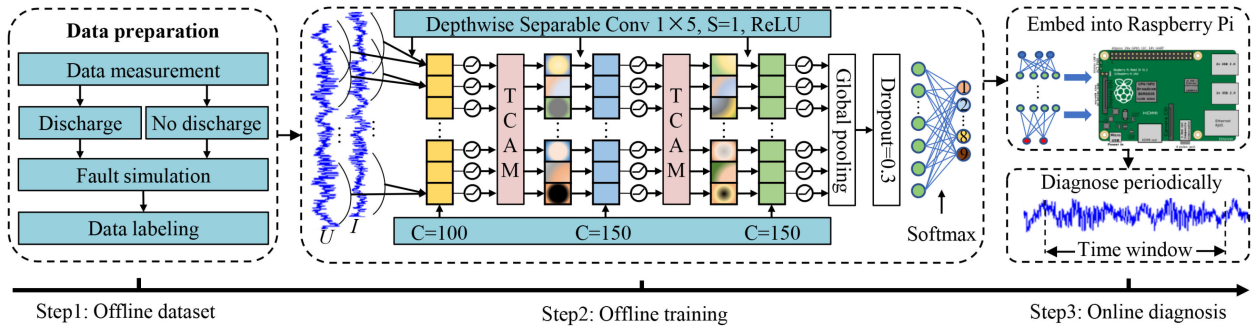


FIG. 3. Flowchart of the proposed end-to-end fault diagnosis method. (U denotes voltage values. I denotes current values. S and C represent strides and channel numbers of convolution operations.).

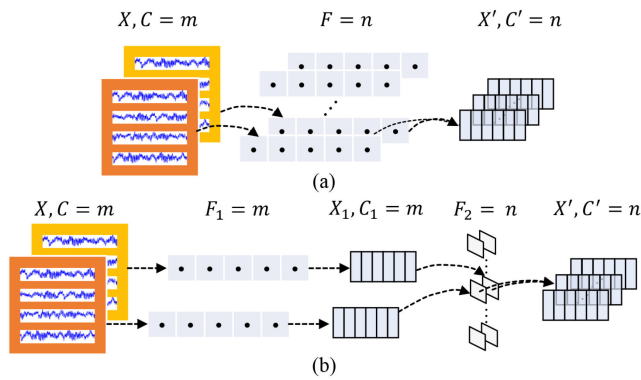


FIG. 4. Comparison of (a) regular 1D convolution and (b) depthwise separable 1D convolution. X and X' denote input and output feature maps. C and C' denote the number of input channels and output channels. F represents the number of filters.

are applied. The parameters $Params_{reg}$ and Floating Point Operations (FLOPs) $FLOPs_{reg}$ of the regular convolution are calculated by:

$$Params_{reg} = (C \times k + 1) \times C' \quad (2)$$

$$FLOPs_{reg} = k \times 1 \times C \times C' \times W \times 1 \quad (3)$$

The parameters $Params_{sep}$ and FLOPs $FLOPs_{sep}$ of the depthwise separable convolution are calculated by:

$$Params_{sep} = C \times k + C' \times (C \times 1 + 1) \quad (4)$$

$$FLOPs_{sep} = k \times 1 \times C \times W \times 1 + C \times C \times W \times 1 \quad (5)$$

Therefore, the FLOPs can be reduced by:

$$\frac{FLOPs_{sep}}{FLOPs_{reg}} = \frac{1}{C'} + \frac{1}{k} \quad (6)$$

For instance, assume that $X \in \mathbb{R}^{60 \times 2}$, each filter size is (5×1) , and C' is 100. The parameters decrease from 5,100 to 350, and FLOPs decrease approximately 5 times.

C. ATTENTION MODULE

Different feature maps recognize fault features to different degrees. Thus, the diagnostic accuracy can be improved by strengthening the features with high contribution rate and suppressing the invalid features. By recalibrating the input

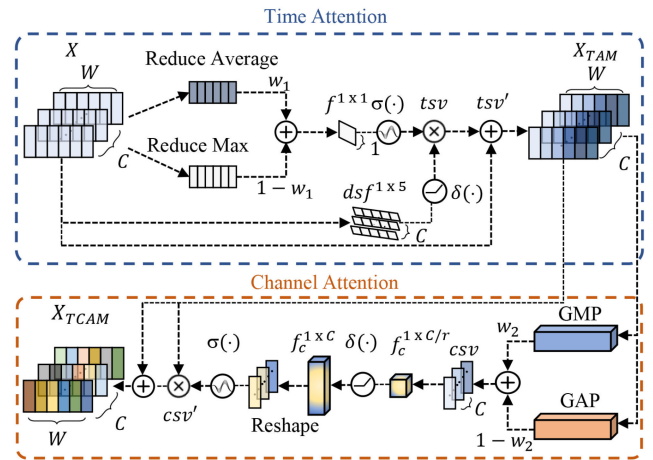


FIG. 5. Proposed time and channel attention module. f , dsf , and f_c denote convolution, depthwise separable convolution, and full connection operations. σ and δ represent sigmoid and ReLU activations. w_1 and w_2 denote weight parameters.

features, the attention module contributes to distinguishing different channels and time data. In this paper, the TCAM is proposed. The basic structure of the TCAM is shown in Fig. 5, which consists of a time attention module (TAM) and a channel attention module (CAM).

1) TIME ATTENTION MODULE

The TAM is proposed to enable the network to learn the temporal correlation of fault data. Assume that the input features $X = [x^1, x^2, \dots, x^W]$, where $x^j \in \mathbb{R}^{1 \times C}$ denotes the j th temporal data index and $j = [1, 2, \dots, W]$. First, TAM takes the average value and max value from X along the channel axis (namely F_{ra}^W and F_{rm}^W) to generate efficient features X_{avg} and X_{max} . Then, the projections of the sum of weighted X_{rm} and X_{ra} are obtained by a 1×1 convolution operation with one channel. The sigmoid function namely $\sigma(\cdot)$ is applied to generate the timewise statistics vectors $tsv(X) \in \mathbb{R}^{1 \times W}$, as shown in (7).

$$tsv(X) = \sigma(f^{1 \times 1}(w_1 \times F_{ra}^W(X) + (1 - w_1) \times F_{rm}^W(X))) \quad (7)$$

where $f^{1 \times 1}$ denotes a convolution operation with the filter size of 1×1 .

To keep the input and output identical dimensions, $tsv'(X) \in \mathbb{R}^{1 \times WC}$ is achieved by multiplying $tsv(X)$ and $s(X) \in \mathbb{R}^{1 \times WC}$, which is obtained by a depthwise separable convolution operation with a filter size of 1×5 and C channels (namely $dsf^{1 \times 5}$). To decrease the reduction in the response value of features caused by repeated feature recalibration operations, a residual connection [36] is used to optimize TAM. Finally, the output of TAM is shown in (8).

$$X_{TAM} = X + tsv'(X) = X + \delta(sf^{1 \times 5}(tsv(X))) \quad (8)$$

where δ represents the ReLU function.

