IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 15, 2017, accepted December 17, 2017, date of publication December 21, 2017, date of current version March 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2785763

Data-Based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM

SENLIN ZHANG, (Member, IEEE), YIXING WANG, MEIQIN LIU, (Senior Member, IEEE), AND ZHEJING BAO, (Member, IEEE)
College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China
Corresponding author: Meiqin Liu (liumeiqin@zju.edu.cn)

This work was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ15F030001.

ABSTRACT Power system faults are significant problems in power transmission and distribution. Methods based on relay protection actions and electrical component actions have been put forward in recent years. However, they have deficiencies dealing with power system fault. In this paper, a method for data-based line trip fault prediction in power systems using long short-term memory (LSTM) networks and support vector machine (SVM) is proposed. The temporal features of multisourced data are captured with LSTM networks, which perform well in extracting the features of time series for a long-time span. The strong learning and mining ability of LSTM networks is suitable for a large quantity of time series in power transmission and distribution. SVM, with a strong generalization ability and robustness, is introduced for classification to get the final prediction results. Considering the overfitting problem in fault prediction, layer of dropout and batch normalization are added into the network. The complete network architecture is shown in this paper in detail. The parameters are adjusted to fit the specific situation of the actual power system. The data for experiments are obtained from the Wanjiang substation in the China Southern Power Grid. The real experiments prove the proposed method's improvements compared with current data mining methods. Concrete analyses of results are elaborated in this paper. A discussion of practical applications is presented to demonstrate the feasibility in real scenarios.

INDEX TERMS Data mining, power system faults, recurrent neural networks, support vector machines.

I. INTRODUCTION

The primary goal is to ensure the reliability and stability of the power system as power grids expand and loads increase. High-accuracy fault prediction in power systems increases the operational reliability and stability, and can help to prevent huge losses resulting from power accidents. Fault prediction is the process of analyzing and mining historical data to predict whether or not there is a fault in the power system so that measures are taken to prevent accidents and ensure system recovery. This is a core technology and a maintenance security approach, and more advanced than fault diagnosis, in helping to make reasonable decisions to prevent faults and reduce the adverse effects. Line trip faults are one of the most common faults in a power system, and they have been researched actively in recent years [1], [2]. If reclosing is unsuccessful, it will result in large-scale power outages and property losses. Consequently, line trip fault prediction is a significant and valuable research subject.

In past studies of power system faults, various artificial intelligence methodologies were proposed including expert systems [3], [4], Bayesian networks [5], [6], rough sets [7], [8], Petri nets [9], [10], neural networks [11], [12], etc. Moreover, researchers have focused on relay protection actions and electrical component actions over the past few decades [13], [14]. A new analytical model was developed to take into account the possible malfunctions in protective relays and circuit breakers based on the existing analyses. It improved the accuracy of fault diagnosis results to a further degree [13]. An analytical model was provided based on the improved temporal constraint network [14]. System fault event reasoning and diagnosis were formulated as an optimization problem where the fault hypotheses were tested. However, these processes have their deficiencies in dealing with power system faults. The result is negatively affected if there are malfunctions in the relay protection and electrical components. Furthermore, the method based on protective relays and circuit breakers approaches the problem after the faults have happened. It cannot do much for predicting whether there is a fault in the power transmission and distribution. However, the electrical measurement data is first-hand information about faults. Analyzing the historical data can help to predict if there will be faults in the power system and help in making a corresponding decision to prevent them. In general, making good use of the electrical measurement data can improve the performance of fault prediction and ensure the reliability and stability of a power system. Research on data-based methodologies has just begun to appear in recent years. It has become a valuable and urgent subject at present.

Research studies on fault prediction based on artificial neural networks (ANNs) have been proposed in recent years. An improved prediction method for optimized ANN based on multilevel genetic algorithms was proposed to improve the fault forecasting model accuracy [15]. In [16], an artificial neural network-based methodology was proposed for early online prediction of any transient instability, which has a great impact on the performance of wide area protection and control systems. Another example is [17], which presented a new methodology based on ANN to detect and predict faults in a boilerąŕs burner system in a power plant. However, there are large quantities of temporal information during power transmission and distribution, which contributes a great deal to fault prediction but cannot be extracted by ANN. Recurrent neural networks (RNNs), a deep learning method, are shown to have a strong ability to capture the hidden correlations occurring in the big data in applications for image captioning [18], [19], voice conversion [20], [21] and natural language processing [22], [23]. They also show good performance in dealing with faults [24], [25]. However, the original RNN has the problem of a vanishing gradient because the later nodes perception of the previous nodes decreases. Long short-term memory (LSTM) networks [26], as an improved network architecture, were proposed to solve the problem mention above. Compared with conventional RNNs, LSTM networks perform well in extracting the features of time series for a longer time span. One example is [27], which proposed an approach based on an LSTM network to get good diagnosis and prediction performances in the cases of complicated operations, hybrid faults and strong noise. In [28], the use of the LSTM network was proposed to accomplish timely detection and identification of faults based on the commonly available measurement signals. The result showed that the LSTM network was better suited for the railway track circuit fault detection and identification tasks than the convolutional network. Moreover, the prediction ability was also proved in [29], which proposed a novel traffic forecast model based on the LSTM network. The compared results validated that the proposed LSTM network can achieve a better performance. In general, the LSTM network is an improved RNN, which deals better with longer time series. However, the research on data-based fault prediction in power systems using LSTM networks is still in the beginning stages.

During the fault prediction process, classification is an essential part. Support vector machine (SVM) is a discriminant classifier defined by a hyperplane. The applications based on SVM were proved to be feasible in [30] and [31]. Due to the good robustness and generalization performance of SVM, it is used for classification of the features captured by the LSTM networks in this paper.

In the previous research, methodologies for a stacked autoencoder for power system fault diagnosis were proposed [32], [33]. The results of simulations prove the feasibility of fault prediction based on deep learning methods. In this paper, a method for data-based line trip fault prediction in power systems using LSTM networks and SVM is proposed to capture the temporal features of the data and increase the performance compared to conventional methods. When applied to each line in a power system, the proposed methodology can detect which line is experiencing the fault. Furthermore, the proposed methodology can be put into practice in power grids for fault prediction to improve fault prevention work and reduce the losses caused by power accidents. The main contributions are outlined as follows.

1) The correlation with temporal information between line trip faults and measurement data is mined for fault prediction before faults happen through LSTM networks and SVM.

2) Measurements for each line including current, voltage, and active power are chosen as the inputs to obtain more comprehensive information. The temporal information is fused with the merging layer through three LSTM subnetworks. The results validate the improved performance.

3) The features, captured through LSTM networks, are put into the SVM classifier to estimate if there is a fault in the power system, which improves the accuracy of the fault prediction markedly.

4) The proposed methodology can work in practice and the parameters can be learned offline and updated online to fit the new operating status. It is an improved methodology for real-time fault prediction.

The rest of the paper is organized as follows. The detailed problem formulation is described in Section II. Section III introduces the architectures and algorithms for RNNs, LSTM networks and SVM. In Section IV, the improved model for power system fault prediction based on LSTM networks and SVM is proposed. Then, the simulation experiments are presented to prove the feasibility and advantages of the proposed methodology in Section V. Finally, the conclusion is given in Section VI.

