Fault-cause identification method based on adaptive deep belief network and time–frequency characteristics of travelling wave

Hanqing Liang†, Yadong Liu‡, Gehao Sheng†, Xiuchen Jiang†

†Department of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, People’s Republic of China
‡E-mail: hyd@sjtu.edu.cn

Abstract: Accurate fault-cause identification is highly important to the fault analysis of overhead transmission lines (OTLs). In order to improve the efficiency and accuracy of fault identification, this study proposes a fault identification method based on the ADBN (adaptive deep belief network) model and the time–frequency characteristics of a travelling wave. According to the mechanisms of different OTL faults, the appropriate time–frequency characteristic parameters of the fault current travelling wave were selected as the input of the ADBN model, and the fault-type labels were selected as the output. The ADBN model introduces the idea of adaptive learning rate into CD (contrastive divergence) algorithm and improves its performance with self-adjusting learning rate. The parameters of the ADBN model were pre-trained with the improved CD algorithm and adjusted by back propagation algorithm with the labels of the samples. The performance of the ADBN model was verified by field data, and the accuracy of fault identification was analysed under different model parameters, characteristic parameters, and sample sizes. The results showed that the model helps to characterise the inherent relationship between characteristic parameters and fault causes, and the proposed method can effectively identify different fault causes in OTLs.

1 Introduction

Fault identification of overhead transmission lines (OTLs) is of great significance to fault diagnosis, formulation of late line-defence schemes, and fault-accident manoeuvres [1]. The traditional fault identification method is realised by fault phase selection components. The principle of the method is to judge the fault line according to the characteristics of the electrical quantity during the fault time, and then the fault cause can be identified based on the meteorological conditions and manual inspection. The common fault root causes can be categorised as branches, mountain fire, animals, lightning, icing, wind, and external damage [2–4].

The traditional fault identification method is easy to implement, but its identification criterion is vulnerable to the influence of different factors such as system operation plan, fault location, and transition resistance. Also, the identification results are easily affected by subjective factors. With the increase of fault cases, the characteristics of electrical quantities are diversified due to different types of lines, voltage levels, fault locations, and operating conditions. These diverse characteristics point to higher requirements for fault diagnosis. Moreover, the on-site manual inspection expends lots of manpower and resources. Thus, a data-driven method is needed to directly identify the fault cause with high accuracy.

Fuzzy logic is a common data-driven approach to enable modelling of subjective and imprecise information, and it has been applied to fault diagnosis of transmission lines [5, 6]. However, fuzzy-logic-based methods often need to be constructed manually by domain experts and require a high degree of robustness to ensure that the rules work. Another class of methods known as decision tree has been adopted for fault classification in transmission lines, and it requires less feature extraction [7–9]. The main limitation of decision-tree-based methods is that they rely heavily on pre-established decision logic. In recent years, with the development of machine learning methods, neural networks (NNs), support vector machines (SVMs), and other methods have been applied to fault identification of transmission lines and some results have been achieved [10–13]. Specifically, NNs are one of the most frequently applied methods in fault diagnosis [14, 15]. Machine learning methods overcome the limitations of conventional systems and reduce the dependence on prior expert knowledge. Also, they are not affected by the system operation or fault location [16]. However, there are also some problems with existing machine learning methods. The NN-based methods have strong learning ability but can easily get stuck at a local optimum, and convergence is slow when dealing with large sample data. The SVM-based methods have outstanding performance when dealing with small sample data, but they are less efficient in dealing with the multi-classification problem of fault diagnosis. Furthermore, existing machine learning methods mainly classify faults from the perspective of line selection, and few studies have focused on a method to directly classify and identify the fault root cause. Moreover, the existing fault-cause identification method has certain limitations in identifying the number of fault types or the recognition accuracy [2–4].

Due to the deficiency of learning ability and processing efficiency, the current methods still have limitations in promotion and application. The deep belief network (DBN) model was proposed in 2006 [17]. The research shows that the feature data obtained by the model are more representative of the original data and more conducive to classification and identification. Layer-by-layer training methods can be used to train the deep NN. The DBN methods can learn more features and parameters without overfitting, and overcome the shortcomings of being time-consuming. The method has outstanding performance in feature recognition, classification, and prediction. Furthermore, the DBN method has been applied in fault classification modelling for power transformers [18]. The results show that it has strong capability of extracting features and excellent fault-tolerance characteristics. With the development of fault detectors on OTLs and the accumulation of fault cases, the DBN model has a large dataset for training to enable reliable fault-cause identification.