2) CHANNEL ATTENTION MODULE

Different from the TAM (which aims to emphasize and suppress temporal signals), the CAM focuses on fault-related feature maps by exploiting the interchannel relationship of features. To improve the diagnostic accuracy of SA1D-CNN, the CAM is applied to the weight output of TAM in this paper. Assume that the input is represented as $X = [x_1, x_2, \dots, x_C]$, where $x^j \in \mathbb{R}^{W \times 1}$ denotes the j th channel and $j = [1, 2, \dots, C]$. To aggregate temporal information, both Global Average Pooling (GAP) and Global Max Pooling (GMP) operations are applied to generate two different channel descriptors: F_{gap}^C and F_{gmp}^C . Then the sum of weighted descriptors are forwarded to a shared network to produce channelwise statistics vectors ($csv(X) \in \mathbb{R}^{1 \times C}$), as shown in (9).

$$csv(X) = w_2 \times F_{gmp}^C(X) + (1 - w_2) \times F_{gap}^C(X) \quad (9)$$

The channel recalibration vectors ($csv'(X) \in \mathbb{R}^{1 \times C}$) are obtained by multiplying X and the results of two fully connected operations (namely $f_c^{1 \times C/r}$ and $f_c^{1 \times C}$) with different hidden activation sizes and activation functions (ReLU and sigmoid).

$$csv'(X) = X \times \sigma(f_c^{1 \times C}(\delta(f_c^{1 \times C/r}(X)))) \quad (10)$$

where r denotes the reduction ratio, which contributes to reducing parameters and computation.

Similar to TAM, the residual connection is also used to reduce the reduction in the feature response value in CAM. The output of CAM is $X_{CAM} = X + csv'(X)$.

Finally, the output of the TCAM (the combination of TAM and CAM) can be represented as:

$$X_{TCAM} = F_{CAM}(F_{TAM}(X)) \quad (11)$$

IV. DATA PREPARATION AND VALIDATION SETUP

In this section, the experimental setup and data preparation are introduced to investigate the effectiveness of the proposed SA1D-CNN for sensor fault diagnosis.

A. EXPERIMENTAL SETUP

To acquire a sensor fault diagnosis dataset of the SIL-IoTs, a series of experiments are designed. The SIL-IoTs system is illustrated in Fig. 6. All experiments are performed indoors to

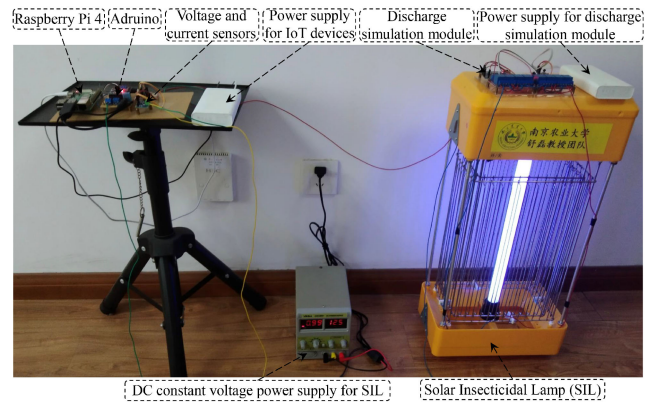


FIG. 6. The SIL-IoTs testedbed. A DC constant voltage power supply provides controllable and stable voltage for the SIL. Different discharge frequencies are realized by a program from the discharge simulation module.

reduce the impact of environmental factors on the experimental results. In our previous work [11], we discovered that the initial voltage of the circuit (fully charged 12 V DC battery) is approximately 12.5 V, and the lamp does not light until the circuit voltage reaches 8 V. The distance between IoT devices and the SIL is set to 50 cm to reduce the impact of electromagnetic interference [9]. The discharge frequencies are set to 0, 0.5, 1, 2, 4, and 10 times per second by the program, where 0 times per second denotes the no discharge condition.

For each experiment, we collect two types of measurements (i.e., voltage value and current value). Every experiment duration is set to 10 min with sampling frequency of 10 times per second; thus, 60,000 samples are obtained for each discharge frequency with 10 voltage segments. To balance the sample number of different conditions, the no discharge samples are expanded five times. Thus, a total of six hundred thousand samples are obtained.

Due to the same feature shape of the testing samples (for inference on the IoT nodes) and training samples, the feature shape of a sample should be small considering the limitation in the IoT nodes. Hence, every 60 consecutive data points are segmented as one sample. The temporal data ($600,000 \times 2$) are reshaped to the input data ($10,000 \times 60 \times 2$). The next step is to simulate sensor faults and label data.

B. SIMULATION OF FOUR SENSOR FAULT TYPES

As mentioned in Section II, four types of sensor faults can be simulated and injected into the original input data to simulate different sensor states. The experiments indicate that the voltage value ranges from 8 V to 12.5 V in normal conditions. The current value ranges from 900 mA to 1000 mA. According to the method in [35], we obtain the dataset, which contains five fault rates (10%, 20%, 30%, 40%, and 50%) and four sensor fault types. To simulate outlier faults, we set the voltage value as a random value from 0 V to 35 V and the current value as a random value from -2000 mA to 2500 mA. To simulate gain faults, we set the value of β as a random value from 1

to 2. The value of ω as a random value ranges from 0 V to 1 V for voltage and from 0 mA to 100 mA for current. We randomly extract continuous samples from normal data and replace them with the sample value when a stuck fault occurs to simulate stuck faults. To simulate offset faults, we set the value of ω as a random value from 3 V to 6 V for voltage and from 300 mA to 600 mA for current signal. The value of η is set the same as the value of ω for the gain fault. The one-hot encoding method is adopted to label data. After excluding some malformed data, the SIL-IoTs fault diagnosis dataset with dimensions of (99449, 60, 2) for features and (99449, 10) for labels is finally obtained.

C. VALIDATION SETUP

The SA1D-CNN is implemented in the TensorFlow library with Python3.8 [37]. Network training is performed on a PC with a Windows 10 operating system, an Intel Core i5-10400 CPU, and 16 GB RAM. Network testing is performed on a Raspberry Pi 4 with Raspbian operating system, a 64-bit quad-core Cortex-A72 processor, and 8 GB RAM. To accelerate the convergence speed, the z -score standardization method is used as shown in (12). In addition, we adopt the cross-entropy loss function, Adam optimization algorithm with a learning rate of 0.001, a batch size of 16, and a dropout equal to 0.3 before classification.

$$x = \frac{x_0 - \mu}{\sigma} \quad (12)$$

where x_0 denotes the original signal, μ denotes the mean value of samples, and σ denotes the standard deviation of the samples.