II. PROBLEM STATEMENT

Line trips are a common fault, which can lead to massive blackouts. In recent years, relay protection actions and electrical component actions have been used for fault diagnosis. For applicability and stability, the fault relevance should be considered. The most common reasons for line trip faults include aging and damage of the distribution equipment, bad insulation, weather changes, and so on. There is a gradual process of distribution line resistance before the line trip faults occur. The electrical measurements will change according to some rules, which includes current, voltage, active power, and the reactive power of users during the process. The proposed approach aims to capture the features of this process

FIGURE 1. A simple flow chart for the solution to fault prediction.

to detect faults. Therefore, the fault relevance has to be mined between the fault records and the electrical measurements. Assume that P is the result of fault prediction, where $P = 1$ representing that there is a fault and $P = 0$ representing normal operations, which is the label for each sample in the network training. A simple flow chart for the solution to the problem in this paper is shown in Fig. 1. Feature mining for big data in power systems is a significant and difficult problem in the process. With respect to LSTM networks and SVM, the key problem is in designing the parameters and network architectures for high-accuracy fault prediction. Specifically, a practical experiment was carried out in this work with real-world data obtained from the power supply administration division of the China Southern Power Grid.

III. ARCHITECTURES AND ALGORITHMS

In this section, the architecture and algorithms for the proposed methodology are introduced in detail. The LSTM network is an improved network based on RNNs, which introduces a core unit called a cell. The improvement helps to solve the vanishing gradient problem of RNNs. Moreover, SVM performs well and has a certain robustness in classification.

A. RECURRENT NEURAL NETWORK

A recurrent neural network is an improved class of artificial neural network using the temporal information of the input data, where connections between units form a directed cycle within the same layer. In contrast, a conventional neural network only has connections between the layers. The units in a layer have no connection. The network does not transmit the temporal information so the performance in dealing with time series may be poor.

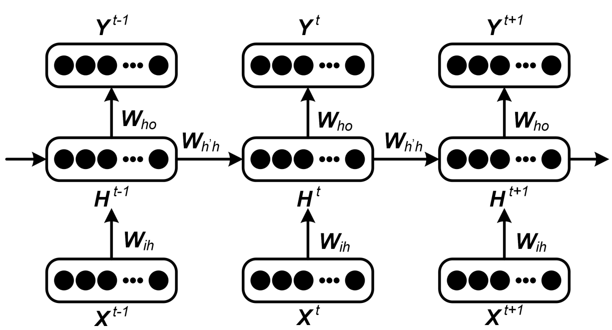

FIGURE 2. A simple RNN structure, where *X* is the input unit, *H* is the hidden unit, *Y* is the output unit, and *W* is the weight matrix.

A simple structure for an RNN is shown in Fig. 2. The process of forward propagation can be concluded from Fig. 2, given by eqs. (1)-(3).

$$a_h^t = \sum_{i=1}^{I} w_{ih} x_i^t + \sum_{h'=1}^{H} w_{h'h} s_{h'}^{t-1} \tag{1}$$

$$s_h^t = f_h(a_h^t) \tag{2}$$

$$a_o^t = \sum_{h=1}^{H} w_{ho} s_h^t \tag{3}$$

The notation w is the weight; a is the sum calculated through weights; f is the activation function; s is the value after calculation by the activation function; t represents the current time of the network; i is the number of input vectors; h is the number of hidden vectors in t time; h' is the number of hidden vectors in $t-1$ time; and o is the number of output vectors.

The hidden layer units receive not only the data input, but also the output of the hidden layer from the last time. Therefore, the network can remember the previous information and apply it to the calculation of the current output. The RNN structure creates an internal state of the network to exhibit dynamic temporal behavior and has a better approach in dealing with time series analysis. Correspondingly, the RNN needs to be trained by back propagation through time.

Unfortunately, the original RNN is affected by the vanishing gradient problem because the later nodesąŕ perception of the nodes from the previous time step decreases. The performance declines when the network structure becomes deep and complex.

B. LONG SHORT-TERM MEMORY

To solve the vanishing gradient problem, a long short-term memory block is introduced into the RNN to remember the values for the case of either long or short durations of time. Specifically, the hidden units of the RNN are replaced by LSTM blocks containing three gates, which are used to control the flow of information into or out of their memory. A peephole LSTM block with cell, input, output, and forget gates is shown in Fig. 3.

The theoretical deduction process for an LSTM block is given by eqs. (4) through (12).

$$a_j^t = \sum_{i=1}^{I} w_{ij} x_i^t + \sum_{h=1}^{H} w_{hj} s_h^{t-1} + \sum_{c=1}^{C} w_{cj} m_c^{t-1} + b_j \tag{4}$$

$$s_j^t = f(a_j^t) \tag{5}$$

$$a_k^t = \sum_{i=1}^{I} w_{ik} x_i^t + \sum_{h=1}^{H} w_{hk} s_h^{t-1} + \sum_{c=1}^{C} w_{ck} m_c^{t-1} + b_k \tag{6}$$

$$s_k^t = f(a_k^t) \tag{7}$$

$$a_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} s_h^{t-1} + b_c \tag{8}$$

$$d_c^t = s_k^t m_c^{t-1} + s_j^t g(a_c^t) \tag{9}$$

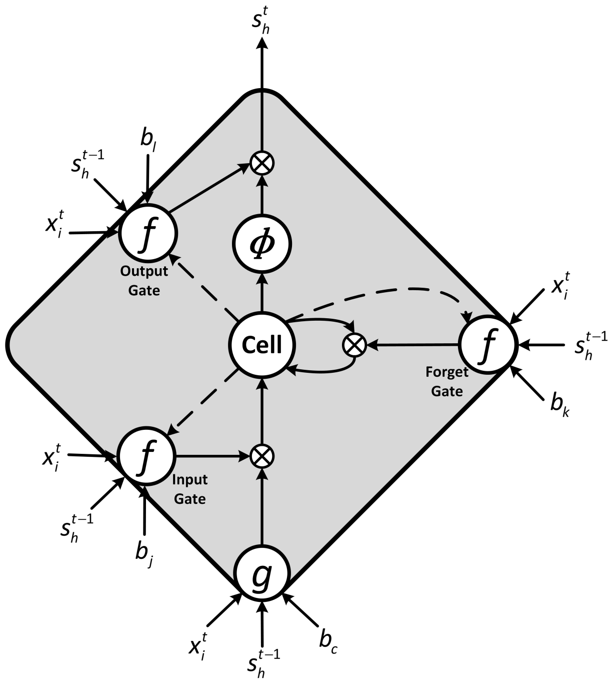

FIGURE 3. **A peephole LSTM block with cell input, output, and forget gates, where all edges have fixed unit weight. The dashed line represents the peephole connection between the current and the last time step.**

$$a_l^t = \sum_{i=1}^{I} w_{il} x_i^t + \sum_{h=1}^{H} w_{hl} s_h^{t-1} + \sum_{c=1}^{C} w_{cl} m_c^t + b_l \tag{10}$$

$$s_l^t = f(a_l^t) \tag{11}$$

$$s_h^t = s_l^t \phi(d_c^t) \tag{12}$$

The symbols in eqs. (4)-(12) are illustrated as follows. The notation m is the input from cell to input gate, j is the number of the input gate vector, k is the number of the forget gate vector, c is the number of the cell vector, l is the number of the output gate vector, d is the value of the cell, and f, g and ϕ are the activation functions.

