Novel Data-Driven Approach Based on Capsule Network for Intelligent Multi-Fault Detection in Electric Motors

Chen, Jianjun; Hu, Weihao; Cao, Di; Zhang, Man; Huang, Qi; Chen, Zhe; Blaabjerg, Frede

Published in:
IEEE Transactions on Energy Conversion

DOI (link to publication from Publisher):
10.1109/TEC.2020.3046642

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Novel Data-Driven Approach Based on Capsule Network for Intelligent Multi-Fault Detection in Electric Motors

Jianjun Chen, Weihao Hu, *Senior Member, IEEE,* Di Cao, Man Zhang, Qi Huang, *Senior Member, IEEE,* Zhe Chen, *Fellow, IEEE*, and Frede Blaabjerg, *Fellow, IEEE*

***Abstract*—With the steady development of technology, electric motors (EMs) have become one of the most important components in modern industry. To ensure stable industrial production, detecting and classifying the EM faults is crucial. A novel intelligent deep-learning-based multi-fault detection method for EMs under varying working conditions is proposed in this paper. This method involves two steps: first, a 2D convolution network without pooling layer is proposed to extract features from raw EM data. In addition, a long short-term memory (LSTM) network is applied to extract the fault features for comparison. Second, a capsule network (Caps-Net) based on a dynamic routing algorithm is used as a classifier to realize intelligent multi-fault detection and improve the generalization performance of the proposed model. The proposed method is applicable to raw physical signals of EMs, which improves the overall efficiency of the fault detection. Moreover, the proposed method has a strong generalization ability. The simulation results demonstrate that the proposed approach can achieve higher accuracy than various benchmark methods. Moreover, its fault detection accuracy is higher than those of other state-of-the-art models under two working conditions, in which the load type and size of the EM are changed, respectively.**

***Index Terms*—Electric machine (EM), Intelligent multi-fault detection, Capsule network, Date-Driven, Current signal.**

I. Introduction

IN modern industrial manufacturing, electric motors (EMs) are widely applied to pumps, fans, machine tools, compressors, mechanical arms, and other devices. With the development of renewable energies, EMs play an important role in the fields of wind power generation and electric vehicles. As one of the most basic members of modern industry, EMs usually operate under severe and varying working conditions. Consequently, these machines suffer frequently from various emergency faults, which may cause malfunction and huge economic losses. In addition, the operation and maintenance costs increase with the wide application of EMs. To ensure the economic and reliable operation of the mechanical equipment and industrial system and to increase the operational effectiveness of complex and expensive mechatronic systems [1]-[3] the monitoring of the working states of EMs has become increasingly important [4].

A. Related Works

To mitigate the impact of faults of EMs on industrial systems and improve the operation efficiency of industrial production, an effective method for EM condition monitoring and fault diagnosis must be established [5], [6]. Recently, with the rapid development of artificial intelligence, machine learning-based approaches have been applied in many fields, particularly for feature learning from big data [7], [8]. For example, Google's AlphaGo and AlphaGo zero [9] use deep neural networks for cognitive learning from big game data or Go rules and have made a great breakthrough. According to [10], artificial neural networks have been successfully used for the diagnosis of electric drives, and it was stated in [11] that the powerful feature learning ability of data-intensive machine-learning methods can also be used for EM fault detection.

EM faults can be classified into two categories: mechanical and electrical faults [12]. Both can alter the physical signals of EMs [13]. Therefore, realizing fault diagnosis by analyzing the physical signals of EMs is feasible. Traditional EM fault diagnosis is realized by signal processing methods, such as complex matrix operation, [14], Fourier transform [15], and wavelet packet transform (WPT) [16]. In [17], a proposed wavelet transform-based fault detection method was applied for the diagnosis of EM bearing faults. Furthermore, a motor fault detection method based on spectral kurtosis coupled with k-nearest neighbor (k-NN) was proposed in [18]. Although these methods enable the fault diagnosis of EMs, they require manual data feature extraction, which is influenced by personal experience. As a result, the fault detection accuracy of these methods depends largely on personal signal processing knowledge and the ability of the designer.

Several researchers have focused on the operation signals of

This work was supported by the Sichuan Science and Technology Program under Grant 2020JDJQ0037 and 2020YFG0312. (Corresponding author: Weihao Hu)

Jianjun Chen, Weihao Hu, Di Cao, Man Zhang, Qi Huang are with the Power System Wide-area Measurement and Control Sichuan Province Key Laboratory, School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China Chengdu 611731, China (e-mail: chjj@std.uestc.edu.cn; whu@uestc.edu.cn; caodi@std.uestc.edu.cn; zhangman@uestc.edu.cn; hwong@uestc.edu.cn);

Zhe Chen and Frede Blaabjerg are with the Department of Energy Technology, Aalborg University, Aalborg, Denmark (e-mail: zch@et.aau.dk; fbl@et.aau.dk)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

EMs and proposed data-driven-based approaches for the fault detection of EMs. For instance, Chen et al. [19] pointed out that deep learning-based methods can achieve a better performance in fault diagnosis than other traditional methods and human experts. Sai et al. [1] proposed an EM fault detection method that applies three different fault feature learning approaches: dual tree complex WPT (DTCWPT), convolutional neural network (CNN), and recurrent neural network (RNN). Subsequently, they applied the support vector machine (SVM) and dense network as a classifier to realize the incipient faults diagnosis and classification of EMs. In [20], the DTCWPT method was adopted to extract the features of the fault signal and identify the faults with a multiple classifier. However, these methods require complex mathematical calculation models, which increases the computational cost of the fault diagnosis. In addition, they require a prior manual feature extraction process. Therefore, a fully data-driven intelligence method for EM fault detection without human intervention is necessary.

With the fast-growing field of machine learning, many researchers have investigated EM fault detection based on this method. By comparing several different algorithms, such as Bayesian learning, k-NN, bagging, and artificial neural networks, Ignacio et al. [21] presented an experimental comparative evaluation of machine learning techniques for EM rotor fault diagnosis. Based on the analysis of the rotor current signal, Fangzhou et al. [22] applied the SVM for the motor gear tooth fault diagnosis of doubly-fed induction generators. Moreover, a deep belief networks dislocated time-series CNNs (DTS–CNNs) were used for EM bearing fault detection in [23]. An improved CNN model, ResNet, can also be used for EMs fault detection, which has strong feature extraction ability and easy for training [24]. Although the previously mentioned methods enable EM fault diagnosis, the number of faults identified by these methods is limited. Moreover, there are some machine learning-based methods for EM multi-fault diagnosis. In [25], asynchronous induction motor winding fault identification based on combined finite-element and neural networks by analyzing the magnetic signature was proposed, and Wenjun et al. [26] proposed a convolutional discriminative feature learning method for induction motor fault detection, including the diagnosis of, for example, winding faults and broken rotor bars; it applies a back-propagation-based neural network to extract fault features and applies an SVM machine classifier to identify different fault conditions. Furthermore, Siyu et al. [27] developed a deep transfer learning approach for machine fault diagnosis, which can also diagnose different kinds of EM faults. LSTM can also be used for EM fault diagnosis. In [28], R. Sabir et al. proposed a LSTM based method for motor bearing fault detection, which applied LSTM for feature extraction and faults classification. However, when the EM working condition changes, this fault detection model must be retrained.

EM fault detection is a hot topic in research. Although the previously presented methods are of great significance for EM fault detection, they have some limitations. First, owing to the limited generalization ability, these EM fault diagnosis methods are based on constant working conditions (for instance, [25] and [26]). When the working condition of the EM changes, the performance of these methods becomes deteriorated. Second, these methods require massive data for training and complex hand-crafted mathematical operations to preprocess the raw data for feature extraction; for example, Sai et al. [1] and Jinxiu et al. [20] preprocessed data by the DTCWPT method. These methods are time-consuming and increase the computational cost. In addition, data preprocessing relies on personal prior knowledge and experience, and the quality of the extracted features has great influence on the performance of the fault detection model. Third, the number of faults considered for detection and classification is sometimes limited to one or two in some research studies (for example, [21], [22], and [23]); these methods only consider different bearing or rotor faults, which limits the application of the fault detection model in other fault scenarios. Therefore, a method with strong generalization ability, which can detect different kinds of EM faults with fewer data under varying working conditions, must be established.