To evaluate the performance of different models, accuracy, F1-score, model size, testing time, power, and energy consumption measures are applied to evaluate the diagnostic accuracy and lightweight degree of different models with five-fold cross validation in this paper. The diagnostic accuracy metrics, i.e., accuracy and F1-score, were defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (13)$$

$$F_1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (14)$$

where TP, FP, TN, FN denote true positive, false positive, true negative, and false negative samples, respectively. Precision denotes the number of predicted positive samples that are true positive samples shown in (15). As shown in (16), recall denotes the number of positive examples in the dataset that are predicted correctly. The accuracy and F1-score range from zero (the worst) to one (the best).

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (15)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (16)$$

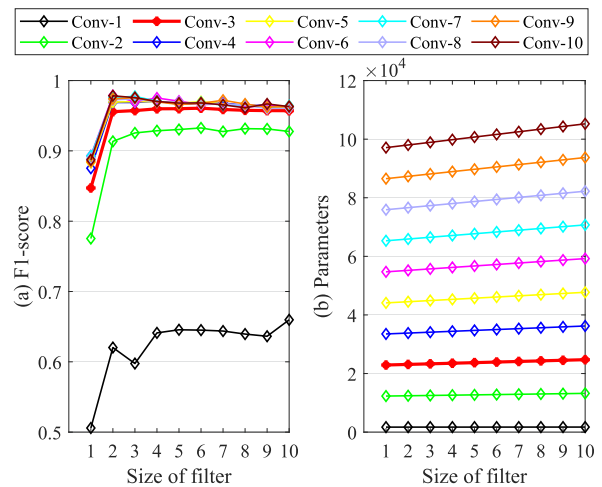


FIG. 7. F1-score and parameters under different numbers of depthwise separable convolution layers and sizes of filter. If the convolution kernel size exceeds the input dimension, "padding=same" is used.

The diagnostic efficiency metrics, i.e., FLOPs (represented in (3)), testing time, and energy consumption are defined as:

$$T = t_2 - t_1 \quad (17)$$

$$P = U \times I \quad (18)$$

where $t_2 - t_1$ denotes the time of inference of a test sample by networks. P denotes the energy consumption of inference of a test sample on a IoT node. U and I denote voltage and current values of Raspberry Pi.

V. PERFORMANCE EVALUATION AND DISCUSSION

A. INFLUENCE OF NEURAL NETWORK STRUCTURE

This section explores the influence of the network structure on the fault diagnosis performance with a fault rate=10%. First, different network depths and widths are discussed. The experiment sets the epochs=30, and the size of filters and the number of depthwise separable convolution layers from 1 to 10. In addition, the number of filters is set to 100. Due to the small sample feature space, the feature maps are padded if the filter size exceeds the input dimension. The F1-score and parameters of the proposed method under different number of depthwise separable layers and sizes of filter are shown in Fig. 7. Obviously, when the number of depthwise separable convolution layers and the size of filters increases from 3 to 10, the performance of the network is not improved (approximately 95% of F1-score). It indicates that the diagnostic accuracy cannot be improved by increasing the depth of the network. Moreover, the model parameters increase from 1712 to 105,230 with increasing the network depth and filter size. Therefore, for data with low dimensions (e.g., 60x2), increasing the network depth and filter size has little help in improving diagnostic accuracy. To balance the diagnostic accuracy and model parameters, the depthwise separable convolution layer and network filter size are set to 3 and 5, which is the basic structure of SA1D-CNN. The details of four basic

TABLE 2 Network Structure and Parameter of Four Basic Models Without Attention Modules

SA1D-CNN		WDCNN		LSTM		DNN	
Layer	Parameter	Layer	Parameter	Layer	Parameter	Layer	Parameter
DSConv 1	input=(*,60,2)	Conv 1	input=(*,60,2)	LSTM 1	input=(*,60,2)	Dense 2	input=(*,60,2)
ReLU	F=100, K=5, S=1	ReLU	F=100, K=10, S=1		H=100	ReLU	U=100
DSConv 2	input=(*,56,100)	Conv 2	input=(*,51,100)	LSTM 2	input=(*,60,100)	Dense 2	input=(*,60,100)
ReLU	F=150, K=5, S=1	ReLU	F=150, K=5, S=1		H=150	ReLU	U=150
DSConv 3	input=(*,52,150)	Conv 3	input=(*,47,150)	LSTM 3	input=(*,60,150)	Dense 3	input=(*,60,150)
ReLU	F=150, K=5, S=1	ReLU	F=150, K=5, S=1		H=150	ReLU	U=150
GAP	input=(*,48,150)	GAP	input=(*,43,150)	GAP	input=(*,60,150)	GAP	input=(*,150)
Dropout	input=(*,150) P=0.3	Dropout	input=(*,150) P=0.3	Dropout	input=(*,150) P=0.3	Dropout	input=(*,150) P=0.3
Dense	input=(*,150) output=(*,10)	Dense	input=(*,150) output=(*,10)	Dense	input=(*,150) output=(*,10)	Dense	input=(*,150) output=(*,10)

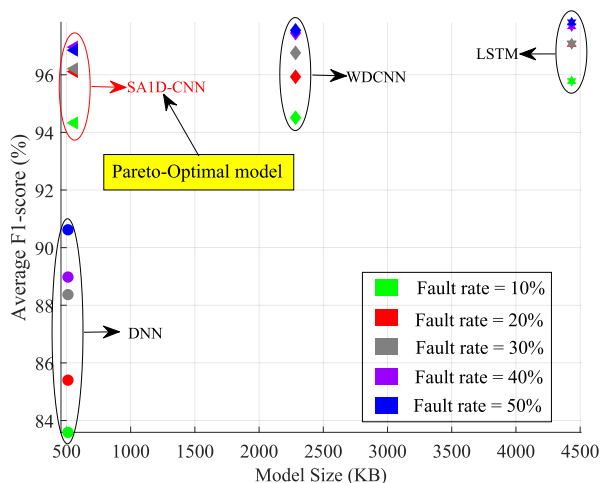


FIG. 8. Performance evaluation and model size of four networks. The results indicate that SA1D-CNN is the Pareto-Optimal model, which has great performance in both detection accuracy and model size.

models without attention modules are shown in Table 2, where DSConv and GAP denote depthwise separable convolution layer and global average pooling layer. F, K, S, H, and U denote number of filter, kernel size, step, number of hidden, and number of unit, respectively.

Second, different baseline networks are compared with SA1D-CNN, i.e., deep convolutional neural networks with wide first-layer kernels (WDCNN [20]), LSTM, and DNN. The experiments set the same numbers of filters and convolution layers for each network. According to [20], the first convolution layer of WDCNN requires a larger filter size. Therefore, the filter size of the first layer is set to 10 for the WDCNN.