Based on the concept of the DBN model, this paper adopts the ADBN (adaptive DBN) to directly identify the fault root cause of OTLs. The ADBN model successfully avoids under-learning and falling into local optimum, which are caused by the fixed learning rate of the traditional model. Based on the mechanism of different fault causes, the paper selects the time–frequency information as the characteristic parameters. The time–frequency information includes the half-wavelength, the amplitude, the steepness, the energy in different time periods, and the energy in different
frequency bands of the fault current travelling wave. Then, the fault diagnosis model is constructed based on ADBN. The characteristic parameters are taken as the input of the model, and the different fault-cause labels are taken as the output of the model. The performance of the ADBN diagnosis model is verified by field data, which is collected by the fault locator on OTLs. Moreover, the accuracy of fault identification is analysed under different model parameters, characteristic parameters, sampling rates, training-set sizes, and sample-set sizes. The results demonstrate that the proposed method can effectively identify different fault root causes of OTLs.

2 Adaptive DBN

The traditional NNs generally contain a single hidden layer. With the increase in the number of layers, the number of the model parameters rapidly increases, and the model training becomes time-consuming. The deep NN requires a large amount of label data during training. It is difficult for the deep network to find the optimal solution when the sample size is small. In addition, it is easy to fall into the local optimal solution. The DBN solves the problem of deep network optimisation by adopting layer-by-layer training [17]. The training gives the whole network a good initial configuration of each hidden-layer unit are independent (no conditions of each hidden-layer unit are independent (no)

A DBN model is composed of several RBMs, and the layer-by-layer training process is carried out from the lower layer to the upper layer. Due to the characteristics of RBM, the training of the DBN model becomes effective. In the DBN model, the training data of the later layers is more representative through the hidden-layer feature extraction, and the problem of insufficient sample size can be solved by generating new data. In order to rapidly train the RBM, the contrastive divergence (CD) algorithm [19] is adopted during the training process. In this paper, we adopt the ADBN to identify the fault root cause. Compared to the traditional DBN model, the ADBN model is improved by the ability to dynamically adjust the learning rate according to the similarities and differences of each iteration direction of the parameters in the CD algorithm. Thus, the ADBN model successfully avoids under-learning and falling into local optima, which are caused by the fixed learning rate of the traditional DBN model.

Like the traditional structure of DBN, ADBN consists of a sequence of RBMs. Fig. 1 shows the structure of the DBN model. The RBM consists of a visible layer (input layer) and a hidden layer (output layer). The visible layer and hidden layer use the bidirectional full connection. Moreover, there is no connection between units in the same layer. Assume that the vectors \( v \) and \( h \) are the states of the explicit element in the visible layer and the implicit element in the hidden layer, respectively. Thus, \( v_i \) is the state of the \( i \)th explicit element and \( h_j \) is the state of the \( j \)th implicit element. Based on a given set of \((v, h)\), its energy function can be expressed as

\[
E(v, h; \theta) = - \sum_{i} a_i v_i - \sum_{j} b_j h_j - \sum_{i,j} v_i h_j
\]

where \( \theta \in \{W_{ij}, a_i, b_j\} \) is the model parameter of RBM, \( a \in \mathbb{R}^n \) is the bias of the explicit element, \( b \in \mathbb{R}^m \) is the bias of the implicit element, and \( W \in \mathbb{R}^{n \times m} \) is the connection weight matrix between the explicit element and the implicit element.

For each state \((v, h)\) of the visible layer and the hidden layer, the joint probability distribution is expressed as

\[
P(v, h; \theta) = \frac{e^{E(v, h; \theta)}}{Z}
\]

where \( Z \) is the partition function, which is the sum of the energy functions in all states of the visible layer and the hidden layer. The expression of \( Z \) is

\[
Z = \sum_n e^{-E(v, h; \theta)}
\]

The edge distribution of the joint probability distribution \( P(v, h; \theta) \) is as follows:

\[
P(v; \theta) = \sum_h P(v, h; \theta) = \sum_h \frac{e^{E(v, h; \theta)}}{Z}
\]

Whether in a visible or hidden layer, the nodes in the same layer are independent of each other. Given the state \( v \) of the visible layer, the probability that the \( j \)th element \( h_j \) of the hidden layer is turned on can be expressed as

\[
P(h_j = 1 | v; \theta) = \sigma \left( b_j + \sum_{i} W_{ij} v_i \right)
\]