B. Contributions

To overcome the previously mentioned disadvantages, a novel fault detection method based on the convolution capsule network (CCaps-Net) is proposed, which is inspired by the dynamic routing between capsules [29], [30]. First, a 2D convolution network for extracting raw data features is proposed. Second, an improved Caps-Net based on a dynamic routing algorithm [30] is used as a classifier for intelligent multi-fault detection, which improves the generalization performance of the fault detection model. Finally, the CCaps-Net intelligent multi-fault detection model is tested under varying conditions. By changing the EM load size and type, two different cross validation studies are conducted. Compared with other state-of-the-art machine learning methods, the EM fault detection model proposed in this paper achieves a better performance. In summary, the main contributions of this paper are as follows:

1) A novel framework for intelligent fault detection and classification of EMs by analyzing the current signal is proposed. In addition, a 2D convolution network for extracting the features of raw data is introduced, and a Caps-Net-based classifier is used for the fault diagnosis and to improve the generalization ability.
2) The proposed method is tested under different working conditions. Two cross validation tests are conducted to demonstrate the generalization ability of the proposed approach. To the best of the authors' knowledge, this is the first study in which the CCaps-Net framework is applied in the field of multi-fault detection of EMs under varying working conditions.
3) Compared with various machine learning-based approaches, the proposed method has higher fault detection accuracy and stronger generalization ability. Moreover, it does not need to be retrained unlike traditional fault detection models under varying working conditions.

The remainder of this paper is structured as follows: the fundamental theoretical basis of the proposed model is presented in Section II. Section III describes the overall framework of the proposed models for EM faults detection, and the case study results and analysis are presented in Section IV. Finally, Section V presents the conclusions.

II. Fundamental Theoretical Basis

A. CNN and LSTM

Convolutional computation models have been widely applied to big data for feature learning, and CNNs are now widely used in various fields, such as video processing and speech recognition [31], etc. The architecture of the traditional CNN is shown in Fig. 1; there are three main layers: the convolution, pooling, and full connection layers. By the convolution and pooling operation of the input data, the feature maps can be obtained by the full connection layer for further processing.

For comparison, the long short-term memory (LSTM) is applied to extract the features of raw time series data of an EM. The LSTM is an improved RNN with better performance. The powerful learning ability of LSTM for time series data mitigates the limitations of traditional hand-crafted methods sufficiently. The inner architecture of an LSTM is shown in Fig. 2. The three gating units and cell are the most important parts of the LSTM. Compared with a general neural network, LSTMs are capable of learning long-term dependencies.

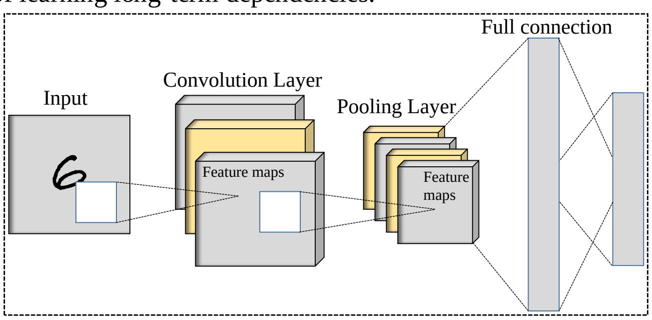


Fig. 1. Architecture of traditional convolution neural network (CNN).

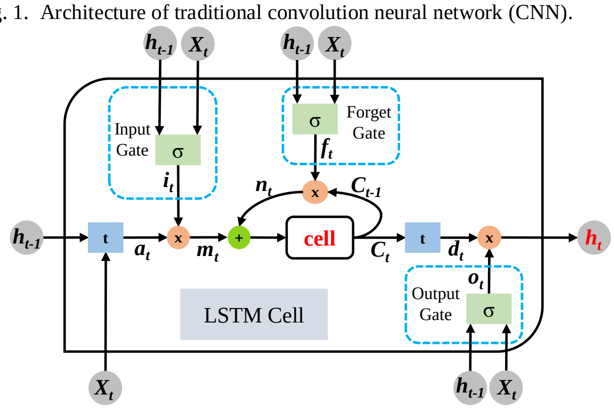


Fig. 2. Inner architecture of long short-term memory cell (LSTM).

The LSTM cell state, C_t, can be described as follows:

$$C_t = m_t + n_t, \tag{1}$$

where m_t and n_t are the intermediate states of the LSTM:

$$m_t = i_t \cdot a_t, \tag{2}$$

$$n_t = f_t \cdot C_{t-1}, \tag{3}$$

where i_t and f_t are the input and forget control gates of the LSTM, respectively; their values are between 0 and 1; a_t is the input state of the LSTM, and C_{t-1} represents the last cell state:

$$i_t = \sigma(W_i \begin{bmatrix} h_{t-1} \\ X_t \end{bmatrix} + b_i), \tag{4}$$

$$f_t = \sigma(W_f \begin{bmatrix} h_{t-1} \\ X_t \end{bmatrix} + b_f), \tag{5}$$

$$a_t = tanh(W_{t1} \begin{bmatrix} h_{t-1} \\ X_t \end{bmatrix} + b_{t1}), \tag{6}$$

where W_i, W_f, and W_{t1} are the weight matrices; b_i, b_f, and b_{t1} are the biases of the neural network; σ is the sigmoid activation function, and $tanh$ is the tanh activation function. The role of f_t is to forget the information in the LSTM cell selectively, and the function of i_t is to record new information into the LSTM cell selectively.

Moreover, the remaining intermediate state d_t can be represented as follows:

$$d_t = tanh(C_t). \tag{7}$$

Finally, the output of the LSTM, h_t, is as follows:

$$h_t = o_t \cdot d_t, \tag{8}$$

where o_t is the output gate of the LSTM (value between 0 and 1). The function of o_t is to determine the output information:

$$o_t = \sigma(W_o \begin{bmatrix} h_{t-1} \\ X_t \end{bmatrix} + b_o). \tag{9}$$

In this study, the feature output of the LSTM is used as the feature input of the fault classifier.

B. Caps-Net

The traditional CNN includes the important pooling layer. However, the max-pooling operation allows neurons in one layer to consider only the most active information in a local pool of the next layer, which causes CNNs to lose important information in the feature transformation. Caps-Net is used to overcome the deficiency of the pooling layer, which replaces the scalar-output features of CNNs with vector output capsules and the max-pooling operation with a dynamic routing algorithm. The architecture of a Caps-Net is shown in Fig. 3.

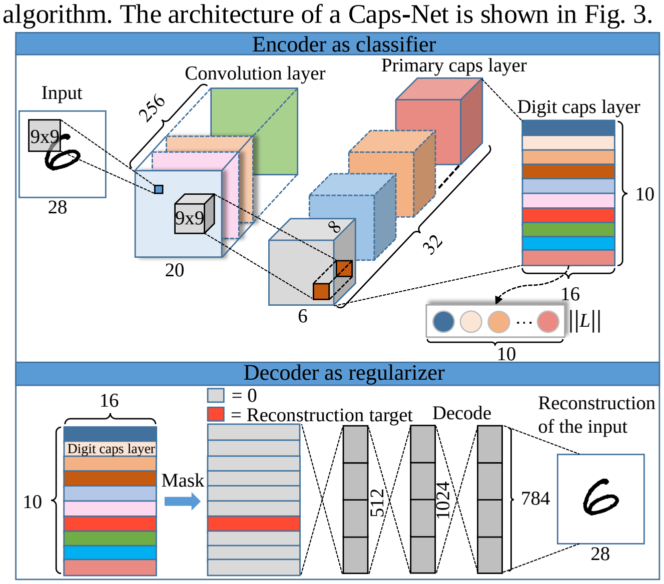


Fig. 3. Architecture of original Caps-Net.

The original Caps-Net consists of two parts: classifier and decoder. The architecture of a capsule classifier has three parts: convolution, primary capsule, and digit capsule layers. Unlike traditional intelligent classifiers, Caps-Net uses the length of the vector in the digit capsule layer to represent the probability of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

the existence of the entity and the orientation of the vector to present the properties of the entity [30]. In addition, the function of the decoder is to reconstruct the input data. Finally, the total loss of the overall Caps-Net includes two parts: the weighted sum of the classification and reconstruction losses.

III. Proposed Framework

To establish a strong generalization method for EM fault diagnosis, a novel intelligent approach (CCaps-Net) and a comparison method (LSTM–Caps-Net) are introduced in this section. The architecture of the proposed EM fault detection model is shown in Fig. 4. The proposed method can be classified into two parts: raw data feature-learning layer based on convolution operation without pooling layer or LSTM and capsule feature classifier layer based on proposed Caps-Net.