Fig. 8 illustrates the performance and model size of four networks, where the Pareto-optimal model denotes an ideal ideal resource allocation state [38]. In this paper, the model size and F1-source are used to estimate the Pareto-optimal model. From these results it is clear that DNN has fewer

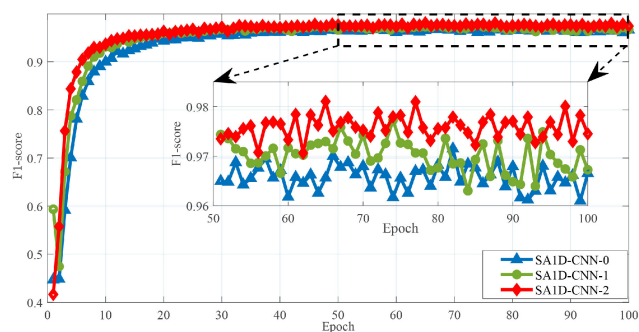


FIG. 9. Performance evaluation of different numbers of TCAM in SA1D-CNN under a fault rate=10%, where the ending number in the model name represents the number of TCAM, e.g., the SA1D-CNN-0 denotes zero TCAM used in SA1D-CNN.

parameters, but its diagnostic accuracy is unacceptable. Although the LSTM has better performance in comparison to others, the largest model size (the number of parameter is 4,434 KB, nearly 8 times larger than SA1D-CNN) leads to higher diagnostic latency and larger computing parameters. Overall, SA1D-CNN is the Pareto-optimal model under different fault rates. To improve the diagnostic accuracy of SA1D-CNN, it is critical to enhance the important features and suppress the features that have little effect on the classification performance.

B. INFLUENCE OF THE NUMBER OF THE TCAM

This section discusses the influence of the number of TCAMs in SA1D-CNN under a fault rate=10%. Fig. 9 shows the F1-score for the three varieties of the SA1D-CNN. Although increasing attention modules increases the model parameters, the F1-score of SA1D-CNN-2 is improved by 1.15% compared with that of SA1D-CNN-0. Therefore, a small increase in the number of parameters is acceptable. In addition, the SA1D-CNN-2 has a faster convergence speed than that of others, which indicates that TCAM can improve the feature extracting ability of SA1D-CNN.

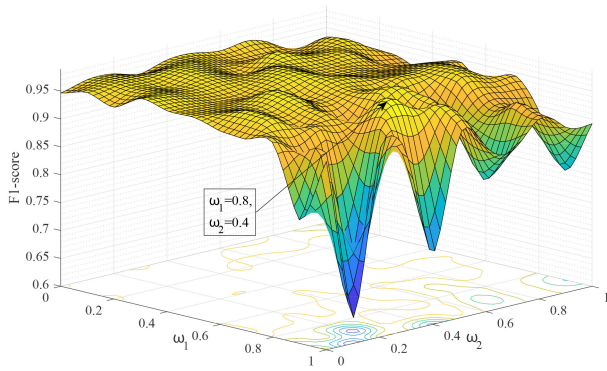


FIG. 10. Performance evaluation of different values of w_1 and w_2 for the TCAM under fault rate=10%.

TABLE 3 Results of the TAM, CAM, and the TCAM Based SA1D-CNN

Fault rate	Metrics	TAM	CAM	TCAM
10%	Accuracy	99.74±0.07	99.7±0.05	99.83±0.03
	F1-score	95.92±0.53	95.09±0.55	96.82±0.49
20%	Accuracy	99.87±0.04	99.86±0.02	99.9±0.03
	F1-score	97.09±0.53	96.39±0.41	97.31±0.51
30%	Accuracy	99.91±0.02	99.9±0.04	99.91±0.04
	F1-score	97.49±0.41	97.24±0.32	97.53±0.28
40%	Accuracy	99.93±0.03	99.9±0.06	99.93±0.02
	F1-score	97.99±0.39	97.45±0.35	98.15±0.42
50%	Accuracy	99.94±0.02	99.91±0.22	99.94±0.01
	F1-score	98.09±0.38	97.62±0.33	98.19±0.31

C. EFFECTIVENESS OF THE TCAM

In this section, optimal weights of the TCAM are selected. Afterward, the performance of SA1D-CNN with TAM, CAM, and TCAM are discussed. In addition, the effectiveness of the TCAM is compared with four state-of-the-art attention modules in SA1D-CNN.

First, the performance of the TCAM under different values of w_1 and w_2 is discussed. The experiment sets fault rate=10%, epoch=30, and two TCAMs. According to Fig. 10, the values of w_1 and w_2 have a great impact on F1-score. The F1-score of the optimal weights ($w_1 = 0.8$, $w_2 = 0.4$) is improved by 1.08% compared with the initial weights ($w_1 = 0.5$, $w_2 = 0.5$), which indicates that extracted features should not be directly added together.

Second, the validities of the TAM, CAM, and TCAM are verified under different fault rates. As summarized in Table 3, slightly superior results are achieved with the TCAM-based SA1D-CNN. In addition, it is worth discussing the interesting fact that the TAM-based SA1D-CNN achieves higher diagnostic accuracy than the CAM-based SA1D-CNN. We speculate that this might be due to the high correlation between temporal changes and fault states. Due to the differences between 1D signals and 2D images (e.g., cyclicity and continuity), 1D-CNN will have better diagnostic accuracy if the model focuses more on the time-serial association rule.

TABLE 4 Network Structure and Parameter of Five Attention Modules Based SA1D-CNN

SENet	Res-SENet	CBAM	JAM	TCAM
Layer	Layer	Layer	Layer	Layer
DSCConv 1	DSCConv 1	DSCConv 1	DSCConv 1	DSCConv 1
ReLU	ReLU	ReLU	ReLU	ReLU
SE 1	Res-SE 1	CBAM 1	JAM 1	TCAM 1
DSCConv 2	DSCConv 2	DSCConv 2	DSCConv 2	DSCConv 2
ReLU	ReLU	ReLU	ReLU	ReLU
SE 2	Res-SE 2	CBAM 2	JAM 2	TCAM 2
DSCConv 3	DSCConv 3	DSCConv 3	DSCConv 3	DSCConv 3
ReLU	ReLU	ReLU	ReLU	ReLU
GAP	GAP	GAP	GAP	GAP
Dropout	Dropout	Dropout	Dropout	Dropout
Dense	Dense	Dense	Dense	Dense

To clearly illustrate the effect of different layers on diagnostic performance, the t-SNE technique [39] is adopted to provide a 2D representation of different layers' output features. As shown in Fig. 11, the distribution of raw signal is turbulent. With more depthwise separable convolution and TCAM layers adopted, the features of ten fault types are gradually separated, where each TCAM layer contributes to extracting the high-level faulty features.