As shown from the structure, LSTM networks can learn when to let errors into or out of the block. When the weights of the input gate take a zero value, no values can get into the block. Moreover, the value cannot get out when the output gate takes zero value. When both gates are closed, the value is trapped in the cell so that the value will not grow or shrink or have an effect on the output of the current time steps. Therefore, in the back propagation process, the gradient can be propagated back across many time steps without exploding and vanishing. Due to the long short-term memory block, LSTM networks have a strong ability to learn the long-range dependencies of time series and performs better in practice compared to the original RNNs.

C. CLASSIFICATION

Classification is the key step in fault prediction. The fundamental steps of the logistic regression classifier and support vector machine are elaborated in detail in this subsection.

1) LOGISTIC REGRESSION CLASSIFIER

The aim of the logistic regression classifier is to learn a 0/1 classification model from the training data features using a logistic function. Suppose that there is a fixed training set $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})\}$, and $y^{(m)} \in \{0, 1\}$, where m is the number of samples. The hypothesis function is given by

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{13}$$

where θ is the weight of the logistic regression classifier to be obtained. The result of the hypothesis function represents the probability P of $y = 1$, so it can be concluded that

$$P(y = 1|x; \theta) = h_\theta(x) \tag{14}$$

$$P(y = 0|x; \theta) = 1 - h_\theta(x) \tag{15}$$

The cost function can be derived by a maximum likelihood estimate, shown in eq. (16).

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} log h_\theta(x^{(i)}) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))] \tag{16}$$

Therefore, a gradient descent method can be used to minimize the cost function.

2) SUPPORT VECTOR MACHINE

SVM is a supervised learning model for classification and regression analysis. The basic rule looks for an optimal hyperplane which is the farthest from the nearest training samples.

For the linear separation in a two-dimensional plane, the classification function can be

$$f(x) = w^T x + b \tag{17}$$

where w, b determines a straight line for classification. According to the relation of geometrical margin $\widetilde{\gamma}$ and functional margin $\hat{\gamma}$ given by

$$\widetilde{\gamma} = \frac{\hat{\gamma}}{\|\omega\|} \tag{18}$$

The problem can be inferred as eq. (19) when the $\hat{\gamma} = 1$.

$$max \frac{1}{\|\omega\|}, s.t., y_i(w^T x_i + b) \geq 1, i = 1, \ldots, n \tag{19}$$

where $y_i = 1$ or -1 is the label of samples, and n is the number of samples. To extend SVM to cases in which the data is not linearly separable, the hinge loss function is given by

$$max(0, 1 - y_i(w^T x + b)) \tag{20}$$

Then the loss function can be concluded as eq. (21).

$$min_{w,b} \sum_{i=1}^{N} max(0, 1 - y_i(w^T x + b)) + \lambda\|\omega\|^2 \tag{21}$$

For linear inseparability, SVM can use a kernel trick and nonlinear mapping algorithm to transform samples from the low-dimensional input space into high-dimensional feature space so that they are linearly separable.

In general, SVM is a novel learning method with solid theoretical foundation, which has good robustness with fewer samples and generalization performance in nonlinear problems compared to logistic regression classifiers. The reason is that the optimization goal of SVM is structural risk minimization instead of empirical risk minimization. It not only ensures the classification accuracy of the samples but also reduces the dimension of the learning model, corresponding to the two terms of eq. (21). In short, it helps to prevent problems of overfitting. Moreover, the computational complexity depends on the number of support vectors instead of the dimension of the sample space, which in some sense avoids the dimension disaster. Therefore, SVM has low computational complexity and good robustness.

IV. METHODOLOGY OF FAULT PREDICTION BASED ON LSTM NETWORKS AND SVM

In this section, we elaborate on the methodology for the line trip fault prediction in power systems based on LSTM networks and SVM. First, the character and source of the data are described in detail. Then the solution to the overfitting problem is discussed. The line trip fault prediction modeling is shown in the last subsection.

A. DATA DESCRIPTION

The real-world historical data was derived from electrical measurements, the equipment ledger, the equipment health, weather, and topology. They were obtained from the power supply administration in the China Southern Power Grid for the years 2012-2014. The electrical measurements including current, voltage, active power, and reactive power of the users are closely related to the faults because of the gradual process of the distribution line resistance. This paper focuses on the correlation between the faults and the electrical measurements. Current, voltage, and active power are selected for the input in the network. They contain all the measurement information so that the phase angle is involved and the reactive power is redundant. During the training process, 500 points of current, voltage, and active power are recorded for samples before the line trip faults or during normal operation. The sampling period is 15 minutes. A normal sample and fault sample for the current are shown in Fig. 4. The difference between them cannot be determined simply from the figures. This is similar for the voltage and active power so the hidden features should be mined though LSTM networks for fault prediction. The samples are time series with temporal information, which are converted into different dimensions for the LSTM networks input. This will be discussed in part of the simulation experiment.

B. DATA PREPROCESSING

Data preprocessing is a fundamental work of data mining. Different types of data have different dimensions. In order to reduce the impact of different magnitudes and dimensions, and increase the convergence rate, the data for users from the power supply administration needs to be preprocessed with 0-1 standardization by

$$y_p = \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}}(x - x_{\min}) + y_{\min} \tag{22}$$

where x is the raw data; y_p is the processed result; $y_{\max}$ and $y_{\min}$ are the maximum and minimum settings of the processed result; and $x_{\max}$ and $x_{\min}$ are the maximum and minimum of the unprocessed data. In the simulation experiment for this paper, $y_{\max}$ and $y_{\min}$ are set by 1 and 0 so that the processed data can be constrained in the [0,1] interval, which is a nonnegative number. After preprocessing, the data is transformed into the same level without changing the inner variations.

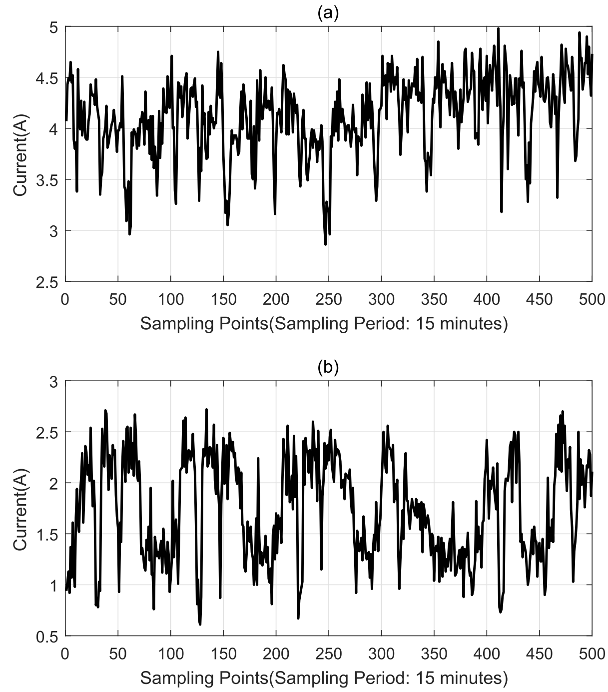

FIGURE 4. (a) Normal sample and (b) fault sample of the current variations.