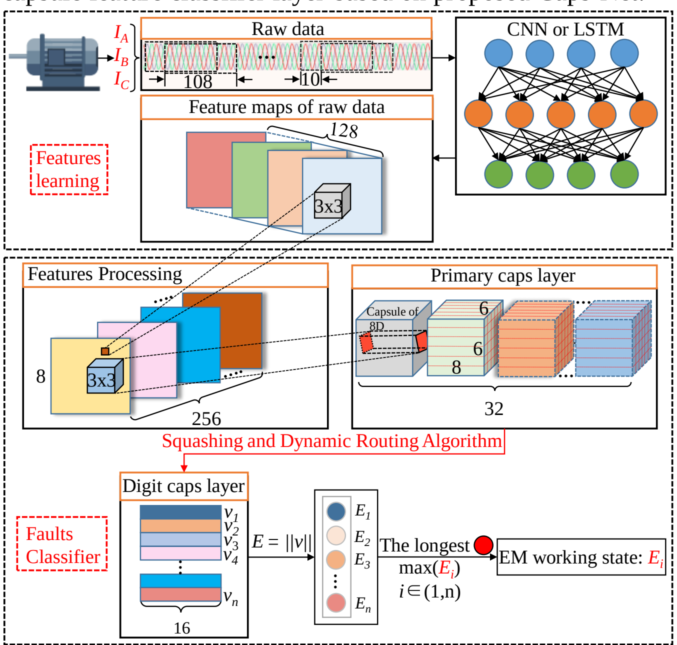


Fig. 4. Architecture of EM multi-fault detection model.

A. Learning of Raw Data Features

In this study, a 9×9 convolution kernel is applied to extract fault features. In the pooling layer of traditional CNNs, features obtained by the convolution operation are replaced by their average or maximum, which may lead to loss of important information. Consequently, the pooling layer of the CNN is removed in this study to ensure the integrity of features. The convolution kernel for EM data feature learning can be mathematically expressed as follows:

$$f_c = act(W_{EM} M_t + b_{EM}), \tag{10}$$

where f_c is used to store the output features of the convolution kernel, which can be applied to further analyses; act represents the activation function; W_{EM} and b_{EM} are the weight matrix and bias, respectively, and M_t is the input EM signal.

According to Section II, the feature learning process of motor data is conducted through the LSTM network. The feature output, h_t, is equal to that in Equation (8), whereas the input of the LSTM network is the motor operation data M_t.

In this study, the output feature shapes of the convolution kernel and LSTM are reshaped to [s, 18, 18], where the variable parameter s is the batch size of the raw data. In general, the output of the last time step of the LSTM is used for subsequent processing. By contrast, this proposed model uses the output features of all time steps. The features of CNN and LSTM learning are directly used as the input of the proposed fault classifier model without any additional processing.

B. Fault Classifier

An improved Caps-Net classifier, which is more suitable for EM fault detection, is proposed. In the fault diagnosis, reconstructing the input data is generally not required. When the complexity of the input data increases, the neural network has difficulties to converge owing to the reconstruction branch, which increases the computational cost [29]. Consequently, to enhance the performance of the fault detection model, a more efficient classifier based on the original capsule classifier is proposed in this paper.

As shown in Fig. 4, two 2D CNNs (two 3×3 convolution kernels with stride of 1) are applied after using the feature learning model to map the EM features to higher dimensions. After the convolution with two CNNs, the raw data features are classified into [6×6×32] primary capsules, and each capsule is an 8D vector. Moreover, each capsule in the [6×6] grid is weight-sharing. The calculation of the 8D vector e_i can be expressed as follows:

$$g_1 = \tanh(W_{g1} f_c + b_{g1}), \tag{11}$$

or

$$g_1 = \tanh(W_{g1} h_t + b_{g1}), \tag{12}$$

$$e_i = \tanh(W_{ei} g_1 + b_{ei}). \tag{13}$$

As shown in Fig. 4, the primary caps layer generates a series of 16D feature vectors in the digit caps layer through "squashing" and "dynamic routing" operations. The length of these vectors represents the probability of the existence of the corresponding working states. The non-linear "squashing" function ensures that the length of these vectors of the digit caps layer is compressed to a value between zero and one. Moreover, the squashing function squash (v_j) can be described as follows:

$$v_j = \left[\left\|l_j\right\|^2 / (1 + \left\|l_j\right\|^2)\right] l_j / \left\|l_j\right\|, \tag{14}$$

where v_j is the output vector of the capsule j (j∈(1, n)), n the number of working states of the EM, and l_j the total input vector. From primary to digit caps layer, the total input l_j is a weighted sum of all prediction vectors $\hat{e}_{j|i}$, and the prediction vector is calculated by multiplying the output 8D vector e_i of a capsule in the primary caps layer by a weight matrix W_{ij}. Therefore, l_j and $\hat{e}_{j|i}$ can be calculated as follows:

$$l_j = \sum_i a_{ij} \hat{e}_{j|i}, \tag{15}$$

$$\hat{e}_{j|i} = W_{ij} e_i, \tag{16}$$

where a_{ij} (i∈(1, 6×6×32)) are coupling coefficients that are determined by the operation process softmax(b_{ij}):

$$a_{ij} = \exp(b_{ij}) \,/\, \sum_k \exp(b_{ik}), \tag{17}$$

where b_{ij} are intermediate variables with initial values of zero, which can be updated discriminatively simultaneously such as the weights W_{ij}. Updating b_{ij} works as follows:

$$b_{ij}^r = b_{ij}^{r-1} + \hat{e}_{j|i} \cdot v_j^r, \tag{18}$$

where r denotes the number of iterations of the dynamic routing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

algorithm. The entire previously presented v_j calculation process is called a "dynamic routing algorithm". The procedures of dynamic routing are summarized in Algorithm 1.

Algorithm 1: Dynamic routing algorithm.

Procedure routing ($\hat{e}_{j|i}$, r, primary caps, digit caps)
  for all capsules i in primary caps layer and capsules j in digit caps layer: set b_{ij} to 0, T to 3.
  for r = 1 : T **do**
    for all capsules i in primary caps layer: set a_i to softmax(b_i);
    for all capsules j in digit caps layer: set l_j to $\sum_i a_{ij}\hat{e}_{j|i}$;
    for all capsules j in digit caps layer: set v_j to squash(v_j);
    for all capsules i in primary caps layer and capsules j in digit caps layer: set b_{ij} to $b_{ij} + \hat{e}_{j|i} \cdot v_j$.
  return v_j

After the operation of the dynamic routing algorithm between the primary and digit caps layers, the proposed model calculates the lengths of the vectors v_j to present the probability that different EM working states exist, and the state corresponding to the longest vector is chosen.

To increase the distances of the interclass and reduce simultaneously intra-class variations of the proposed model, a margin loss function is introduced to optimize the proposed model. The model is trained by minimizing the margin loss function (19).

$$L_k = Y_n \max\left(0, m^+ - \left||v_n|\right|\right)^2 + \lambda(1 - Y_n)\max(0, \left||v_n|\right| - m^-)^2, \tag{19}$$

where Y_n represents the index value of the data sample labels; different values represent different working states of EMs. In this study, m^+, m^-, and λ are set to 0.9, 0.1, and 0.25, respectively [30], and the Adam optimizer with a learning rate of 0.001 is applied to optimize the model and parameter update.

Algorithm 2: Training procedures of proposed model for EM.

Input: three-phase current I_A, I_B, I_C; learning rate lr; batch size b; number of iterations for training t; number of iterations in routing algorithm T;
Output: fault detection accuracy; states of motor predicted by model; real states of motor;
Initialize: initialize the weights and biases of the CNN and LSTM $W_{EM}, W_i, W_f, W_o, W_{t1}, b_{EM}, b_i, b_f, b_o$; initialize the classifier parameters $W_{g1}, W_{ei}, b_{g1}, b_{ei}$ and routing algorithm parameters b_{ij}, W_{ij};
  for iterations in range t **do**
    # obtain motor working state feature:
    $f_c = act(W_{EM}M_t + b_{EM})$ or $h_t = o_t \cdot d_t$;
    # feature processing
    $g_1 = \tanh\left(W_{g1}f_c + b_{g1}\right)$
    or $g_1 = tanh\left(W_{g1}h_t + b_{g1}\right)$;
    # primary caps layer: obtain 8D vectors e_i
    $e_i = \tanh(W_{ei}g_1 + b_{ei})$;
    # digit caps layer: obtain fault feature vectors v_j
    for r = 1 : T **do**
      routing ($\hat{e}_{j|i}$, r, Primary caps, Digit caps)
    return v_j;
    # calculate length of vector v_j
    $S_j = ||v_j||$;
    # determine the longest vector and output the index j, which corresponds to different motor states
    J = tf.argmax(S_j, axis = 1);
  end for
  output the fault detection accuracy and EM working states.