Third, we discuss the performance of the TCAM and four attention module (as baseline attention modules) based SA1D-CNN, including SENet [40], Res-SENet [40], CBAM [41], and JAM [25]. The network structure of SA1D-CNN with different attention modules are shown in Table 4. According to related works, the reduction ratios of SENet, Res-SENet, CBAM, and JAM are set to 8, 8, 8, and 2, respectively. As shown in Fig. 12, the results demonstrate that the TCAM-based SA1D-CNN has the best performance with different fault rates. In addition, an increase in the fault rate has a positive impact on the improvement in diagnostic accuracy due to the balanced samples. However, fault conditions occupy a small part of all conditions in the real world, which indicates that the imbalance of samples is inevitable [42]. Therefore, compared with the others, the nearly 2% improvement of the TCAM with a 10% fault rate demonstrates that the TCAM can effectively improve the fault feature learning ability of SA1D-CNN in the real case.

The testing results of four basic models and five kinds of attention modules based SA1D-CNN are shown in Fig. 13. The results indicate that SA1D-CNN has higher accuracy than WDCNN and DNN. TCAM shows the most significant improvement in accuracy than the others.

D. EFFECTIVENESS OF THE TCAM AND SA1D-CNN ON RASPBERRY PI

In this section, the effectiveness of the TCAM and SA1D-CNN on Raspberry Pi is discussed. The mode size (h5

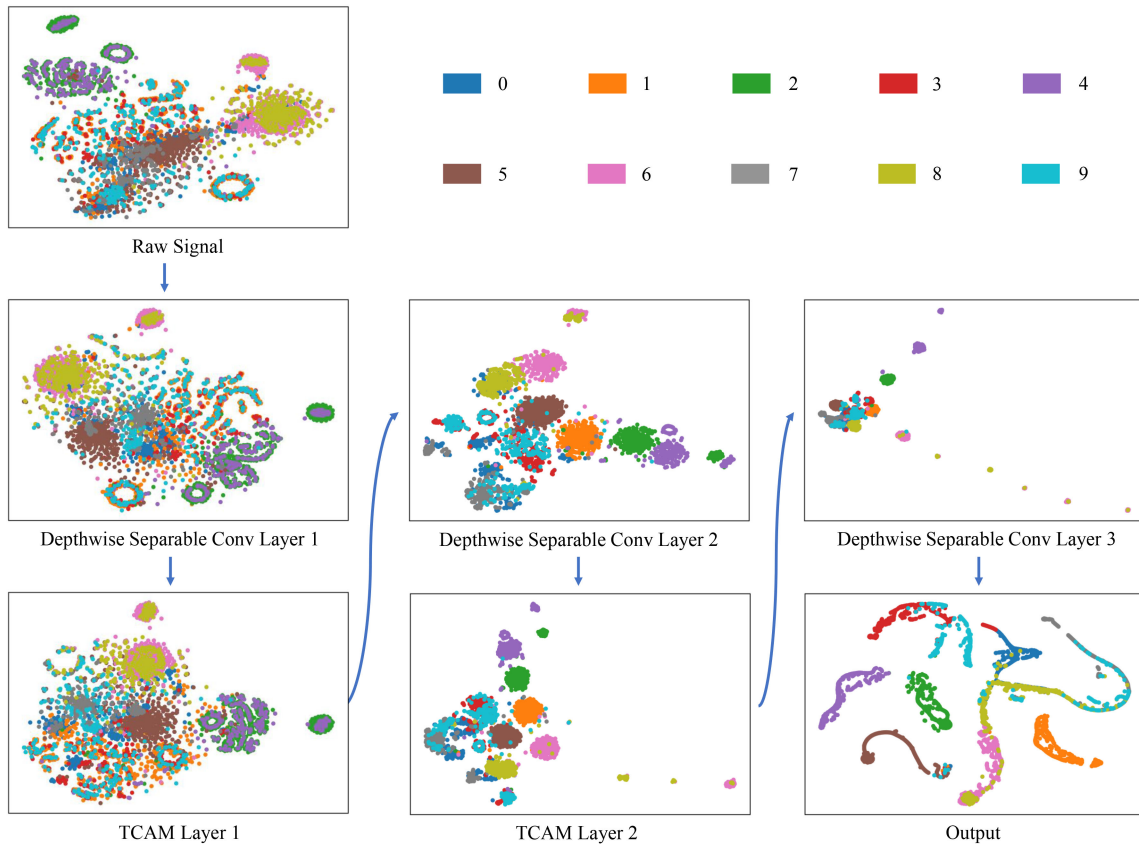


FIG. 11. Visualization of different layer of the SA1D-CNN, where different colors denote different fault labels in Table I.

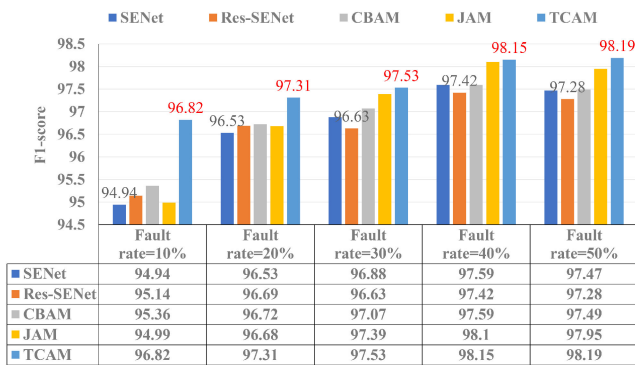


FIG. 12. Average F1-score using different attention modules-based SA1D-CNN.

and TFLite formats), testing time, energy consumption, and FLOPs of four basic models with three layers and SA1D-CNN with different attention modules are summarized in Tables 5 and VI, where larger T indicates lower diagnostic latency. P denotes the average operational power for classifying one sample on Raspberry Pi. FLOPs denotes the computation complexity of fault diagnosis models. The results of Table V indicate that SA1D-CNN has the lowest diagnostic latency and energy consumption compared with other networks. Although the DNN model size is slightly smaller than that

TABLE 5 Results of Four Basic Models Without Attention Modules on Raspberry Pi

	h5 (KB)	TFLite (KB)	T (s)	P (W)	FLOPs (M)
SA1D-CNN	560	171	7.93	4.86	3.9
WDCNN	2,284	752	21.56	5.31	17
LSTM	4,434	1505	288.54	4.11	0.45
DNN	512	162	10.47	4.25	4.56