Given the state \( h \) of the hidden layer, the probability that the \( i \)th element \( v_i \) of the visible layer is turned on can be expressed as

\[
P(v_i = 1 | h; \theta) = \sigma \left( a_i + \sum_{j} W_{ij} h_j \right)
\]

In (5) and (6), \( \sigma(x) \) is the activation function, and its common expression is

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]

It is necessary to determine the criteria for activation and start-up by setting a threshold [19]. The threshold settings are as follows:

\[
h_j = \begin{cases} 
1, & p(h_j = 1 | v) > \xi \\
0, & p(h_j = 1 | v) < \xi 
\end{cases}
\]

where \( \xi \) represents the probability threshold that discriminates the state of the hidden-layer neurons (on or off). The range of \( \xi \) is from 0.5 to 1 [19].

The process from (5) to (8) is a Gibbs sampling process. Given a training sample, the RBM is trained to adjust the parameter \( \theta \) so that the probability distribution under the control of \( \theta \) is as close as possible to the distribution of the training data. The larger the \( P(v; \theta) \), the greater the probability that the visible layer is turned on. That is, the hidden layer has fully learned the characteristics of the data in the visible layer. Therefore, maximising \( P(v; \theta) \) can be used when training the RBM. According to the principle of gradient
rizing, the maximum value of \( P(v|\theta) \) can be obtained by adjusting the connection weight \( \theta \). The logarithm derivative of \( P(v|\theta) \) can be obtained as

\[
\frac{\partial \log P(v|\theta)}{\partial \theta} = \left( \frac{\partial E(v, h|\theta)}{\partial \theta} \right)_{P(h,v|\theta)} - \left( \frac{\partial E(v, h|\theta)}{\partial \theta} \right)_{P(v,h|\theta)}
\]

where \( \langle \cdot \rangle_p \) is the mathematical expectation of the distribution \( P \).

Due to the existence of the partition function \( Z \), solving the joint probability distribution is complicated. In order to rapidly train the RBM, the CD algorithm was proposed by Hinton [19]. The CD algorithm is a more effective learning algorithm compared with Gibbs sampling. For each training sample, the probability distributions of \( P(h|v, \theta) \) and \( P(v, h|\theta) \) are denoted by \( q \) and \( p \), respectively. The partial derivatives of the log-likelihood function with respect to the visible layer and the hidden layer.

\[
\frac{\partial \log P(v|\theta)}{\partial W_{ij}} = \langle v^i h^j \rangle_q - \langle v^i h^j \rangle_h
\]

\[
\frac{\partial \log P(v|\theta)}{\partial a_i} = \langle v^i \rangle_q - \langle v^i \rangle_h
\]

\[
\frac{\partial \log P(v|\theta)}{\partial b_j} = \langle h^j \rangle_q - \langle h^j \rangle_h
\]

Since the computation of \( \langle \cdot \rangle_q \) is both time-consuming and labour-intensive, the number of Gibbs samples is reduced to \( k \) according to the CD-\( k \) criterion and the updating formula of the parameter set \( \theta \in \{ W_{ij}, a_i, b_j \} \) is

\[
W_{ij} = W_{ij} + \eta (\langle v^i h^j \rangle_q - \langle v^i h^j \rangle_h)
\]

\[
a_i = a_i + \eta (\langle v^i \rangle_q - \langle v^i \rangle_h)
\]

\[
b_j = b_j + \eta (\langle h^j \rangle_q - \langle h^j \rangle_h)
\]

where \( \eta \) is the learning rate. A large number of studies have shown that setting the value of \( k \) as 1 can achieve a smooth distribution of the visible layer and the hidden layer.

In the CD-\( k \) algorithm, each RBM requires multiple iterations, and the parameters are not necessarily updated in the same direction after each iteration. Therefore, a fixed learning rate will cause the ‘premature’ phenomenon or difficulty converging. In this paper, the ADBN algorithm was designed by modifying the fixed learning rate to the adaptive learning rate. The adaptive learning rate was constructed according to the similarities and differences of the parameter-update directions between two consecutive iterations of the RBM training process. The update mechanism of the adaptive learning rate is

\[
\eta = \begin{cases} 
\eta_0, & \Delta > 0 \\
\rho \eta, & \Delta < 0 
\end{cases}
\]

where \( \Delta = \langle v^i h^j \rangle_q - \langle v^i h^j \rangle_h \cdot \langle v^i h^j \rangle_q - \langle v^i h^j \rangle_h \), and \( \alpha \) and \( \beta \) are the increasing coefficient and the decreasing coefficient of the learning rate, respectively. \( \Delta \) is the product of the parameter variations in two successive RBM iterations. If the parameters are updated in the opposite direction after two consecutive iterations, the learning rate will decrease. If the parameters are updated in the same direction after two iterations, the learning rate will increase.