The model is trained by minimizing the margin loss function and optimized by the Adam optimizer.

Algorithm 2 presents the training procedure of the proposed model.

IV. Analysis of Results

The case study and analysis of the results are presented in this section. The numerical simulation and model training are completed on a 64-bit PC with Intel Core i9-7900X CPU of 3.8 GHz, 32 GB RAM, and an RTX2080 Ti GPU with 11 GB VRAM. The proposed model is implemented in the Pycharm software platform with the GPU version Tensorflow.

The proposed method is compared with LSTM–Caps-Net and other state-of-the-art methods: CNN [6], [23], SVM [26], LSTM-ATT [32] and ResNet [33]. More specifically, a two-layer CNN with max-pooling operation model is applied for comparison. The kernel function and parameters c and gamma of the SVM are RBF, 0.9, and 20, respectively. Only the SVM without a combination with other feature extraction methods is applied in this study. The LSTM–ATT model is a combination of the attention mechanism model and a three-layer LSTM. A ResNet with ten residual units is also applied for the performance comparison, each residual unit contains four layers of neurons. It evaluates the performance of the proposed and other state-of-the-art methods based on the average accuracy of the fault identification. The accuracy can be calculated as follow:

$$\mathrm{A} = \sum \frac{n_{dec}}{n_{batch}} \times 100\%, \tag{20}$$

where n_{dec} is the number of EM states correctly detected by the model, n_{batch} represents the number of batches of data input into the neural network in each test round. The detailed hyper-parameters of the proposed method and state-of-the-art methods are summarized in appendix section. The hyper-parameters tuning of the models is based on the Grid Search method [34]. In this paper, we set the learning rate in the range of 0.1~10^{-8} and reduce ten times at each tuning time, which is a general range of neural network parameters tuning. In addition, we set the number of neural layer in the range of 1~5 and increase 1 layer at each tuning time. Similarly, other hyper parameters are also set to a general range. After all the hyper parameters are set to a specific range, these parameters are tuned by Grid Search.

A. EM Working States and Training Dataset Description

The proposed method is applied for a multi-fault diagnosis of a three-phase asynchronous induction motor under varying working conditions. The data of the motor under different working conditions are simulated by a commercial software, *ANSYS Electronics Suite*, which can uniquely simulate the electromagnetic performance of the circuit and system design, and can evaluate the temperature, vibration and other key mechanical effects to help users design innovative electrical and electronic products faster and more economically [35]. The detailed motor parameters are shown in Table I.

Three different types of loads are applied to the motor: constant power (CP), constant torque (CT), and linear torque (LT) loads. In addition, three varying load torques are applied. To ensure operation efficiency of the motor, the loads are 68,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

70, and 72 Nm, respectively. Therefore, there are six different kinds of motor operation datasets for the fault detection model training and testing. In addition, seven induction motor health states are considered, including one normal state and six fault states: inter-turn short circuit (ISC), broken rotor bar (BRB), load missing (LM), open-phase (OP), and rotor dynamic and static eccentricity (RDE and RSE, respectively) states. The health states of EM are simulated by changing the external circuit structure of the EM model in *ANSYS Electronics Suite* software [35]. The descriptions of the seven EM health states are listed in Table II.

TABLE I
MOTOR PARAMETERS

No.	PARAMETER	Setting
1	Rated power	11 kW
2	Rated speed	1458 rpm
3	Rated load torque	75 Nm
4	Rated voltage	380 V
5	Rated frequency	50 Hz
6	Efficiency & power factor	0.85 and 0.86

TABLE II
DESCRIPTION OF SEVEN DIFFERENT WORKING STATES OF EM

Label	Health State	Description
0	Normal	Normal working state
1	ISC	Phase A with inter-turn short circuit of 2 turns
2	BRB	4 broken rotor bars in induction motor, 2 neighboring broken bars, another 2 bars in opposite direction;
3	LM	When the motor works normally for 0.4 s, the load is suddenly missing;
4	OP	When the motor works normally for 0.4 s, phase A experiences a phase failure;
5	RDE	Rotor dynamic eccentricity: the rotation track of the rotor is a circle, and the center of the circle is a point;
6	RSE	Rotor static eccentricity; the center of the rotation track of the rotor is a circle.

The motor current is one of the most commonly used and easily measured values. Different motor fault types can be detected by analyzing the features of the three-phase current (I_A, I_B, I_C) of the motor in different fault states with the neural network model. The sampling frequency of the EM working signal is 1 kHz. There are six EM working conditions in total, and each working condition has seven health states. The length of each time-series data point is 4000. In total, there are 168000 samples. It can be observed in Fig. 4 that a sub-sampling window containing 108 raw data points is collected to be one sample, and there are 10 data points interval between each sub-sampling window. A dataset is a combination of all these samples with different EM health states. In this study, when the training and test data of the model are in the same condition, 80% of the motor operation dataset are selected for training, 10% for validation under training phase, and the remaining 10% for testing. When the training and test data come from different working conditions, 80% of the dataset are selected for model training and 20% for validation under training phase. Under this condition, 10% of samples generated under other loading condition are randomly selected as the test set. Validation is similar to test, but different from the model final test, the purpose of validation is to observe the performance of fault diagnosis model in real time during training, and its results can be used to judge whether to end the training ahead of time, which can save the time cost. In each training round, 450 batches of EM data are sent to the fault detection model. The shape of each batch of EM raw signal input sent to the fault detection model is [108, 3].

B. *Analysis of Multi-fault Detection Result*

First, the performance of different fault detection methods under different load types is compared. The results are shown in Table III. The average test time of each batch of data of CNN is 0.06s, which spends the least time. The proposed method follows closely, which has 0.08s average test time and much less than LSTM-Caps-Net and LSTM-ATT models. Despite the average test time of each batch of data of CNN is 0.02s shorter than the proposed method, the fault detection performance of CCaps–Net is much better than CNN method.

The method proposed in this paper has a better fault detection performance with an average accuracy of $98.39 \pm 0.09\%$. Moreover, LSTM–Caps-Net has a high average accuracy of $94.35 \pm 0.16\%$. The performance of the CNN and ResNet method is relatively high but also weaker than that of the method proposed in this paper; the average fault detection accuracy of the CNN is $92.37 \pm 0.43\%$ and the ResNet is $95.11 \pm 0.20\%$ because the output of these methods is scalar, whereas the output of the capsule network is a vector with direction. Therefore, the capsule network can carry more data features. In addition, because some important data features are lost owing to the existence of a pooling layer in the traditional CNN, its accuracy of the fault detection is lower than that of the method proposed in this paper. When the number of fault types increases, the dimension of the data label increases. Consequently, the classification ability of the SVM that has not been combined with other feature extraction models is weakened. This is the reason why the average accuracy of the SVM is only $80.72 \pm 1.18\%$. Compared with that of the proposed model, the performance of the LSTM–ATT model is only $83.73 \pm 0.96\%$ under the three different load type conditions. This is because the feature extraction ability of the LSTM is weaker than that of the proposed method. Furthermore,

TABLE III
FAULT CLASSIFICATION ACCURACY OF PROPOSED AND OTHER STATE-OF-THE-ART METHODS UNDER DIFFERENT LOAD TYPE CONDITIONS WITH 72 NM LOAD TORQUE

Methods	Fault Detection Accuracy under Different Load Types				Average test time
	CP	CT	LT	Average	
CCaps-Net	**98.33 ±0.11%**	**98.56 ±0.07%**	**98.28 ±0.09%**	**98.39 ±0.09%**	0.08s
LSTM–Caps-Net	94.76 ±0.12%	94.22% ±0.21%	94.07% ±0.15%	94.35 ±0.16%	0.62s
LSTM–ATT	82.45 ±1.12%	85.16% ±0.86	83.59 ±0.89%	83.73 ±0.96%	0.48s
SVM	79.17 ±1.33%	81.73 ±1.14%	81.26 ±1.08%	80.72 ±1.18%	0.12s
CNN	92.88 ±0.43%	91.56 ±0.39%	92.66 ±0.47%	92.37 ±0.43%	0.06s
ResNet	95.13 ±0.22%	95.34 ±0.18%	94.85 ±0.21%	95.11 ±0.20%	0.33s

the attention mechanism model causes over-fitting, which decreases the EM fault detection accuracy of the LSTM–ATT model, and the activation function of the LSTM cell affects the fault detection accuracy of the LSTM-based model. For example, under CP condition with 72 Nm load, for the LSTM–Caps-Net model, when the activation function has been changed to ReLU (Fig. 5 shows the illustrations of ReLU and the default LSTM activation function Tanh), the average fault detection accuracy of the model decreases from 94.76% to 81.22%. Fig. 6 shows the accuracy curve of the last 20 episodes of the validation dataset during the LSTM-based model training for the case in which the parameters of the model stabilize.