TABLE 6 Results of Five Attention Modules Based SA1D-CNN on Raspberry Pi

	h5 (KB)	TFLite (KB)	T (s)	P (W)	FLOPs (M)
SENet [40]	698	209	11.76	4.51	3.97
Res-SENet [40]	698	210	12.02	4.49	3.98
CBAM [41]	696	214	22.12	4.12	4.04
JAM [25]	4,669	1,526	44.78	5.25	21.4
TCAM	1,126	351	26.01	4.33	7.65

of SA1D-CNN, fully connected operations of DNN result in higher diagnostic latency. In particular, the LSTM is not suitable for diagnosing faults in IoT nodes due to the high energy consumption and diagnostic latency. The results in Table 6 suggest that the TCAM has a relatively small model size, although the high diagnostic accuracy of the TCAM

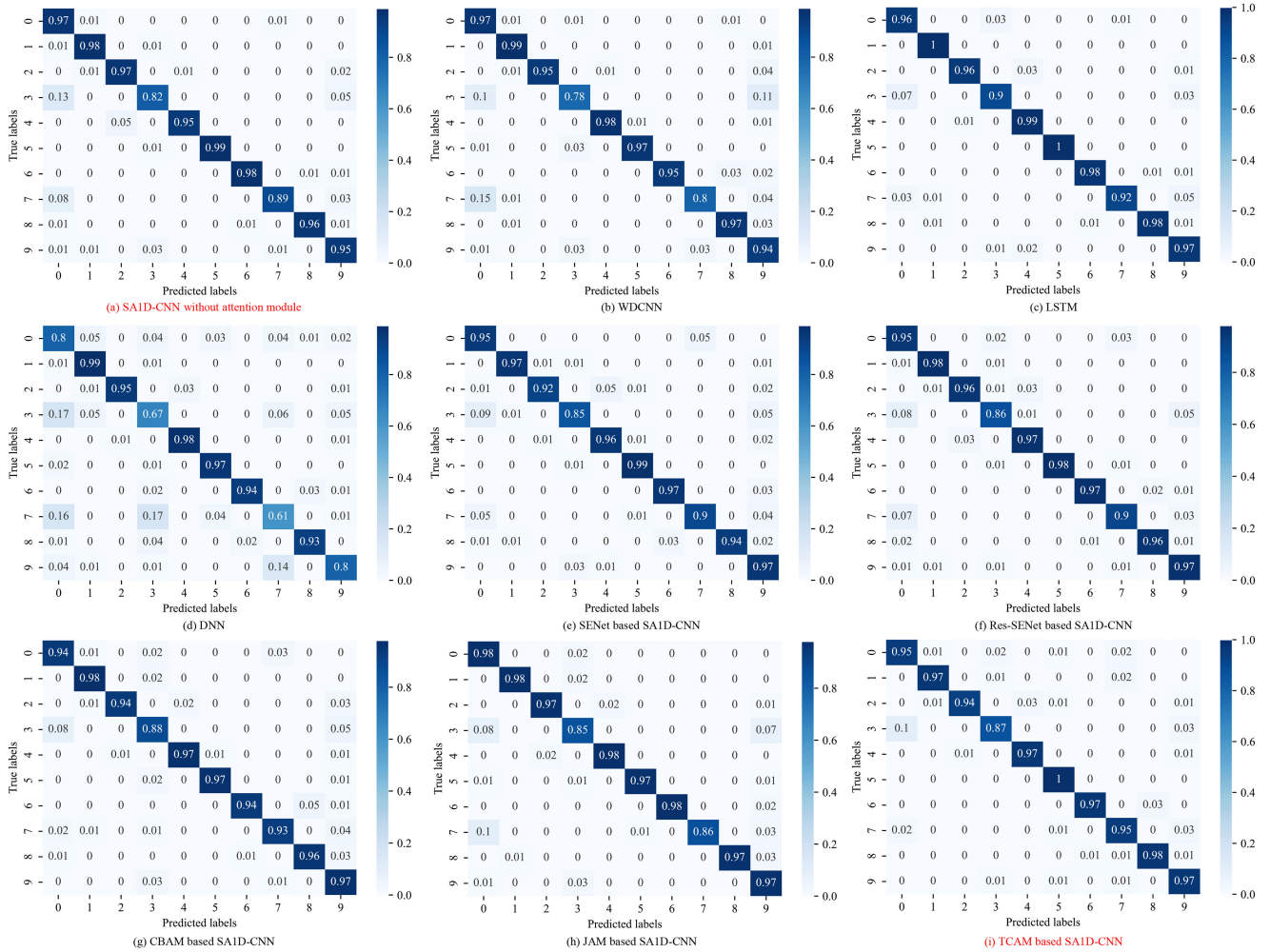


FIG. 13. Testing results of four basic models: (a)-(d) and five kinds of attention modules based SA1D-CNN: (e)-(i), where the proposed methods are shown as red content.

leads to higher complex computation and energy consumption. Nonetheless, the TCAM-based SA1D-CNN represents the best compromise between diagnostic accuracy, model size, and energy consumption.

VI. CONCLUSION

High voltage discharge released by the SIL results in fluctuations in the voltage and current measurements, which should be distinguished from sensor faults. In this study, a set of experiments and fault simulations are designed to simulate sensor faults and generate sensor fault datasets in the SIL-IoTs system. A separable convolution-based lightweight 1D-CNN architecture is developed to diagnose sensor faults with a reduced data transmission, which is implemented and validated in SIL-IoTs nodes. Furthermore, the time and channel attention module is presented to extract high-level fault-related temporal signals and feature maps. The results demonstrated the effectiveness of the proposed method in diagnosing sensor faults of SIL-IoTs with reasonable model size and diagnostic delay. The proposed method can be adopted in related scenarios with resource constraint devices.

The limitation of the proposed method lies in the lack of priority for a variety of faults. Serious faults should be detected at the earliest stage. A multi-grade fault diagnosis scheme may be developed to overcome this limitation by aggregating multiple diagnostic models with different targets. In the future, we will study hierarchical fault diagnosis schemes by evaluating fault severity and online training for the self-adaptation of online fault diagnosis methods.

REFERENCES

- [1] O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, and X. Wang, "Internet of Things for the future of smart agriculture: A comprehensive survey of emerging technologies," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 4, pp. 718–752, Apr. 2021, doi: [10.1109/JAS.2021.1003925](https://doi.org/10.1109/JAS.2021.1003925).
- [2] U. Saeed, S. Ullah Jan, Y.-D. Lee, and I. Koo, "Machine learning-based real-time sensor drift fault detection using Raspberry PI," in *Proc. Int. Conf. Electron., Inf., Commun.*, 2020, pp. 1–7, doi: [10.1109/ICEIC49074.2020.9102342](https://doi.org/10.1109/ICEIC49074.2020.9102342).
- [3] E. Mousavinejad, X. Ge, Q.-L. Han, T. J. Lim, and L. Vlacic, "An ellipsoidal set-membership approach to distributed joint state and sensor fault estimation of autonomous ground vehicles," *IEEE/CAA J. Automatica Sinica*, vol. 8, no. 6, pp. 1107–1118, Jun. 2021, doi: [10.1109/JAS.2021.1004015](https://doi.org/10.1109/JAS.2021.1004015).