According to (8), the probability that the hidden-layer neurons being turned on will decrease when \( \zeta \) increases. Then, the update direction of the two weights is more likely to be consistent after two consecutive Gibbs iterations [20–22]. It can be obtained that

\[
P((f_2 f_1) - (f_2 f_1)) \times ((f_2 f_1) - (f_2 f_1)) > 0) \propto \xi
\]

where \( f_1 \) denotes the state of the input layer. \( f_2 \) denotes the reconstructed state of input layer after \( k \) iterations. \( f_3 \) denotes the state of the hidden layer obtained by \( f_1 \). \( f_4 \) denotes the state of hidden layer after \( k \) iterations.

According to (13)–(16), it can be obtained that

\[
\alpha \propto P((f_2 f_1) - (f_2 f_1)) \times ((f_2 f_1) - (f_2 f_1)) > 0
\]

It can be seen from (17) and (18) that the mathematical relationship between \( \zeta \) and \( \alpha \) is \( \alpha \propto \xi \). In addition, the weight is updated once during a Gibbs sampling process, while the binary sampling of the intermediate state is performed twice. The updated weight in each sampling process is proportional to the state-sampling value. Then, the approximate relationship between \( \zeta, \alpha, \) and \( \beta \) is obtained as follows: \( \alpha \propto \zeta, \beta \propto \xi \). Thus, the range of \( \alpha \) was set as \([1, 2]\), and the range of \( \beta \) was set as \([0, 1]\). When the learning rate adaptively changes according to (16), the error correction signal in the supervised learning process will change adaptively according to the similarities and differences between two consecutive update directions. Thus, the convergence speed and accuracy of the CD algorithm are improved.

### 3 Fault diagnosis based on ADBN model and time–frequency characteristics

#### 3.1 Data sources

The fault data used in this paper includes: (i) the online monitoring data of OTLs, (ii) the tripping report of OTLs, and (iii) the fault current travelling wave data of OTLs. The data was provided by the Electric Power Research Institute (EPRI) at Shandong, Shanghai, Guangdong and Anhui. The OTL voltage levels varied from 110 to 750 kV. The fault travelling waves were collected by the distributed system installed on the substation and the transmission line. The tripping report covered faults from 2009 to 2016. The typical travelling waves of different faults are shown in Fig. 2. The travelling-wave data is collected by the fault detector installed on OTLs, and the sampling frequency is 1 MHz. The fault-cause is determined by the subsequent off-line processing of the travelling-wave data. Each group of travelling-wave data corresponds to a trip report. The files that contain the travelling-wave data can be read by MATLAB (matrix laboratory). The corresponding label of the wave data can be obtained from the tripping report.

Based on the above data, the fault sample set of OTLs was composed. As shown in Table 1, a total of 2734 typical fault samples were collected. The fault causes were divided into nine classes: tree-caused fault (TF), mountain fire-caused fault (MFF), flotage-caused fault (FF), icing-caused fault (IF), windage yaw-caused fault (WYF), external damage-caused fault (EDF), animal-caused fault (AF), shielding failure (SF), and back flashover (BF). The sample set was divided into a training set and a testing set. The number of training samples was set to 80% of the total number of samples.

#### 3.2 Selection of characteristic parameters

The transient current travelling wave contains a wealth of fault information, such as polarity, amplitude, fault time, fault location, transition resistance and so on. These pieces of information can provide an important basis for fault diagnosis of OTLs. For different fault types, there are some differences in the time–frequency characteristics of travelling waves. Based on the mechanism of different fault types, this paper selected the appropriate time–frequency characteristic parameters as the input of ADBN model.