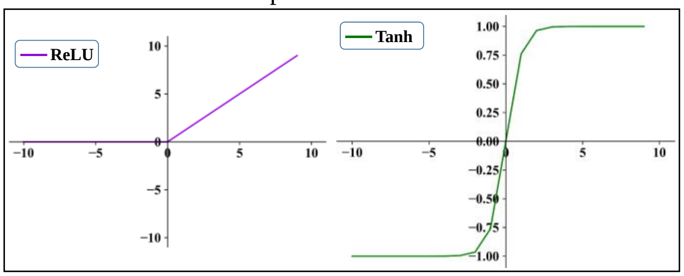


Fig. 5. ReLU and Tanh functions.

The accuracy of the LSTM–Caps-Net of the last 20 episodes is as follows:

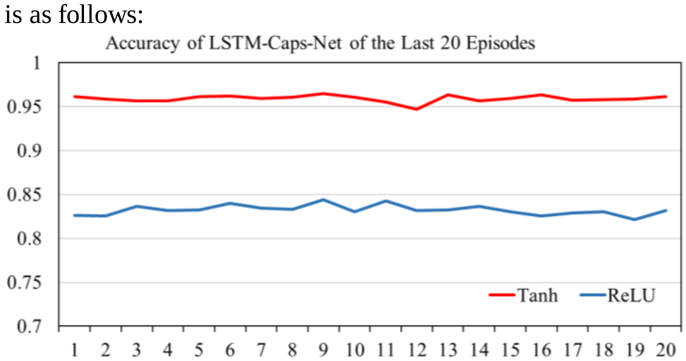


Fig. 6. Impact of LSTM cell activation function on performance of LSTM–Caps-Net fault detection accuracy under CP condition with 72 Nm load.

It can be observed from Fig. 6 that when the activation function is changed to ReLU, the faults detection accuracy decreases a lot. This is because gradient explosion occurred in the neural network, which leads to the decrease of the fault detection accuracy of the model. Moreover, the input data of the model is three-phase sinusoidal current, and there have many negative values in the sampling data, which will cause some neurons to be 0 when using ReLU. Tanh can solve this problem, which is the reason why its faults detection accuracy is higher than the LSTM based model using ReLU function.

In order to explore the efficiency of the proposed method, a fivefold cross validation is carried out. Take the dataset under constant power condition with 72Nm load as an example, the proposed method is compared with the ResNet model and the results are shown in Fig. 7. It presents that the fault detection performance of the proposed method is higher than the ResNet. In addition, the fault detection accuracy is analyzed by the analysis of variance (ANOVA) [36], the ANOVA table is presented in Table IV. It can be observed that the P value is 6.60273e-10, which is much smaller than the significant level value 0.005. Moreover, compared with other methods, the fault detection accuracy of the proposed method is much higher. Above all, it demonstrates that the proposed method has significant advantages for EM fault detection.

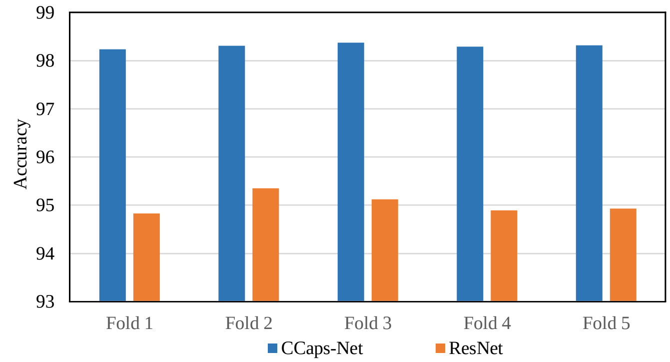


Fig. 7. Fault detection accuracy of fivefold cross validation.

TABLE IV
THE VARIANCE ANALYSIS RESULTS OF THE PROPOSED METHOD AND RESNET

Source	SS	df	MS	F	Prob>F
Columns	26.9616	1	26.9616	1134.03	6.60273e-10
Error	0.1902	8	0.0238		
Total	27.1518	9			

The results of different methods under varying load types are shown in Table V, where "CP→CT" indicates that the fault detection model is trained under CP load dataset and tested under CT load dataset. Due to limited space, this paper only presents the results under different load types conditions of 72 Nm load torque. The fault detection average accuracy of the proposed model is 94.72±0.12% and the accuracy of LSTM-Caps-Net is 91.78±0.39%, respectively. In addition, the faults classification accuracy of the other four state-of-the-art models is 81.22 ± 0.94%, 79.59 ± 1.21%, 91.33 ± 1.30% and 92.91±0.69% respectively, which are lower than the method proposed in this paper. It can be seen from Table V that when the load type changes, the accuracy of the other four models is lower than that of CCaps-Net model. This might be because when the load type of EM changes, the features of the EM data also change. And these state-of-the-art models have weak generalization, which leads to the decrease of fault detection accuracy. CNN without pooling operation retains all features of EM data, which are further processed and stored in the capsule vectors of Caps-Net. So that the CCaps-Net model has stronger generalization ability. However, due to the limited learning ability of LSTM to a large number of data, the accuracy of LSTM based method is still lower than that of CCaps-Net despite the existence of Caps-Net.

The confusion matrices of CCaps-Net under different load type conditions are shown in Fig. 9 of appendix section. In the confusion matrix (a), the misjudgment of the OP state as an ISC state causes the main error of the "CT→LT" scenario. As shown in the distribution of the confusion matrices (b), (c), (d), and (f), the fault detection error of the CCaps-Net model originates mainly from the fact that the ISC and RDE states are misjudged as OP and RSE states in "CP→LT", "CP→LT", "CT→LT", and "LT→CT" scenarios. In matrices (a), (c), (d), and (e), the misjudgment of the LM and ISC states as normal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

states is the error source of the CCaps-Net model. Fig. 8 presents the t-SNE for CCaps-Net with 72 Nm constant power load. It shows that the features of different EM states are distinguishable, and only few states have been misjudged. The error originates mainly from misjudging a few ISC and OP states. In addition, the RSE state is misjudged as RDE state, which causes a fault detection error in the proposed model. The fault classification accuracy of different methods under varying working conditions is shown in Table VI. The results demonstrate that the proposed approach has the best performance; its average fault detection accuracy reaches 98.63%. Compared with Table III, except for CCaps-Net, the accuracy of the other four fault diagnosis models has decreased. This might be because the features of the motor data are more complex under different load size conditions.

TABLE V
ACCURACY OF DIFFERENT METHODS UNDER VARYING LOAD TYPE CONDITIONS WITH 72 NM LOAD TORQUE

Methods	Fault Detection Accuracy Under Different Load Conditions						
	CP→CT	CP→LT	CT→CP	CT→LT	LT→CP	LT→CT	Average
CCaps-Net	**94.52±0.07%**	**95.29±0.14%**	**94.74±0.09%**	**94.67±0.11%**	94. 48±0.21%	**94.44±0.07%**	**94.72±0.12%**
LSTM–Caps-Net	88.61±0.33%	89.73±0.42%	92.44±0.34%	92.77±0.51%	**94. 63±0.28%**	92.66±0.46%	91.78±0.39%
LSTM–ATT	80.79±1.21%	81.81±0.98%	81.28±0.88%	82.49±0.74%	79.29±1.01%	81.64±0.79%	81.22±0.94%
SVM	77.91±1.55%	77.02±1.23%	80.75±1.08%	81.57±1.02%	80.55±0.98%	79.75±1.37%	79.59±1.21%
CNN	93.33±1.02%	93.77±1.13%	90.22±1.53%	90.44±1.42%	89.33±1.62%	90.89±1.08%	91.33±1.30%
ResNet	92.79±0.42%	94.94±0.84%	93.11±0.63%	94.08±0.58%	91.82±0.79%	90.71±0.86%	92.91±0.69%

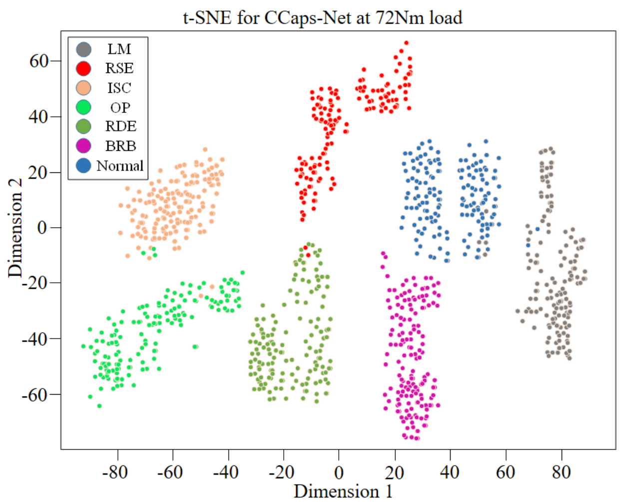


Fig. 8. CCaps-Net feature visualization based on t-SNE at 72 Nm constant power load.