- [4] Z. Zhang, H. Li, C. Wu, and Q. Zhou, "Finite frequency fuzzy H_∞ control for uncertain active suspension systems with sensor failure," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 4, pp. 777–786, Jul. 2018, doi: [10.1109/JAS.2018.7511132](https://doi.org/10.1109/JAS.2018.7511132).
- [5] P. M. Papadopoulos, V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Scalable distributed sensor fault diagnosis for smart buildings," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 3, pp. 638–655, May 2020, doi: [10.1109/JIAS.2020.1003123](https://doi.org/10.1109/JIAS.2020.1003123).
- [6] F. Yang, L. Shu, K. Huang, K. Li, G. Han, and Y. Liu, "A partition-based node deployment strategy in solar insecticidal lamps Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 11, pp. 11223–11237, Nov. 2020, doi: [10.1109/JIOT.2020.2996514](https://doi.org/10.1109/JIOT.2020.2996514).
- [7] F. Yang, L. Shu, Y. Yang, G. Han, S. Pearson, and K. Li, "Optimal deployment of solar insecticidal lamps over constrained locations in mixed-crop farmlands," *IEEE Internet Things J.*, vol. 8, no. 16, pp. 13095–13114, Aug. 2021, doi: [10.1109/JIOT.2021.3064043](https://doi.org/10.1109/JIOT.2021.3064043).
- [8] X. Guo et al., "Two-hop energy consumption balanced routing algorithm for solar insecticidal lamp Internet of Things," *Sensors*, vol. 22, no. 1, 2022, Art. no. 154, doi: [10.3390/s22010154](https://doi.org/10.3390/s22010154).
- [9] K. Huang, K. Li, L. Shu, X. Yang, T. Gordon, and X. Wang, "High voltage discharge exhibits severe effect on ZigBee-based device in solar insecticidal lamps Internet of Things," *IEEE Wireless Commun.*, vol. 27, no. 6, pp. 140–145, Dec. 2020, doi: [10.1109/MWC.001.2000082](https://doi.org/10.1109/MWC.001.2000082).
- [10] K. Huang, K. Li, L. Shu, and X. Yang, "Demo abstract: High voltage discharge exhibits severe effect on ZigBee-based device in solar insecticidal lamps Internet of Things," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2020, pp. 1266–1267, doi: [10.1109/INFOCOMWKSHP50562.2020.9162736](https://doi.org/10.1109/INFOCOMWKSHP50562.2020.9162736).
- [11] X. Yang, L. Shu, K. Huang, K. Li, and H. Yao, "Poster abstract: Insecticidal performance simulation of solar insecticidal lamps Internet of Things using the number of falling edge trigger," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2021, pp. 1–2, doi: [10.1109/INFOCOMWKSHP51825.2021.9484570](https://doi.org/10.1109/INFOCOMWKSHP51825.2021.9484570).
- [12] X. Yang et al., "Characteristics analysis and challenges for fault diagnosis in solar insecticidal lamps Internet of Things," *Smart Agriculture*, vol. 2, no. 2, pp. 17–27, 2020, doi: [10.12133/j.smartag.2020.2.2.202005-SA002](https://doi.org/10.12133/j.smartag.2020.2.2.202005-SA002).
- [13] G. Li, Y. Li, H. Chen, and W. Deng, "Fractional-order controller for course-keeping of underactuated surface vessels based on frequency domain specification and improved particle swarm optimization algorithm," *Appl. Sci.*, vol. 12, no. 6, 2022, Art. no. 3139, doi: [10.3390/app12063139](https://doi.org/10.3390/app12063139). [Online]. Available: <https://www.mdpi.com/2076-3417/12/6/3139>
- [14] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots," *Appl. Sci.*, vol. 11, no. 23, 2021, doi: [10.3390/app112311202](https://doi.org/10.3390/app112311202).
- [15] M. Van, S. S. Ge, and D. Ceglarek, "Fault estimation and accommodation for virtual sensor bias fault in image-based visual servoing using particle filter," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1312–1322, Apr. 2018, doi: [10.1109/TII.2017.2723930](https://doi.org/10.1109/TII.2017.2723930).
- [16] R. Xiong, Q. Yu, W. Shen, C. Lin, and F. Sun, "A sensor fault diagnosis method for a lithium-ion battery pack in electric vehicles," *IEEE Trans. Power Electron.*, vol. 34, no. 10, pp. 9709–9718, Oct. 2019, doi: [10.1109/TPEL.2019.2893622](https://doi.org/10.1109/TPEL.2019.2893622).
- [17] Y. Yu, Y. Zhao, B. Wang, X. Huang, and D. Xu, "Current sensor fault diagnosis and tolerant control for VSI-based induction motor drives," *IEEE Trans. Power Electron.*, vol. 33, no. 5, pp. 4238–4248, May 2018, doi: [10.1109/TPEL.2017.2713482](https://doi.org/10.1109/TPEL.2017.2713482).
- [18] W. Deng, Z. Li, X. Li, H. Chen, and H. Zhao, "Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022, doi: [10.1109/TIM.2022.3159005](https://doi.org/10.1109/TIM.2022.3159005).
- [19] H. Cui, Y. Guan, and H. Chen, "Rolling element fault diagnosis based on VMD and sensitivity MCKD," *IEEE Access*, vol. 9, pp. 120297–120308, 2021, doi: [10.1109/ACCESS.2021.3108972](https://doi.org/10.1109/ACCESS.2021.3108972).
- [20] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, 2017, doi: [10.3390/s17020425](https://doi.org/10.3390/s17020425).
- [21] D. She, M. Jia, and M. Pecht, "Weighted entropy minimization based deep conditional adversarial diagnosis approach under variable working conditions," *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 5, pp. 2440–2450, Oct. 2021, doi: [10.1109/TMECH.2020.3040175](https://doi.org/10.1109/TMECH.2020.3040175).
- [22] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019, doi: [10.1109/ACCESS.2018.2888842](https://doi.org/10.1109/ACCESS.2018.2888842).
- [23] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020, doi: [10.1109/TII.2019.2943898](https://doi.org/10.1109/TII.2019.2943898).
- [24] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019, doi: [10.1109/TIE.2018.2877090](https://doi.org/10.1109/TIE.2018.2877090).
- [25] H. Wang, Z. Liu, D. Peng, and Y. Qin, "Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5735–5745, Sep. 2020, doi: [10.1109/TII.2019.2955540](https://doi.org/10.1109/TII.2019.2955540).
- [26] D. Shi, Y. Ye, M. Gillwald, and M. Hecht, "Designing a lightweight 1D convolutional neural network with bayesian optimization for wheel flat detection using carbody accelerations," *Int. J. Rail Transp.*, vol. 9, no. 4, pp. 311–341, 2021.
- [27] F. Iandola and K. Keutzer, "Small neural nets are beautiful: Enabling embedded systems with small deep-neural-network architectures," *Proc. 12th IEEE/ACM/IFIP Int. Conf. Hardware/Software Co-design Syst. Synthesis Companion*, New York, NY, USA: Association for Computing Machinery, 2017, doi: [10.1145/3125502.3125606](https://doi.org/10.1145/3125502.3125606).
- [28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1800–1807, doi: [10.1109/CVPR.2017.195](https://doi.org/10.1109/CVPR.2017.195).
- [29] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [30] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6848–6856, doi: [10.1109/CVPR.2018.00716](https://doi.org/10.1109/CVPR.2018.00716).
- [31] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520, doi: [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474).
- [32] A. Howard et al., "Searching for mobilenetv3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324, doi: [10.1109/ICCV.2019.00140](https://doi.org/10.1109/ICCV.2019.00140).
- [33] S. Ma, W. Liu, W. Cai, Z. Shang, and G. Liu, "Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions," *IEEE Access*, vol. 7, pp. 57023–57036, 2019, doi: [10.1109/ACCESS.2019.2912072](https://doi.org/10.1109/ACCESS.2019.2912072).
- [34] B. Wang, Y. Lei, N. Li, and T. Yan, "Deep separable convolutional network for remaining useful life prediction of machinery," *Mech. Syst. Signal Process.*, vol. 134, 2019, Art. no. 106330.
- [35] X. Fu, Y. Wang, W. Li, Y. Yang, and O. Postolache, "Lightweight fault detection strategy for wireless sensor networks based on trend correlation," *IEEE Access*, vol. 9, pp. 9073–9083, 2021, doi: [10.1109/ACCESS.2021.3049837](https://doi.org/10.1109/ACCESS.2021.3049837).
- [36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31th AAAI Conf. Artif. Intell.*, San Francisco, California, USA: AAAI Press, 2017, pp. 4278–4284.
- [37] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016. [Online]. Available: <https://arxiv.org/abs/1603.04467>
- [38] W. Deng et al., "An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems," *Inf. Sci.*, vol. 585, pp. 441–453, 2022, doi: [10.1016/j.ins.2021.11.052](https://doi.org/10.1016/j.ins.2021.11.052).
- [39] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [40] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: [10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372).
- [41] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. IEEE Eur. Conf. Comput. Vis.*, V. M. Ferrari, C. Hebert Sminchisescu, and Y. Weiss, Eds., Cham: Springer, 2018, pp. 3–19.
- [42] X. Gu, Y. Zhao, G. Yang, and L. Li, "An imbalance modified convolutional neural network with incremental learning for chemical fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3630–3639, Jun. 2022, doi: [10.1109/TII.2021.3112988](https://doi.org/10.1109/TII.2021.3112988).