Lighting strikes of OTLs are divided into two fault causes, SF and BF. The fault currents of these two types of faults have short wave-tail times. Whether SF or BF, the fault current flowing through the line is mainly composed of two parts. One part is the lightning current that directly enters the line after the shunt, and the other part is the lightning current that flows into the ground and then reflects into the line. The over-voltage of BF has a large
amplitude and steep wave head. Additionally, there are large differences between the three-phase voltages of BF. After the fault phase is struck by lightning, the voltage rapidly rises, causing insulator flashover. Due to the phase-to-phase coupling and corona effects, the non-fault phase voltages fluctuate with high frequency in response to the fault phase voltage, and the voltage returns to normal after the end of the oscillation. Compared with BF, the lightning current is lower during SF, and thus the over-voltage amplitude of the SF is smaller than BF. Furthermore, the lightning current of SF and BF fault also flow in different directions before the flashover of the insulator. In the event of SF, the lightning current is injected into the lightning strike point. Except for the part flowing along the conductor, most of the lightning current is discharged to the tower after the insulation breakdown. In the event of BF, the lightning current is evacuated into two parts: one part flows from the top of the tower to the corner of the tower, and the other part is discharged to the line due to insulation breakdown. In addition, before the insulation breakdown in the event of BF, the lightning current flows through the lightning conductor first, and a reverse polarity pulse is induced on the faulty line. Based on the above analysis, it can be seen that there is a certain difference between the fault travelling wave of BF and SF.

The wave-tail time of the lightning strike current is <40 μs, and the measured results are generally within 20 μs. On the other hand, the current travelling wave has a longer wave-tail time and smaller amplitude compared with the non-lightning strike. Therefore, the lightning strike fault and non-lightning strike fault can be distinguished based on the amplitude and half wavelength of the travelling wave. For non-lightning faults, the fault cause can also be determined based on the characteristics of the transient travelling wave. Based on the analysis of real on-site fault data and fault mechanisms, the following conclusions can be drawn:

(i) TFs are caused by the contact or insufficient distance between the branches and the wires. The falling edge of the wave head is slower, and the rising edge is steeper than other high-impedance grounding faults. There is intermittent flashover before the main discharge peak. The amplitude of the initial travelling wave is small, and the value can be as low as the ampere level. (ii) MFFs are mainly caused by the thermal ionisation of air. Both the rising edge of the wave head and the falling edge of the wave tail are slow. The amplitude of the initial travelling wave is small, and there is obvious pre-discharge characteristic. (iii) The cause of FF is due to the suspension of the floaters on the conductor, ground wire, or tower so that the potential between the wire and ground is less than the safe value. The rising edge of the wave is steep. The amplitude of the travelling wave is large, and the value can reach several kilo-amperes. The waveform often exhibits bifurcation phenomenon. (iv) The reason for the occurrence of IF is that the insulator surface is covered with ice and the insulation performance is reduced. The waveform has low harmonic content before flashover. The amplitude of the initial travelling wave is large, and the rising edge is steep. However, the amplitude is smaller than other metallic short-circuit faults. (v) WYFs are caused by strong winds that reduce the distance between the wire and the tower or the lightning conductor to less than the safety value. The degree of waveform similarity is high, and the reflected wave is obvious. The amplitude of the initial travelling wave is large. (vi) The EDF is caused by the short distance between the crane and the conductor. The fault presents a metallic short-circuit feature. The rising edge of the wave head is steep, and the falling edge of the wave tail is

<table>
<thead>
<tr>
<th>Fault cause</th>
<th>Total samples</th>
<th>Training samples</th>
<th>Test samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF</td>
<td>301</td>
<td>241</td>
<td>60</td>
</tr>
<tr>
<td>MFF</td>
<td>164</td>
<td>131</td>
<td>33</td>
</tr>
<tr>
<td>FF</td>
<td>191</td>
<td>153</td>
<td>38</td>
</tr>
<tr>
<td>IF</td>
<td>219</td>
<td>175</td>
<td>44</td>
</tr>
<tr>
<td>WYF</td>
<td>137</td>
<td>110</td>
<td>27</td>
</tr>
<tr>
<td>EDF</td>
<td>774</td>
<td>619</td>
<td>155</td>
</tr>
<tr>
<td>AF</td>
<td>205</td>
<td>164</td>
<td>41</td>
</tr>
<tr>
<td>SF</td>
<td>396</td>
<td>317</td>
<td>79</td>
</tr>
<tr>
<td>BF</td>
<td>345</td>
<td>276</td>
<td>69</td>
</tr>
<tr>
<td>total</td>
<td>2734</td>
<td>2186</td>
<td>546</td>
</tr>
</tbody>
</table>

Fig. 2 Typical waveforms of different faults
the steepest in non-lightning faults. The amplitude of the initial travelling wave can be as high as several kilo-amperes. (vii) The AF is mainly caused by the short distance between the conductor and the animals or the nests, leading to the short circuit of the insulators. The rising edge of the wave is slow. The amplitude of the waveform is small, but it is larger than other high-impedance grounding faults. The waveform contains high-frequency harmonic components before and after the flashover. According to the above analysis, the waveforms of different fault types have certain differences in amplitude, half-wavelength, steepness, energy in different periods, and energy in different frequency bands. The wavelet transform (WT) [23] has the adaptability for time–frequency window, which can fully reflect the correlation between the fault occurrence time and the fault cause.