Table VII shows the detection results at constant power and varying loads, where "72→70Nm" indicates that the fault detection models are trained with the motor dataset with 72 Nm torque load and tested with the dataset with 70 Nm torque load. The comparison shows that the proposed method outperforms other methods, which exhibit a higher average fault detection accuracy of 88.40±0.65%. In particular, the fault detection accuracy of CCaps-Net is at least about 6% higher than that of the other state-of-the-art models.

TABLE VI
FAULT CLASSIFICATION ACCURACY OF DIFFERENT METHODS UNDER DIFFERENT WORKING CONDITIONS WITH CONSTANT POWER LOAD

Methods	Fault Detection Accuracy Under Different Working Conditions			
	68 Nm	70 Nm	72 Nm	Average
CCaps-Net	**98.88 ±0.05%**	**98.67 ±0.12%**	**98.33 ±0.07%**	**98.63 ±0.08%**
LSTM–Caps-Net	94.18 ±0.23%	90.89 ±0.37%	94.76 ±0.34%	93.28 ±0.31%
LSTM–ATT	80.62 ±0.76%	80.66 ±0.83%	82.45 ±0.69%	81.24 ±0.76%
SVM	80.85 ±0.98%	76.80 ±1.65%	79.17 ±1.23%	78.94 ±1.29%
CNN	90.11 ±0.34%	90.48 ±0.55%	92.88 ±0.29%	91.16 ±0.39%
ResNet	93.49 ±0.18%	95.97 ±0.09%	94.52 ±0.15%	94.66 ±0.14%

TABLE VII
ACCURACY OF DIFFERENT METHODS AT CONSTANT POWER WITH VARYING LOADS

Methods	Fault Detection Accuracy Under Different Load Conditions						
	72→70Nm	72→68Nm	70→72Nm	70→68Nm	68→72Nm	68→70Nm	Average
CCaps-Net	**93.44±0.44%**	**84.88±0.75%**	**90.52±0.47%**	**85.33±0.63%**	**88.28±1.02%**	**87.94±0.59%**	**88.40±0.65%**
LSTM–Caps-Net	81.42±1.32%	75.15±2.14%	82.67±1.14%	79.89±1.26%	76.81±0.98%	75.94±0.79%	78.65±1.27%
LSTM–ATT	79.89±2.01%	77.75±2.97%	81.60±1.73%	77.36±2.12%	76.56±2.94%	75.02±3.12%	78.03±2.48%
SVM	62.43±4.36%	40.43±10.49%	63.70±4.97%	45.65±9.97%	33.49±8.57%	41.43±11.52%	47.86±8.31%
CNN	71.63±5.63%	71.96±6.67%	78.74±7.34%	79.41±7.05%	78.52±8.78%	80.81±3.86%	76.85±6.56%
ResNet	82.59±0.86%	80.97±1.03%	81.71±0.82%	81.16±1.27%	83.77±0.87%	83.14±0.67%	82.22±0.92%

C. Analysis of Multi-fault Detection Result with Light Load

The fault detection results of the model at heavy load (68, 70, and 72 Nm) have been previously analyzed. The diagnosis results of the different methods under a light load dataset are shown in Table VIII for a load size of 40 Nm, and those of varying loads (40 → 45 Nm) are shown in Table IX.

According to the table, the fault detection accuracy decreases for all methods at 40 Nm and varying loads. However, the proposed approach achieves the highest fault detection accuracy because when the EM load size decreases, the current of the EM decreases, which improves the total harmonic distortion (THD) performance of the current signal. The increase in the THD makes it difficult for the fault detection models to extract the features of the EM current data. This

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

decreases the accuracy of the fault diagnosis of most models.

TABLE VIII
FAULT CLASSIFICATION ACCURACY OF DIFFERENT METHODS AT CONSTANT POWER AND 40 NM EM LOAD TORQUE

Methods	Fault Detection Accuracy Under 40 Nm EM Load Condition
CCaps-Net	94.67±0.23%
LSTM–Caps-Net	90.13±0.29%
LSTM–ATT	78.92±0.88%
SVM	77.06±1.31%
CNN	89.68±0.57%
ResNet	92.48±0.64%

TABLE IX
FAULT CLASSIFICATION ACCURACY OF DIFFERENT METHODS AT CONSTANT POWER AND 40 NM EM LOAD TORQUE

Methods	Fault Detection Accuracy Under Varying EM Load Conditions
	40 → 45 Nm
CCaps-Net	87.11±0.49%
LSTM–Caps-Net	79.07±0.93%
LSTM–ATT	70.85±1.34%
SVM	47.35±5.74%
CNN	81.95±0.76%
ResNet	82.15±0.83%

The results in Sections B and C demonstrate that compared with other state-of-the-art methods, the proposed method has a higher fault detection accuracy and stronger generalization ability. Thus, it can be used for EM multi-fault detection under varying working conditions. In addition, in this study, data features are directly extracted without complex hand-crafted mathematical operations, which increases the efficiency of the fault diagnosis and reduces the impact of human interventions on data processing. In practical application, the method proposed in this paper can detect various types of faults of EMs by pure current signal acquisition without changing the original controller structure or using external measuring equipment (such as an oscilloscope). This reduces measurement and time costs effectively.

V. CONCLUSION

In this paper, a novel data-driven multi-fault detection model based on CCaps-Net is proposed for the state detection of EMs under varying working conditions. The CNN without pooling layer retains the important features of EM data and improves the accuracy of the fault identification. The powerful feature learning ability of Caps-Net enables storing more data information. By combining the capsule network with CNN, the proposed model obtains a higher average multi-fault detection accuracy of at least 98%. Moreover, CCaps-Net has a stronger generalization ability under varying working conditions than other state-of-the-art methods. The EM fault detection accuracy of CCaps-Net under different load size conditions is at least 10% higher than those of other methods.