XING YANG (Graduate Student Member, IEEE) received the M.S. degree in control engineering from the Nanjing University of Information Science and Technology, Nanjing, China, in 2018. He is currently working toward the Ph.D. degree in agricultural electrification and automation from the College of Engineering, Nanjing Agricultural University, Nanjing, China.

Since November 2021, he has been a Visiting Ph.D. Student with the University of Lincoln, Lincoln, U.K.. He has authored or coauthored papers

in several prestigious journals/conferences, such as IEEE WIRELESS COMMUNICATIONS, IEEE/CAA JOURNAL OF AUTOMATICA SINICA, and IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. His research interests include fault diagnosis in wireless sensor networks, agricultural Internet of Things reliability, and machine learning algorithm.



ZHIQIANG HUO received the B.Sc. and M.Sc. degrees from the School of Information Engineering, University of Geosciences, Beijing, China, in 2013 and 2016, respectively, and the Ph.D. degree from the University of Lincoln, Lincoln, U.K., in 2020. He is currently a Research Associate with the Institutional Research Information System, University College London (UCL), London, U.K. He is also a member of CHIMERA at UCL, which dedicates to improving the healthcare and outcomes of critically ill patients. He has authored or coauthored more than 30 publications in peer-reviewed IEEE leading journals and conferences, such as IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, IEEE NETWORK, and IEEE ACCESS. His research interests include time-series complexity analysis, data mining, machine learning, and artificial intelligence.

thored more than 30 publications in peer-reviewed IEEE leading journals and conferences, such as IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, IEEE NETWORK, and IEEE ACCESS. His research interests include time-series complexity analysis, data mining, machine learning, and artificial intelligence.



LEI SHU (Senior Member, IEEE) received the B.S. degree in computer science from South Central University for Nationalities, Wuhan, China, in 2002, the M.S. degree in computer engineering from Kyung Hee University, Seoul, South Korea, in 2005, and the Ph.D. degree in computer engineering from the Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland, in 2010.

Till 2012, he was a Specially Assigned Researcher with the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. He is currently a Distinguished Professor with Nanjing Agricultural University, Nanjing, China, and a Lincoln Professor with the University of Lincoln, Lincoln, U.K. He is also the Director of the NAU-Lincoln Joint Research Center of Intelligent Engineering, Nanjing, China. Currently, he is serving as Chair of IEEE Industrial Electronics Society – TC on Cloud and Wireless Systems for Industrial Applications. In addition, he is serving as Specialty Chief Editor of Frontiers in Plant Science (Sustainable and Intelligent Phytoprotection).

His current research interests include wireless sensor networks and Internet of Things. His H-index is 66 and i10-index is 277 in Google Citation, total citation of his papers is more than 16738 times.



YU ZHANG (Member, IEEE) received the B.Eng. degree in engineering mechanics from Tongji University, Shanghai, China, in 2004, and the Ph.D. degree in civil engineering from the University of Nottingham, Nottingham, U.K., in 2011.

She is currently a Senior Lecturer of Digital Engineering with Loughborough University, Loughborough, U.K. She held several academic positions, including a Research Fellow, Lecturer, and Senior Lecture with the School of Engineering, University of Lincoln, Lincoln, U.K., from 2011 to 2019. In 2019, she joined the Department of Aeronautical and Automotive Engineering, Loughborough University. Her main research interests include condition monitoring, novelty detection, fault diagnosis and the development of data analysis and machine learning algorithms and especially for industrial applications.



KAILIANG LI received the B.S. degree from the Guangdong University of Petrochemical Technology, Maoming, China, in 2015. He is currently a Research Assistant with the College of Engineering, Nanjing Agricultural University, Nanjing, China. He has authored or coauthored more than 10 papers in related conferences, and journals in the areas of sensor networks and Internet of Things. His research interests include wireless sensor networks and Internet of Things.