C. OVERFITTING PROBLEM

Overfitting is the key problem in fault prediction because of the characteristics of the sample. During the stable operation of a power system, faults rarely happen. The number of fault samples is small. Therefore, the training can easily fall into an overfitting problem. During the network training process, when the number of iterations increases, the network may have a good fit for the training set and small loss in the training set, but the fitting to the validation set is poor. Dropout and batch normalization are the effective solutions for such an overfitting problem. The theories are discussed in this subsection.

1) DROPOUT

The basic dropout method is that during the forward propagation of networks, the neurons stop working with

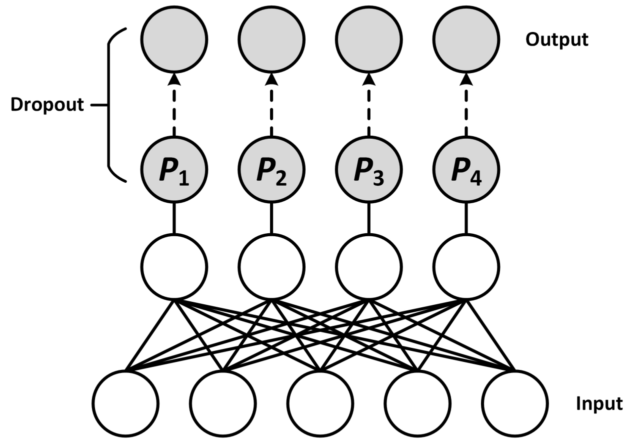

FIGURE 5. The schematic diagram of one layer dropout, where *P* is the probability of presenting neurons.

probability P. The schematic diagram is shown in Fig. 5. Random neurons do not work, so the situation of a better performance in some fixed combination is avoided. The network can learn some common knowledge for better generalization performance. Moreover, training the network with dropout can be considered as training multiple subnetworks. The output is the average of all the subnetworks. The subnetworks from each iteration are basically not repeated. Therefore, it can avoid a scenario where the trained network is excessively fitted to the training set and accelerates the convergence rate. In the end, the accuracy of the fault detection is improved.

2) BATCH NORMALIZATION

The essence of batch normalization is that data normalization is done at the intermediate layer during the training of each batch. The mean and standard deviation of output in the normalization layers are 0 and 1. The nature of the learning process is to study the distribution features of the data. Once the distribution of the training set is different from that of the test set, the generalization ability of the network is greatly reduced. Moreover, if the distribution of each batch of training data is different, the network must learn to adapt to different distributions for each iteration, which will reduce the speed of training the network and the performance of the network. Therefore, using batch normalization can improve the generalization performance and accelerate the convergence rate.

D. MODELING OF FAULT PREDICTION

For line trip fault prediction, the major task is detecting whether there is a fault or not during the operation of a power system. The overall structure of the model is shown in Fig. 6. The samples are labeled as normal or fault according to the fault records. Five hundred points of current, voltage, and active power are recorded for samples before line trip faults or during normal operation. Then, the 500-dimensional vector is translated into multiple input vectors with time steps for LSTM networks. The temporal features are captured through three LSTM networks. The last moment of the LSTM output units are obtained for merging in a fusion layer. In order to retain information about each measurement, the form of the fusion is set as concatenating. Layers of dropout and batch normalization are added after the three LSTM networks to deal with the overfitting problem in data-based fault prediction. Then the trained, merged temporal features are put into the SVM classifier for fault prediction results, as it has better performance in classification. The structure and parameters of the model are designed and adjusted according to multiple tests with samples for satisfactory results, which are shown in the next section. This is the key problem in data-based fault prediction in power systems. Moreover, if there are other types of associated data about faults recorded in the power system, the proposed network can be extended with the LSTM subnetworks to exploit more information for fault prediction. The improved performances in a practical experiment are shown in Section V.

TABLE 1. Samples from the simulation experiment for case 1.

Samples Label	Normal	Fault
Number of Training Samples	2050	2050
Number of Testing Samples	510	510

V. SIMULATION RESULTS

The simulation condition and results from the practical experiment are discussed in this section, which proves that the proposed method has improved performance.

A. EXPERIMENTAL CONDITION AND EVALUATION METHOD

The experiments were done with data from the Wanjiang substation in the China Southern Power Grid, shown in Fig. 7. There are 36 feeders connecting to the load side in the Wanjiang substation, called Feeder 1. Feeder 2, ... , Feeder 36. The current, voltage, and active power of the users are recorded under these lines.

In evaluating the performance of the trained network, testing is a significant step. A K-fold cross validation (K-CV) is a classical approach to evaluate the trained network. In K-fold cross-validation, the original samples are randomly and averagely partitioned into K subsamples. Each subsample is a testing set, and the rest of the K-1 subsamples are the training set. The experiments are repeated K times with these subsamples. The average performance of the K models is considered the final performance. The advantage of this method is that all observations are used for both training and testing so that overfitting and lack of learning are avoided to enable persuasive results.

Generally, K is 5. In this simulation, the number of training samples and test samples is shown in Table 1.

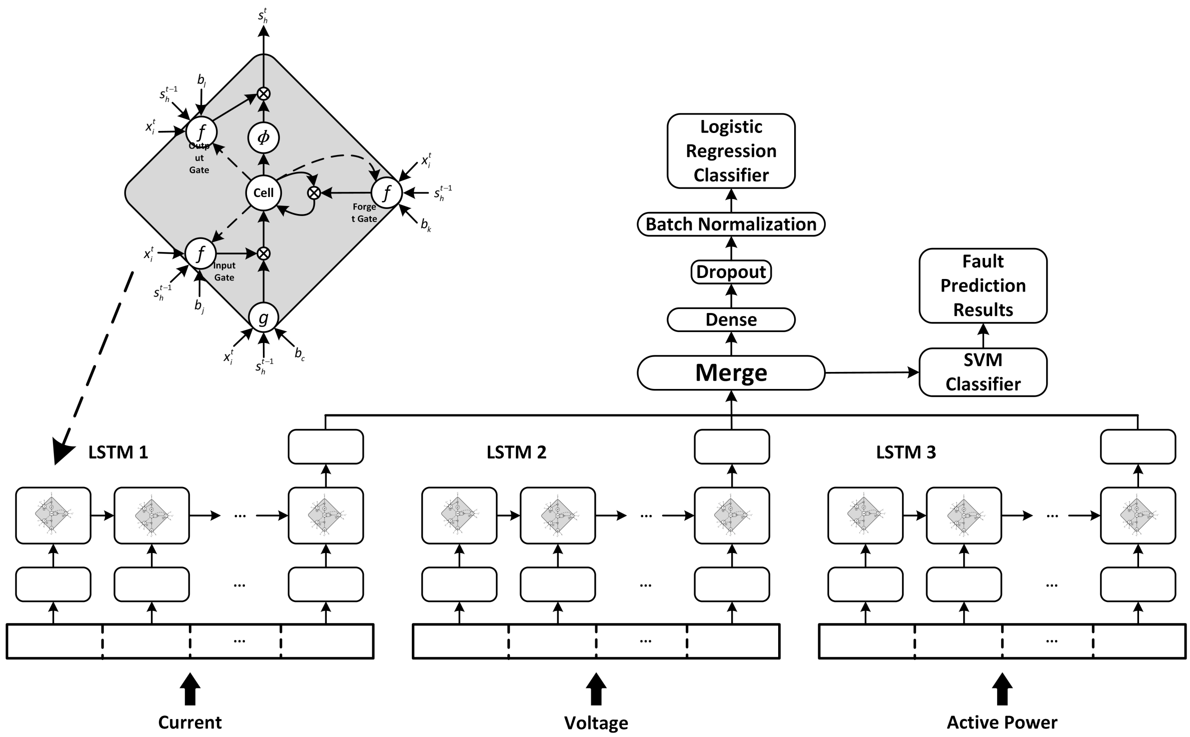

FIGURE 6. Proposed model for the data-based line trip fault prediction, where detailed layer connections are indicated.