In this paper, we utilise the cubic B-spline wavelet to decompose and reconstruct the travelling wave signal. In addition, there is a certain correlation between the fault occurrence time and the fault cause. For example, (i) AF often occurs in the daytime in spring or summer, (ii) SF and BF often occur during summer nights, (iii) TF often occurs in the autumn, and (iv) IF often occurs in the winter. In summary, the characteristic parameters were selected as follows:

\[
M_i = \int_0^{t_{i,0}} f(t) \, dt, \quad M_{\text{total}} = \int_0^{t_{\text{total}}} f(t) \, dt,
\]

\[
M_{\text{mid}} = \int_0^{t_{\text{mid}}} f(t) \, dt, \quad M_i = \int_0^{t_i} f(t) \, dt,
\]

\[
T_1 = \frac{T_{\text{total}}}{M_{\text{total}}}, \quad T_2 = \frac{M_{\text{mid}}}{M_{\text{total}}}, \quad T_3 = \frac{I_m}{U_{\text{line}}(t_{i,0} - t_i)},
\]

\[
T_4 = \frac{I_m}{U_{\text{line}}(t_{i,0} - t_i)}, \quad T_5 = \frac{E_1}{\sum_{j=1}^{n} E_j}, \quad T_6 = \frac{e_j}{\sum_{j=1}^{n} e_j}, \quad T_7 = \frac{M_i}{M_k},
\]

\[
T = [T_0, T_1, T_2, T_3, T_4, T_5, T_6, T_7]
\]

where \( T \) is the characteristic vector and \( f(t) \) is the travelling-wave signal acquired in the fault phase. \( T_{i,0} \) is a three-digit binary number that indicates the fault occurrence time. The first two digits indicate the seasons (00 – spring, 01 – summer, 10 – autumn, and 11 – winter), and the last digit indicates the day and night (0 – day, 1 – night). \( t_0 \) is the start time of the selected waveform. \( t_{\text{total}} \) is the end time of the selected waveform. \( t_i \) is the start time of the initial travelling wave. \( t_{i,0} \) is the end time of the initial travelling wave. \( t_{\text{mid}} \) is half of the total sampling time. \( t_i \) is the characteristic moment and is a self-defined value that is between \( t_{\text{mid}} \) and \( t_{\text{total}} \). It is mainly used to compare the time–frequency characteristics before and after the initial wave head. \( t_{i,0} \) is the peak value of the initial wave head. \( t_i \) is the time corresponding to the peak value. \( t_{i,0} \) is the time corresponding to the half-peak value \((t_{i,0} < t_i)\). \( n \) is the wavelet decomposition level. \( E_j \) is the energy in the \( j \)th frequency band after WT and its time range is \([t_{i,j}, t_{i,j+1}]\). \( e_j \) is the energy in the \( j \)th frequency band after WT and its time range is \([t_0, t_k] \).

line voltage. \( M_k \) is the maximum ratio of peak steepness to line voltage. In this paper, \( t_0 \) and \( t_{\text{total}} \) are 0 and 1200 \( \mu \text{s} \), respectively.

According to (19), it can be seen that \( T_1 \) and \( T_2 \) reflect the time-domain energy characteristics in different periods. \( T_3 \) and \( T_4 \) reflect the steepness and amplitude characteristics in different periods. \( T_5 \) and \( T_6 \) reflect the frequency-domain energy characteristics in different periods.

The softmax classifier was used for the output of ADBN. For \( n \) independent classifications, the corresponding labels for output were set to 1, 2, ..., \( n \). Each label corresponding to one category. The maximum output value of \( n \) outputs (normalised between 0 and 1) was set to 1, and the rest of the output values were set to 0. When the output value is 1, the corresponding label represents the ideal category. In this paper, the labels were set to correspond with the fault causes as follows: 1 – TF, 2 – MFF, 3 – FF, 4 – IF, 5 – WYF, 6 – EDF, 7 – AF, 8 – SF, 9 – BF.