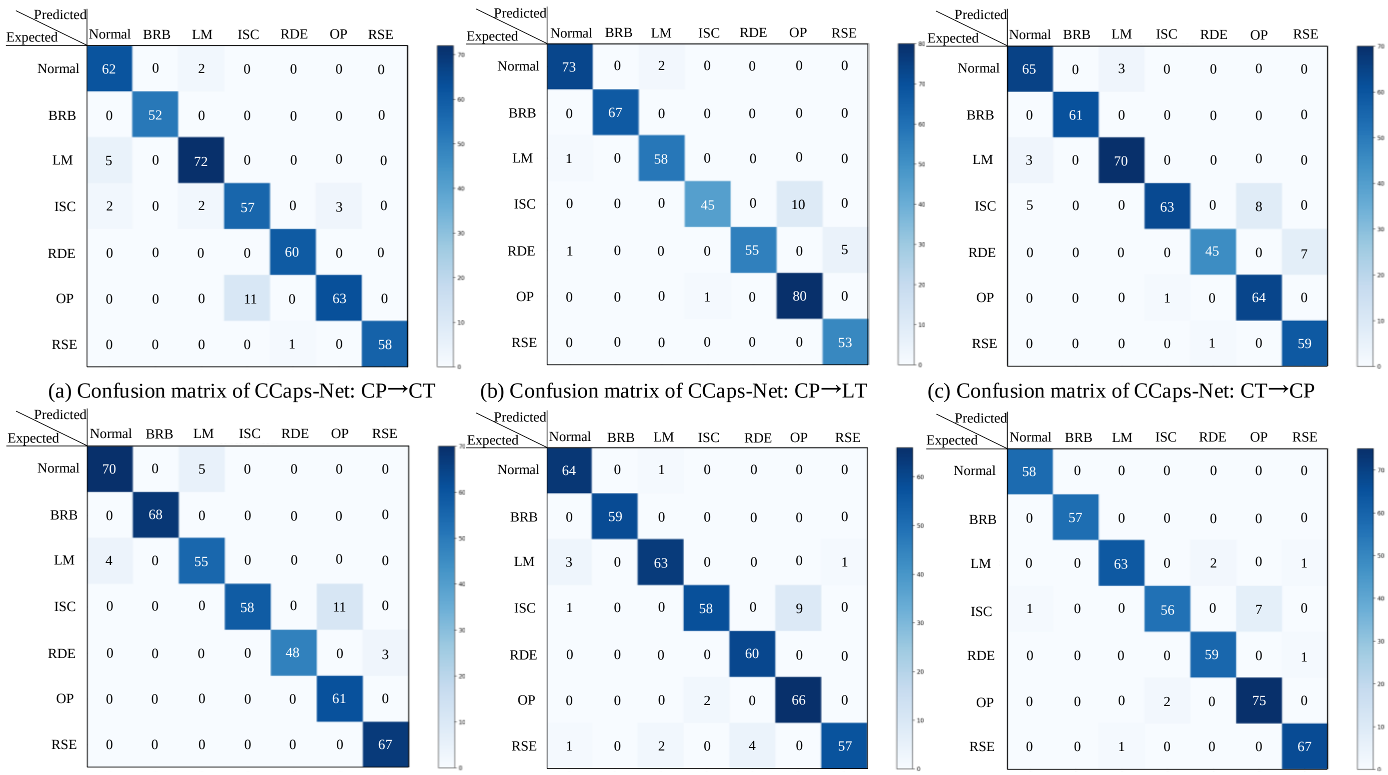
Building an experimental motor platform is particularly time-consuming during the COVID-19 pandemic. Currently, the method proposed in this paper is only being tested on simulation data. When testing with the same dataset, the methods proposed in this paper have the best performance. Many researches are based on simulations (for example, [4] and [29]). Moreover, the number of fault types of the available public data sets is insufficient. In the future, the experimental motor platform will be built and the proposed method will be verified.

APPENDIX

TABLE X
THE HYPER-PARAMETERS OF THE PROPOSED AND STATE-OF-THE-ART METHODS

Hyper-parameters	**CCaps-Net**	**LSTM–Caps-Net**	**LSTM–ATT**	**SVM**	**CNN**	**ResNet**
Learning rate	0.001	0.001	0.0001	/	0.0001	0.001
N_layer	4	7	4	/	4	10 Residual units
Neural cell	128/256/32/7	18/18/18/128/256/32/7	128/128/128/7	/	32/64/1024/7	(64/128/256/512)*10
Epoch	**200**	400	400	/	300	400
Mini-batch	450	450	450	/	450	450
Optimizer	AdamOptimizer	AdamOptimizer	AdamOptimizer	/	AdamOptimizer	AdamOptimizer
Loss function	Margin loss	Margin loss	Softmax cross entropy	/	Softmax cross entropy	Softmax cross entropy
Initialization	Zeros init	Zeros init	Zeros init	/	Zeros init	Zero init
Activation function	ReLU/squashing	Tanh/squashing	Tanh/softmax	/	ReLU/softmax	ReLU/softmax
C	/	/	/	0.9	/	/
Kernel	/	/	/	RBF	/	/
gamma	/	/	/	20	/	/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion



(a) Confusion matrix of CCaps-Net: CP→CT (b) Confusion matrix of CCaps-Net: CP→LT (c) Confusion matrix of CCaps-Net: CT→CP

(d) Confusion matrix of CCaps-Net: CT→LT (e) Confusion matrix of CCaps-Net: LT→CP (f) Confusion matrix of CCaps-Net: LT→CT

Fig. 9. Confusion matrices of CCaps-Net fault detection results under varying load type conditions: (a) "CP→CT", (b) "CP→LT", (c) "CT→CP", (d) "CT→LT", (e) "LT→CP", and (f) "LT→CT".

References

[1] J. You, S. Yin and H. Gao, "Fault detection for discrete systems with network-induced nonlinearities," *IEEE Trans. Ind. Inform.*, vol. 10, no. 4, pp. 2216-2223, Nov. 2014.

[2] J. Hang, S. Ding, J. Zhang, M. Cheng, W. Chen and Q. Wang, "Detection of interturn short-circuit fault for PMSM with simple fault indicator," *IEEE Trans. Energy Convers.*, vol. 31, no. 4, pp. 1697-1699, Dec. 2016.

[3] G. Zhang, W. Hu, D. Cao, et al., "Deep Reinforcement Learning Based Approach for Proportional Resonance Power System Stabilizer to Prevent Ultra-Low-Frequency Oscillations," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5260-5272, Nov. 2020.

[4] I. V. Abramov, Y. R. Nikitin, A. I. Abramov, E. Sosnovich and P. Božek, "Control and diagnostic model of brushless DC motor," *Journal of Electrical Engineering*, vol. 65, no. 5, pp. 277-282, Sep. 2014.

[5] O. Poncelas, J. A. Rosero, J. Cusido, J. A. Ortega and L. Romeral, "Motor fault detection using a rogowski sensor without an integrator," *IEEE Trans. Ind. Electron.*, vol. 56, no. 10, pp. 4062-4070, Oct. 2009.

[6] G. Cirrincione, R. R. Kumar, A. Mohammadi, S. H. Kia, P. Barbiero and J. Ferretti, "Shallow Versus Deep Neural Networks in Gear Fault Diagnosis," *IEEE Trans. Energy Convers.*, vol. 35, no. 3, pp. 1338-1347, Sept. 2020.

[7] M. I. Jordan, T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-60, Jul. 2015.

[8] D. Cao, W. Hu, J. Zhao, et al., "A Multi-agent Deep Reinforcement Learning Based Voltage Regulation using Coordinated PV Inverters", *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4120-4123, Sept. 2020.

[9] D. Silver, J. Schrittwieser, K. Simonyan, et al., "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354-359, Oct. 2017.

[10] I. V. Abramov, A. I. Abramov, Y. R. Nikitin, E. Sosnovich, P. Božek and V. Stollmann, "Diagnostics of electrical drives," 2015 International Conference on Electrical Drives and Power Electronics (EDPE), Tatranska Lomnica, 2015, pp. 364-367.

[11] F. Lv, C. Wen, Z. Bao and M. Liu, "Fault diagnosis based on deep learning," 016 American Control Conference (ACC), Boston, MA, USA, 2016, pp. 6851-6856.

[12] S. Nandi, H. A. Toliyat and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.

[13] R. R. Schoen, B. K. Lin, T. G. Habetler, J. H. Schlag and S. Farag, "An unsupervised, on-line system for induction motor fault detection using stator current monitoring," *IEEE Trans. Ind. Appl.*, vol. 31, no. 6, pp. 1280-1286, Nov.-Dec. 1995.

[14] Z. Feng, X. Chen and M. J. Zuo, "Induction motor stator current AM-FM model and demodulation analysis for planetary gearbox fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 15, no. 4, pp. 2386-2394, April 2019.

[15] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," *IEEE Trans. Ind. Inform.*, vol. 9, no. 2, pp. 760-771, May 2013.

[16] R. Razavi-Far, M. Farajzadeh-Zanjani and M. Saif, "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors," *IEEE Trans. Ind. Inform.*, vol. 13, no. 6, pp. 2758-2769, Dec. 2017.

[17] S. Singh and N. Kumar, "Detection of bearing faults in mechanical systems using stator current monitoring," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1341-1349, Jun. 2017.

[18] J. Tian, C. Morillo, M. H. Azarian and M. Pecht, "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis," *IEEE Trans. Ind. Electro*., vol. 63, no. 3, pp. 1793-1803, Mar. 2016.

[19] C. Lu, Z. Wang, W. Qin and Jian Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377-388, Jan. 2017.

[20] J. Qu, Z. Zhang and T. Gong, "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion," *Neurocomputing*, vol. 171, pp. 837-853, Jan. 2016.