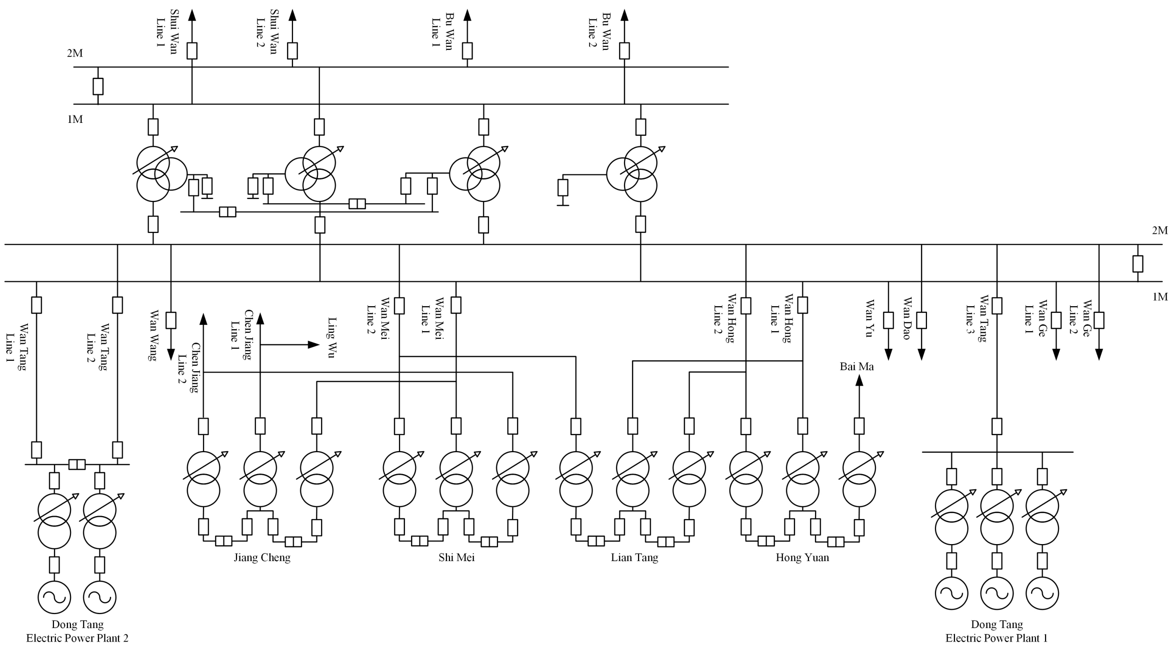

FIGURE 7. Primary electrical system in Wanjiang above 110 kv, including electric power plants, transmission buses, converting stations, and user loads.

The performance of the fault diagnosis can be observed through the accuracy rate which is the ratio of proper diagnosis samples and all test samples. The training is stopped if the accuracy does not improve in multiple epoches. Then the best accuracy of the epoches is considered the result in an experiment. The average accuracy of repeated experiments is recorded as the final result.

B. FAULT PREDICTION BASED ON THE LSTM NETWORKS

The experiment for fault prediction based on LSTM networks is discussed in this subsection, the model of which is shown in Fig. 6 with SVM removed. The parameters of LSTM networks are set as in Table 2 for better performance through multiple experiments, which should be adjusted in practical circumstances. The number of epoches is set as 40 because

TABLE 2. Parameters in the experiment.

Parameters	Value
Input Time steps	25
Input Dimension	50
Output Units	1
Batch Size	32
Epoch	40
Optimizer	RMSprop
Learning Rate	0.001
Decay	0.9
Dropout rate	0.5

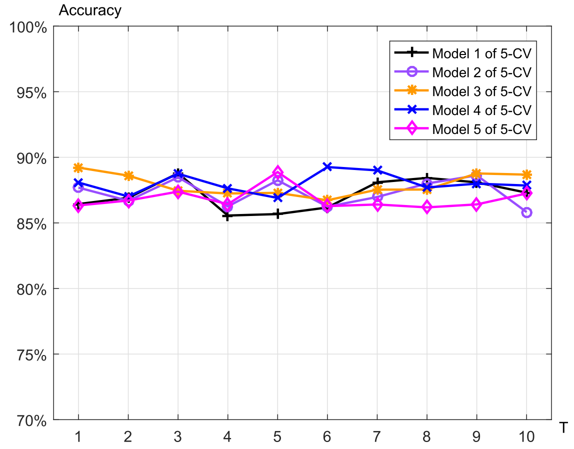

FIGURE 8. Results of fault prediction based on the LSTM fusion network based on 5-CV, where the experiment of each model repeats 10 times.

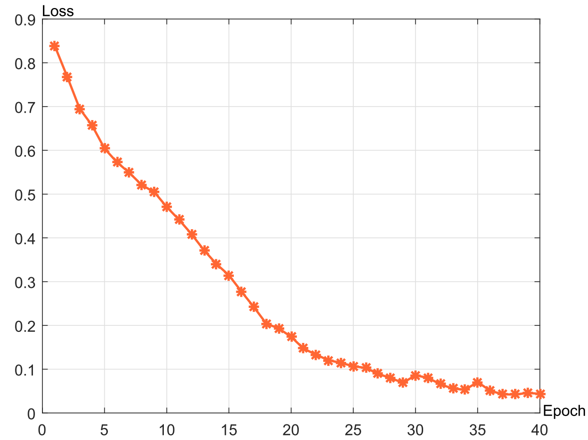

FIGURE 9. The loss in fault prediction through 40 epoches of training.

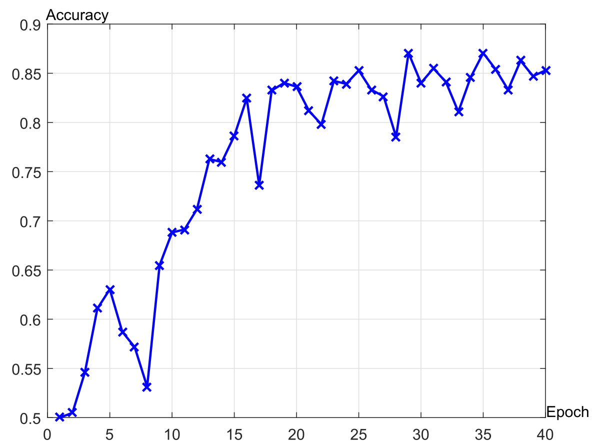

FIGURE 10. The accuracy of the fault prediction through 40 epoches of training.