### 3.3 Steps of fault diagnosis model

The procedures of the fault diagnosis method based on the ADBN model and time–frequency characteristics of the current travelling wave are as follows:

i. Extract the fault travelling waves of different fault types.

ii. Calculate the value of the time–frequency characteristic \( T = [T_0, T_1, T_2, T_3, T_4, T_5, T_6, T_7] \) and use \( T \) as the input characteristic.

iii. Utilise the improved CD algorithm to perform pre-training of DBN model and adaptively adjust the learning rate. Fine-tune the network parameters by back propagation algorithm and stochastic gradient descent.

iv. Utilise the trained parameters to perform diagnostic tests on the data in the testing set.

The overall procedure of the proposed diagnosis model is shown in Fig. 3.

### 4 Case studies and analysis

#### 4.1 Model structure and parameter setting

In this paper, we use the simulation software MATLAB (matrix laboratory) R2014b to implement the ADBN algorithm. The CPU type is ‘Intel Core i3-4160 @ 3.60 GHz’. The program of fault identification algorithm is completed on MATLAB. Firstly, the initial data (including the travelling-wave data and the labels) is imported into MATLAB’s storage units. Secondly, the program calculates the input characteristics based on the wave data. Then, the values of input characteristics and the labels are stored in another unit. Finally, the ADBN model is trained based on the input characteristics and labels. When the trained model is input into new data, the corresponding recognition results can be obtained.

The structure of the ADBN model is shown in Fig. 4. According to the sample data in Table 1, \( T = [T_0, T_1, T_2, T_3, T_4, T_5, T_6, T_7] \) was used as the input characteristic parameter to test the ADBN model. The learning rate of the weight, the learning rate of the visible-layer bias term, and the learning rate of the hidden-layer bias term were set as 0.1. The weight attenuation coefficient was 0.0008. In order to improve the contradiction between the convergence speed and the instability of the back propagation algorithm, the initial momentum term was set to 0.5, and the momentum term was set to 0.9 when the reconstruction error was in a steady-increase state. The initial connection value was a random number that obeyed the normal distribution \( N (0, 0.01) \). The biases of the hidden layer and the visible layer were set as 0. The value of the increase coefficient \( \alpha \) was 1.4, and the value of the decrease coefficient \( \beta \) was 0.7. First, the self-defined time \( t_k \) was set as 850 \( \mu \text{s} \), and the wavelet decomposition level was set as 4. Empirically, the number of training samples is between 50 and 90% of the total sample size. For ease of calculation, the ratio of training samples to test samples is generally 50/50%, 60/40%,
The number of the training cycles was 550, the input was 8 characteristic parameters, and the output was 9 fault causes. The network structure was 10-10-10-10-10-9. With this network structure, the relationships between the self-defined time $t_k$, the wavelet decomposition level $n$, and the diagnostic accuracy rate are shown in Fig. 6. It can be seen that the effect of the diagnosis model was better when $t_k$ was in the range of 900–950 μs and $n$ was in the range of 5–8. Therefore, $t_k$ was adjusted to 900 μs, and $n$ was adjusted to 6.

With the above network structure and parameters, the ADBN model was used to perform fault diagnosis on the test samples, and the obtained confusion matrix is shown in Fig. 7. As can be seen in Fig. 7, the total recognition accuracy of the ADBN model reached 95.4%. The recognition rate of SF was the highest (77 out of 79 cases were correctly identified). The recognition rate of FF was relatively low (35 out of 38 cases were correctly identified).

### 4.2 Comparison of different models and parameters

In order to evaluate the performance of the ADBN model, the fault causes were identified by using the SVM, the back propagation NN (BPNN), the convolutional NN (CNN), the DBN model, and the ADBN model. The sigmoid kernel function was used in the SVM model. The cross-validation method was used to obtain the penalty factor. The crossover probability was 0.8, and the mutation probability was 0.05. The optimal penalty factor was 10. The two kernel parameters of the sigmoid kernel function were 2 and 0.4.

The BPNN model consisted of an input layer, hidden layer, and an output layer. The number of neurons in each layer was 10, 10, and 9, respectively. The learning rate was 0.01, and the learning cycle was 800. The network structure of the CNN model was M-6C-1S-6C-2S-N. M, N, C, and S were the training sample set, output set, convolutional layer, and downsampling layer, respectively. Both convolution kernels of the convolutional layer were 2×2. The learning cycle of the CNN model was 500. The network structure, training cycle, and parameter settings of the DBN model were consistent with the ADBN model.