[21] I. Martin-Diaz, D. Morinigo-Sotelo, O. Duque-Perez and R. J. Romero-Troncoso, "An experimental comparative evaluation of machine learning techniques for motor fault diagnosis under various operating conditions," *IEEE Trans. Ind. Appl.*, vol. 54, no. 3, pp. 2215-2224, May-June 2018.

[22] F. Cheng, J. Wang, L. Qu and W. Qiao, "Rotor-current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier," *IEEE Trans. Ind. Appl.*, vol. 54, no. 2, pp. 1062-1071, Mar.-Apr. 2018.

[23] R. Liu, G. Meng, B. Yang, C. Sun and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1310-1320, Jun. 2017.

[24] C. Wei et al., "Substation Equipment Thermal Fault Diagnosis Model Based on ResNet and Improved Bayesian Optimization," *2019 9th International Conference on Power and Energy Systems (ICPES)*, Perth, Australia, 2019, pp. 1-5.

[25] M. Barzegaran, A. Mazloomzadeh and O. A. Mohammed, "Fault diagnosis of the asynchronous machines through magnetic signature analysis using finite-element method and neural networks," *IEEE Trans. Energy Convers.*, vol. 28, no. 4, pp. 1064-1071, Dec. 2013.

[26] W. Sun, R. Zhao, R. Yan, S. Shao and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1350-1359, June 2017.

[27] S. Shao, S. McAleer, R. Yan and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Inform.*, vol. 15, no. 4, pp. 2446-2455, April 2019.

[28] R. Sabir, D. Rosato, S. Hartmann and C. Guehmann, "LSTM Based Bearing Fault Diagnosis of Electrical Machines using Motor Current Signal," *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA, 2019, pp. 613-618.

[29] Z. Zhu, G. Peng, Y. Chen and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62-75, Jan. 2019.

[30] S. Sabour, N. Frosst, and G. E. Hinton. "Dynamic routing between capsules," in *Proceedings of the Advances in Neural Information Processing Systems*, 2017.

[31] Z. Liu, Z. Wu, T. Li, J. Li and C. Shen, "GMM and CNN hybrid method for short utterance speaker recognition," *IEEE Trans. Ind. Inform.*, vol. 14, no. 7, pp. 3244-3252, July 2018.

[32] J. Chen, W. Hu, D. Cao, et al., "An Imbalance fault detection algorithm for variable-speed wind turbines: A deep learning approach," *Energies*, vol. 12, no. 14, pp. 2764-2778, Jul. 2019.

[33] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp.770-778.

[34] Hatem A. Fayed, Amir F. Atiya. "Speed up grid-search for parameter selection of support vector machines," *Applied Soft Computing*, vol. 80, pp. 202-210, July 2019.

[35] I. M. Abdulbaqi, A. T. Humod, O. K. Alazzawi, "Application of FEM to Provide the Required Database for MCSA Based On-line Fault Detection System on 3-phase Induction Motor using ANSYS Maxwell2D," *Advances in Natural & Applied Sciences*, vol. 10, no. 16, pp. 43-54, Nov. 2016.

[36] M. N. S. K. Shabbir, X. Liang and S. Chakrabarti, "An ANOVA-Based Fault Diagnosis Approach for Variable Frequency Drive-Fed Induction Motors," *IEEE Trans. Energy Convers*., Early Access.

Jianjun Chen received the B.S. from Chongqing University of Technology, Chongqing, China, in 2018. He is currently working toward the Ph.D. degree in control science and engineering at the University of Electronic Science and Technology of China (UESTC). His research interests include mechanical fault diagnosis and deep learning.

Weihao Hu (S'06–M'13–SM'15) received the B.Eng. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, both in electrical engineering, and Ph. D. degree from Aalborg University, Denmark, in 2012.

He is currently a Full Professor and the Director of Institute of Smart Power and Energy Systems (ISPES) at the University of Electronics Science and Technology of China (UESTC). He was an Associate Professor at the Department of Energy Technology, Aalborg University, Denmark and the Vice Program Leader of Wind Power System Research Program at the same department. His research interests include artificial intelligence in modern power systems and renewable power generation. He has led/participated in more than 15 national and international research projects and he has more than 170 publications in his technical field.

He is an Associate Editor for IET Renewable Power Generation, a Guest Editor-in-Chief for Journal of Modern Power Systems and Clean Energy Special Issue on Applications of Artificial Intelligence in Modern Power Systems, a Guest Editor-in-Chief for Transactions of China Electrical Technology Special Issue on Planning and operation of multiple renewable energy complementary power generation systems, and a Guest Editor for the IEEE TRANSACTIONS ON POWER SYSTEM Special Section on Enabling very high penetration renewable energy integration into future power systems. He was serving as the Technical Program Chair (TPC) for IEEE Innovative Smart Grid Technologies (ISGT) Asia 2019 and is serving as the Conference Chair for the Asia Energy and Electrical Engineering Symposium (AEEES 2020). He is currently serving as Chair for IEEE Chengdu Section PELS Chapter. He is a Fellow of the Institution of Engineering and Technology, London, U.K. and an IEEE Senior Member.

Di Cao received the B.S. from North China University of Water Resources and Electric Power, Zhengzhou, China, in 2014. He is currently working toward the Ph.D. degree in control science and engineering at the University of Electronic Science and Technology of China. His research interest includes optimization of distribution network.

Man Zhang (M'20) and received the B.S. and M.S. degrees from Northwestern Polytechnical University, Xi'an, China, in 2012 and 2015, respectively. She prepared the Ph.D. degree at Génie Electrique et électronique de Paris, Gif-Sur-Yvette, France, and received the Ph.D. degree from University Paris-XI, France, in 2019, all in electrical engineering. She is currently a Lecture with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China. Her research interests include numerical and analytical analysis of multiphysic problems, vibration, and noise reduction of SRM

Qi Huang (S'99, M'03, SM'09) was born in Guizhou province in the People's Republic of China. He received his BS degree in Electrical Engineering from Fuzhou University in 1996, MS degree from Tsinghua University in 1999, and Ph.D. degree from Arizona State University in 2003. He is currently a professor at UESTC, the Executive Dean of School of Energy Science and Engineering, UESTC, and the director of Sichuan State Provincial Lab of Power System Wide-area Measurement and Control. He is a member of IEEE since 1999. His current research and academic interests include power system instrumentation, power system monitoring and control, and power system high performance computing.

Zhe Chen (M'95-SM'98-F'19) received the B.Eng. and M.Sc. degrees from Northeast China Institute of Electric Power Engineering, Jilin, China, and the Ph.D. degree from University of Durham, Durham, U.K.

He is a Full Professor with the Department of Energy Technology, Aalborg University, Denmark. He is the Leader of Wind Power System Research Program in the Department of Energy Technology, Aalborg University and the Danish Principle Investigator for Wind Energy of Sino-Danish Centre for Education and Research. His research areas include power systems, power electronics and electric machines; and his main current research interests are wind energy and modern power systems. He has led many research projects and has more than 400 publications in his technical field.

Dr. Chen is an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS, a Fellow of the Institution of Engineering and Technology, London, U.K., a Chartered Engineer in the U.K., and a Fellow of the IEEE.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TEC.2020.3046642, IEEE Transactions on Energy Conversion

Frede Blaabjerg (S'86–M'88–SM'97–F'03) was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. From 1988 to 1992, he got the PhD degree in Electrical Engineering at Aalborg University in 1995. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998. From 2017 he became a Villum Investigator. He is honoris causa at University Politehnica Timisoara (UPT), Romania and Tallinn Technical University (TTU) in Estonia.

His current research interests include power electronics and its applications such as in wind turbines, PV systems, reliability, harmonics and adjustable speed drives. He has published more than 600 journal papers in the fields of power electronics and its applications. He is the co-author of four monographs and editor of ten books in power electronics and its applications.

He has received 32 IEEE Prize Paper Awards, the IEEE PELS Distinguished Service Award in 2009, the EPE-PEMC Council Award in 2010, the IEEE William E. Newell Power Electronics Award 2014, the Villum Kann Rasmussen Research Award 2014, the Global Energy Prize in 2019 and the 2020 IEEE Edison Medal. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON POWER ELECTRONICS from 2006 to 2012. He has been Distinguished Lecturer for the IEEE Power Electronics Society from 2005 to 2007 and for the IEEE Industry Applications Society from 2010 to 2011 as well as 2017 to 2018. In 2019-2020 he serves as President of IEEE Power Electronics Society. He is Vice-President of the Danish Academy of Technical Sciences too. He is nominated in 2014-2019 by Thomson Reuters to be between the most 250 cited researchers in Engineering in the world.