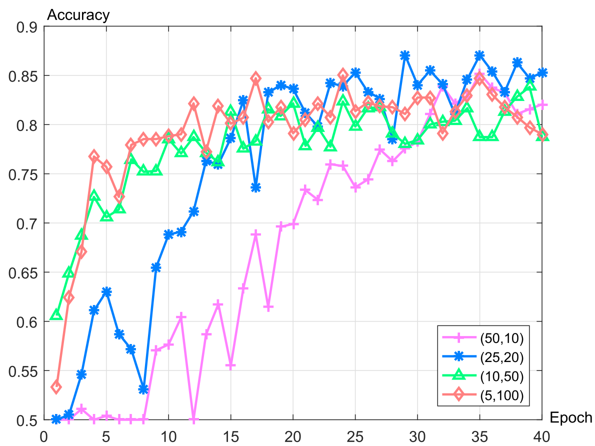

FIGURE 11. Comparative results of different input time steps and dimensions, where T in (T,D) is the time step and D in (T,D) is the dimension of the input vector.

the accuracy is stable according to repeated experiments. The 'RMSprop' optimizer is chosen for its better performance in recurrent neural networks. The results of fault prediction based on the LSTM fusion network based on 5-CV are shown in Fig. 8. The final result is the average accuracy of five trained models: 87.44%. The accuracy of the fault prediction and loss during a training period are shown in Fig. 9 and Fig. 10. The accuracy increases with training while the loss decreases. Repeated experiments ensure the stability of the network. The results prove that the correlation with temporal information between line trip faults and measurement data can be mined for fault prediction. The fault can be detected with high probability but still needs to be improved.

C. INFLUENCE FACTORS FOR IMPROVED PERFORMANCE IN FAULT PREDICTION

The input time step and dimension are important influencing factors because LSTM networks extract temporal features. The experiment results of several representative input time steps and dimensions are shown in Fig. 11. When the input time step reaches above 50, the convergence rate is too slow so it is meaningless for showing results. It can be concluded that the input dimension of (25, 20) is better. The reason is as follows. If the time step is long, it is inevitable that the learned features are lost through the long process. Then, the convergence rate is slow. On the other hand, if the time step is short and the dimension of the input vector is high, the temporal information is lost. Therefore, the input time step and dimension are set as (25, 20), which is suitable in this case.

Overfitting is a difficult problem in data-based fault prediction. Layers of dropout and batch normalization are

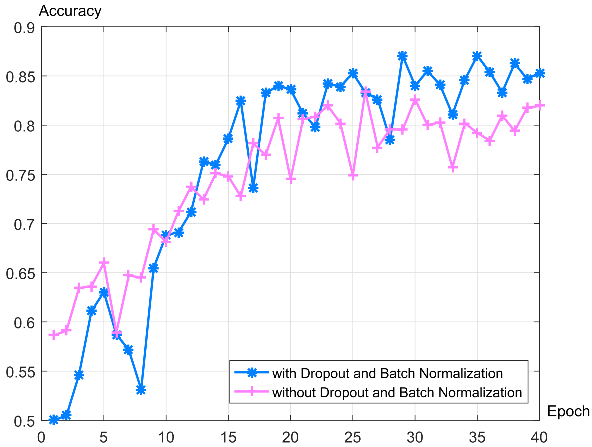

FIGURE 12. Improved results using the dropout and batch normalization in a training process.

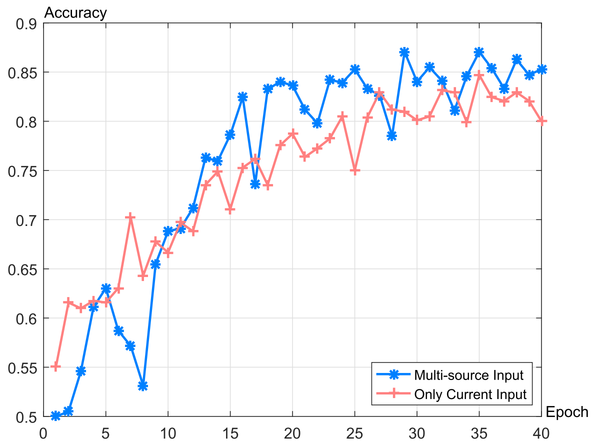

FIGURE 13. Comparison results of multi-sourced information and single inputs in a training process.

introduced into the network to avoid overfitting. With the parameters set as in Table 2, the experiment is carried out to prove the remission of overfitting. The improved results using dropout and batch normalization are shown in Fig. 12. The performance is improved markedly with layers of dropout and batch normalization. The convergence rate is low and unstable at the beginning of the training because the network has more layers and parameters to train. However, the accuracy increases in the latter part. In general, the network with dropout and batch normalization performs better and has a faster convergence rate.

There is a considerable amount of measurement data in the operation of a power system. The temporal features are captured from the usersąŕ data for current, voltage, and active power. The comparison experiments are performed for multi-sourced inputs and single inputs. The results are shown in Fig. 13, where the improvement can be clearly concluded from the two curves. It can be explained as follows. When a fault is about to happen, the gradual state transformation of the power equipment is reflected in all of the measurement

TABLE 3. The accuracy for LSTM networks with improved influence factors in 10 repeated experiments.

Number	1	2	3	4	5
Accuracy	86.8%	88.5%	87.5%	87.7%	88.1%
Number	6	7	8	9	10
Accuracy	86.3%	85.1%	88.3%	86.7%	89.7%

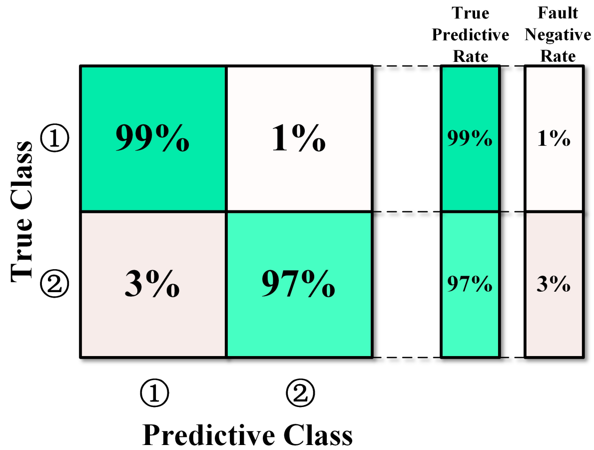

FIGURE 14. Confusion matrix for the results based on the proposed method.

TABLE 4. The accuracy for the proposed method in 10 repeated experiments.

Number	1	2	3	4	5
Accuracy	97.8%	97.5%	97.7%	97.6%	97.8%
Number	6	7	8	9	10
Accuracy	97.7%	97.5%	97.6%	97.7%	97.7%

data including current, voltage, and active power. The clearer features are mined with more adequate information. Therefore, multi-LSTM networks show better performance.

The improvement is proven to be effective in similar experiments in data-based fault prediction using the LSTM networks and SVM. The features are better captured with better performance of LSTM networks, which are then more conducive to classification with the SVM classifier.

The results of repeated experiments based on LSTM networks with improved influence factors are shown in Table 3. The average for the accuracy is 87.5%. It is acceptable but still needs to be improved.

D. RESULTS OF PROPOSED METHOD FOR FAULT PREDICTION

The data-based line trip fault prediction method using LSTM networks with SVM is proposed for increasing the accuracy in fault prediction. The trained merged temporal features are put into the SVM classifier for fault prediction results, shown in Fig. 6. The results from repeated experiments and one of the confusion matrixes are shown in Table 4 and Fig. 14. The average for accuracy is 97.7%, which increases by about 10% compared to the method based on just LSTM

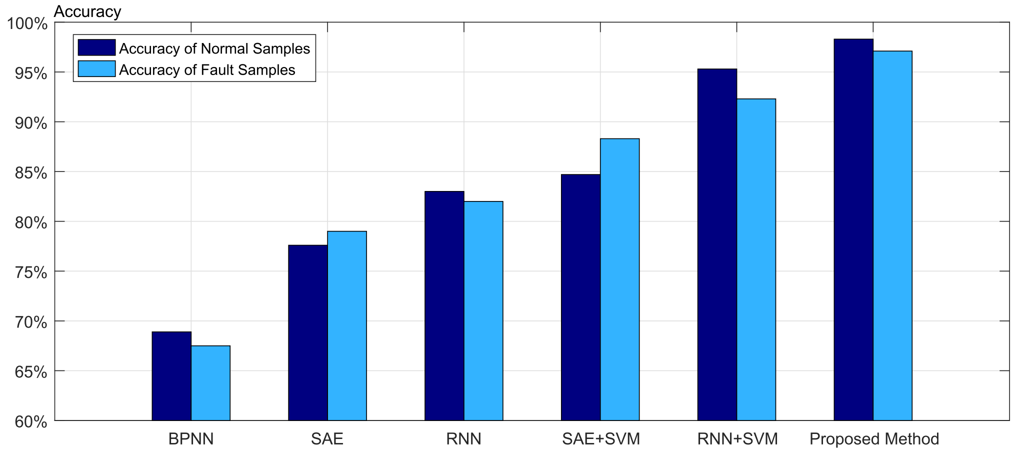

FIGURE 15. Comparison results for the proposed method and data-based mining approaches in the case of data-based line trip fault prediction, which includes back propagation neural networks (BPNNs), stacked autoencoders (SAEs), RNNs and SVM.

TABLE 5. Comparison results for the proposed method and the data-based mining approaches.

Methods	BPNN	SAE	RNN	SAE+ SVM	RNN+ SVM	Proposed Method
Accuracy	68.2%	78.3%	82.5%	86.5%	93.8%	**97.7%**

TABLE 6. The results for the proposed method from the first half of the year 2014.

Month	Jan	Feb	Mar	Apr	May	Jun
Accuracy	96.5%	97.0%	97.3%	97.8%	95.4%	96.6%

networks. Moreover, the results show the true high prediction rate both in normal samples and fault samples derived from the confusion matrix. The improvement relies on SVM, which has good robustness with samples and in generalization performance for nonlinear problems. SVM basically does not involve the probability measure and law of large numbers. Moreover, the optimization goal of SVM is structural risk minimization, which alleviates the overfitting problem. It can ensure the classification accuracy of samples and reduce the dimension of the learning model. These advantages make it suitable for data-based fault prediction in power systems.

The experiment was done with the same samples from Table 1 based on current data mining approaches. The comparison results for the proposed method and the data-based mining approaches are shown in Fig. 15, where the accuracy of the normal and fault test set is involved. It shows that the improvement is marked in both normal and fault situations. The final results are shown in Table 5.

It can be concluded that the proposed method performs much better than other approaches such as back propagation neural networks (BPNNs), stacked autoencoders (SAEs), RNNs and SVM. LSTM networks can extract the temporal information from data depending on the connected hidden layer units but the BPNN and SAE do not have this ability. Compared to RNNs, the LSTM can solve the problem of the vanishing gradient using the LSTM block.

In general, it relies on the stronger learning ability of LSTM networks for time series, and the good robustness and generalization performance of SVM. The fault features are mined from multiple sources of measurement data for fault prediction at high accuracy. Compared to the methods based on relay protection actions and electrical component actions, the proposed data-based method can predict if there will be faults in a power system based on first-hand information. Therefore, the proposed methodology, that of data-based line trip fault prediction in power systems using LSTM networks with SVM, is a noteworthy improvement for ensuring the reliability and stability of a power system.

E. EXPERIMENT AND APPLICATION IN PRACTICAL SITUATIONS

The proposed methodology was proven effective in a repeated data experiment in the last subsection. When the proposed network is built in accordance with a practical situation and trained with historical data, it can work in online fault prediction. The parameters are constantly updated to fit the new operating status on line. In a practical application situation, LSTM networks were trained on historical data for the years 2011-2013 and tested with the real fault records from the first half of the year for 2014 from the Wanjiang substation in Guangdong, China. The results are shown in Table 6. The accuracy of each month is stable above 95%. The results have great significance in practical fault prediction.

The training and testing time for the experiment was at the minute level with a Tesla M40 GPU. For practical application, the related research was done at the State Grid Electricity Research Institute in Shandong, China. Higher configuration computers and clusters are available at the State Grid for dealing with big power data. The hardware requirements for online training and prediction were easy to fulfill using the State Grid facilities. In general, practical fault prediction can be achieved with the proposed method.

VI. CONCLUSION

To increase operational reliability and stability in power systems, a data-based method for line trip fault prediction using LSTM networks with SVM is proposed in this paper. First, 500 sampling points of current, voltage, and active power were recorded for samples before line trip faults or during normal operations. The samples were reshaped into dimensions of (25,20) as the input for LSTM networks. Meanwhile, the samples were processed with 0-1 standardization to the same level. Then, the multi-sourced data was put into LSTM networks for training and fusion. The temporal features were mined through three LSTM networks. To solve the overfitting problem of fault prediction in power systems, layers of dropout and batch normalization were added into the network. Afterwards, the fusion features were put into the SVM classifier for more accurate prediction results. Moreover, the proposed network can be extended with LSTM subnetworks to obtain more information for fault prediction if there is other available data related to faults from the power system. The experiments were done with real data from the Wanjiang substation in the China Southern Power Grid. Specifically, the corresponding experiment results prove the increase in accuracy using fusion of multiple sources, and layers of dropout and batch normalization. In summary, the improvement of the proposed method compared to the current data mining approaches is noteworthy. The accuracy of the line trip fault prediction can reach about 97%. It is of great significance for operational reliability and the stability of a power system. Moreover, the proposed method is demonstrated in practice in the last part of the paper. The hardware requirements can be met with the equipment in a power system. The proposed method for accurate fault prediction is valuable in practical application.

REFERENCES

[1] G. Zheng, G. Radman, W. Guan, and S. Yang, "Wide area phasor measurements based disturbance monitoring for line trip event," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Vancouver, BC, Canada, Jul. 2013, pp. 1–5.

[2] D. Zhou, Y. Liu, and J. Dong, "Frequency-based real-time line trip detection and alarm trigger development," in *Proc. IEEE PES Gen. Meeting*, Washington, DC, USA, Jul. 2014, pp. 1–5.

[3] M. T. Veljko, R. T. Predrag, and M. D. Zeljko, "Expert system for fault detection and isolation of coal-shortage in thermal power plants," in *Proc. Conf. Control Fault-Tolerant Syst.*, Nice, France, Oct. 2010, pp. 666–671.

[4] C. Nan, F. Khan, and M. T. Iqbal, "Abnormal process condition prediction (fault diagnosis) using G2 expert system," in *Proc. Can. Conf. Elect. Comput. Eng.*, Vancouver, BC, Canada, Apr. 2007, pp. 1507–1510.

[5] Z. Yongli, H. Limin, and L. Jinling, "Bayesian networks-based approach for power systems fault diagnosis," *IEEE Trans. Power Del.*, vol. 21, no. 2, pp. 634–639, Apr. 2006.

[6] Y. Brandt, B. W. Jervis, and Y. Maidon, "Circuit multi-fault diagnosis and prediction error estimation using a committee of Bayesian neural networks," in *Proc. IEE Colloq. Test. Mixed Signal Circuits Syst.*, London, U.K., Oct. 1997, pp. 1–7.

[7] S. S. S. Rawat, V. A. Polavarapu, V. Kumar, E. Aruna, and V. Sumathi, "Anomaly detection in smart grid using rough set theory and K cross validation," in *Proc. Int. Conf. Circuits, Power Comput. Technol.*, Nagercoil, India, Mar. 2014, pp. 479–483.

[8] X. Xu and J. F. Peters, "Rough set methods in power system fault classification," in *Proc. Can. Conf. Elect. Comput. Eng.*, Winnipeg, MB, Canada, May 2002, pp. 100–105.

[9] Z. Jiang, Z. Li, N. Wu, and M. Zhou, "A Petri net approach to fault diagnosis and restoration for power transmission systems to avoid the output interruption of substations," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2017.2682185.

[10] S. K. Yang, "A condition-based failure-prediction and processing-scheme for preventive maintenance," *IEEE Trans. Rel.*, vol. 52, no. 3, pp. 373–383, Sep. 2003.

[11] W. Li, A. Monti, and F. Ponci, "Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2651–2665, Nov. 2014.

[12] S. Bhattacharya, "Fault detection on a ring-main type power system network using artificial neural network and wavelet entropy method," in *Proc. Int. Conf. Comput., Commun. Autom.*, Noida, India, May 2015, pp. 1032–1037.

[13] W. Guo, F. Wen, G. Ledwich, Z. Liao, X. He, and J. Liang, "An analytic model for fault diagnosis in power systems considering malfunctions of protective relays and circuit breakers," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1393–1401, Jul. 2010.

[14] Y. Cui, J. Shi, and Z. Wang, "Power system fault reasoning and diagnosis based on the improved temporal constraint network," *IEEE Trans. Power Del.*, vol. 31, no. 3, pp. 946–954, Jun. 2016.

[15] Y. Tao, J. Zheng, T. Wang, and Y. Hu, "A state and fault prediction method based on RBF neural networks," in *Proc. IEEE Workshop Adv. Robot. Social Impacts*, Shanghai, China, Jul. 2016, pp. 221–225.

[16] E. Rakhshani, I. Sariri, and K. Rouzbehi, "Application of data mining on fault detection and prediction in boiler of power plant using artificial neural network," in *Proc. Int. Conf. Power Eng., Energy Elect. Drives*, Lisbon, Portugal, Mar. 2009, pp. 473–478.

[17] M. Mahdi and V. M. I. Genc, "Artificial neural network based algorithm for early prediction of transient stability using wide area measurements," in *Proc. 5th Int. Istanbul Smart Grid Cities Congr. Fair*, Istanbul, Turkey, Apr. 2017, pp. 17–21.

[18] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[19] H. Zuo, H. Fan, E. Blasch, and H. Ling, "Combining convolutional and recurrent neural networks for human skin detection," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 289–293, Mar. 2017.

[20] T. Nakashika, T. Takiguchi, and Y. Ariki, "Voice conversion using RNN pre-trained by recurrent temporal restricted Boltzmann machines," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 3, pp. 580–587, Mar. 2015.

[21] S. Leglaive, R. Hennequin, and R. Badeau, "Singing voice detection with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, South Brisbane, QLD, Australia, Apr. 2015, pp. 121–125.

[22] V.-T. Tran, K.-H. Nguyen, and D.-H. Bui, "A vietnamese language model based on recurrent neural network," in *Proc. 8th Int. Conf. Knowl. Syst. Eng.*, Hanoi, Vietnam, Oct. 2016, pp. 274–278.

[23] W.-W. Liu, W.-Q. Zhang, Y. Shi, A. Ji, J. Xu, and J. Liu, "Improved phonotactic language recognition based on RNN feature reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, May 2014, pp. 5322–5326.

[24] A. I. Moustapha and R. R. Selmic, "Wireless sensor network modeling using modified recurrent neural networks: Application to fault detection," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 5, pp. 981–988, May 2008.

[25] C. Xu, G. Wang, X. Liu, D. Guo, and T.-Y. Liu, "Health status assessment and failure prediction for hard drives with recurrent neural networks," *IEEE Trans. Comput.*, vol. 65, no. 11, pp. 3502–3508, Nov. 2016.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] T. de Bruin, K. Verbert, and R. Babuska, "Railway track circuit fault diagnosis using recurrent neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 523–533, Mar. 2017.

[28] M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," in *Proc. IEEE Int. Conf. Aircraft Utility Syst.*, Beijing, China, Oct. 2016, pp. 135–140.

[29] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Jan. 2017.
[30] S. U. Jan, Y. D. Lee, J. Shin, and I. Koo, "Sensor fault classification based on support vector machine and statistical time-domain features," *IEEE Access*, vol. 5, pp. 8682–8690, May 2017.
[31] M. Bigdeli, M. Vakilian, and E. Rahimpour, "Transformer winding faults classification based on transfer function analysis by support vector machine," *IET Electr. Power Appl.*, vol. 6, no. 5, pp. 268–276, May 2012.
[32] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, Jul. 2017.
[33] Y. Wang, M. Liu, and Z. Bao, "Deep learning neural network for power system fault diagnosis," in *Proc. 35th Chin. Control Conf.*, Chengdu, China, Jul. 2016, pp. 6678–6683.

SENLIN ZHANG (M'11) received the B.E. degree in control theory and control engineering from the Wuhan University of Technology, Wuhan, China, in 1984, and the M.E. degree in control theory and control engineering from Zhejiang University, Hangzhou, China, in 1991. He is currently a Professor with the College of Electrical Engineering, Zhejiang University. His current research interests include textile automation, intelligent systems, and underwater wireless sensor networks.

YIXING WANG received the B.E. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2014, where he is currently pursuing the Ph.D. degree in control theory and control engineering with the College of Electrical Engineering. His research interests include data mining and fault diagnosis in power systems.

MEIQIN LIU (M'09–SM'13) received the B.E. and Ph.D. degrees in control theory and control engineering from Central South University, Changsha, China, in 1994 and 1999, respectively. She was a Post-Doctoral Research Fellow with the Huazhong University of Science and Technology, Wuhan, China, from 1999 to 2001. She was a Visiting Scholar with the University of New Orleans, New Orleans, LA, USA, from 2008 to 2009. She is currently a Professor with the College of Electrical Engineering, Zhejiang University, Hangzhou, China. She has authored over 90 peer reviewed papers, including 44 journal papers. Her current research interests include information fusion, intelligent systems, and nonlinear control.

ZHEJING BAO (M'07) received the B.S. and M.S. degrees in control science and engineering from Shandong University, Jinan, China, in 1996 and 1999, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2007. From 2007 to 2009, she was a Post-Doctoral Research Associate with the College of Electrical Engineering, Zhejiang University, where she is currently an Associate Professor. Her research interests include the simulation, intelligent control, optimal energy management, and planning of microgrid.