The selected characteristic vector $T = [T_0, T_1, T_2, T_3, T_4, T_5, T_6, T_7]$ contained both the time and frequency domain characteristics of the fault travelling waves. $T_1, T_2, T_3, T_4,$ and $T_7$ were the time-domain parameters. $T_5$ and $T_6$ were the frequency-domain parameters. In order to verify the rationality of the proposed characteristic vector $T$, different characteristic vectors were tested as the inputs of the SVM, BPNN, CNN, DBN, and ADBN models. The characteristic vectors used for comparison were $\mathcal{I} = [T_0, T_1, T_2, T_3, T_4, T_7]$ (only including time-domain characteristics), $\mathcal{O} = [T_0, T_5, T_6]$ (only including frequency-domain characteristics), and $\mathcal{P} = [T_0, T_1, T_2, T_3, T_6, T_7]$ (including partial time–frequency characteristics). The structure and parameters of different diagnostic models will change as the input vector changes.

The fault diagnosis results of different models and characteristic vectors are shown in Table 2. The recognition accuracy listed in Table 2 is the means in 15 tests. It can be seen that the ADBN model had a higher recognition accuracy compared with other models. The ADBN model with vector $T$ achieved the best fault diagnosis, the recognition accuracy rates of training and testing sets were 70/30%, 80/20%, or 90/10%. In this paper, the dataset was divided into a training dataset with 80% samples and a testing dataset with 20% samples [24–26]. The input set of the training sample was a $10 \times 2186$ matrix ($T_0$ occupies three rows), and the output set was a $9 \times 2186$ matrix. The input set of the test sample was a $10 \times 546$ matrix, and the output set was a $9 \times 546$ matrix.
were 96.2 and 95.4%, respectively. With the input vector $T$, the training accuracy rate of the ADBN model exceeded those of the SVM, BPNN, CNN, and DBN models by 7.9, 10.5, 4.9, and 2.4%, respectively. Additionally, the testing accuracy rate of the ADBN model exceeded those of the SVM, BPNN, CNN, and DBN models by 8.5, 10.3, 4.9, and 2.5%, respectively. Compared to the DBN model, the accuracy of the ADBN model was improved. In addition, the accuracy of the ADBN model with the $T$ vector significantly increased compared to the ADBN model with the $I$, $O$, or $P$ vectors: the training accuracy rate increased by 14.8, 24, and 8.7%, respectively, and the testing accuracy rate increased by 15.2, 24.2, and 9.1%, respectively. The results show that the proposed characterisation parameters are reasonable, and they can fully describe the fault information contained in different travelling waves.

The total running time of the algorithm mainly includes the calculating time of the input characteristics and the training time of the ADBN model. With the input vectors $I$, $O$, $P$, and $T$, the corresponding times for the characteristic calculation are 17, 15, 21, and 24 s, respectively. The training times corresponding to different models and vectors are listed in Table 3. The $T$ vector contains more fault information, which makes the model more computationally expensive and time-consuming. When the $T$ vector was used for the input characteristics, the training time of the SVM, BPNN, CNN, DBN, and ADBN models were 241, 255, 223, 248, and 206 s, respectively. Although the structure of the ADBN model is complex, it adopts the improved CD algorithm to perform layer-by-layer pre-training and adaptively adjusts the learning rate according to each iteration direction of the parameters. In this way, the problems of under-learning and falling into local optimum can be avoided. Also, the speed of model convergence is significantly improved. Moreover, when we utilise the trained model to perform diagnostic tests on new data, the testing time is <5 s.

### 4.3 Influence of sample sizes

In order to study the effect of the sample size, the number of samples was set to 1000, 1500, 2000, and 2500. The ratio of the number of training samples to the number of testing samples remained unchanged. Using $T$ as the characteristic vector, the obtained recognition accuracy is shown in Table 4. The accuracy rate of the ADBN model gradually increased as the number of samples increased. As the number of samples increased, the characteristic information extracted from the ADBN model was more abundant, and the internal relationship between the electrical quantity and the fault cause could be more accurately characterised. The fault transient electrical quantities of OTLs vary with line length, voltage level, fault location, and operating conditions. With the increase of OTL faults, the sample set will be expanded and improved, which helps to improve the diagnostic effectiveness of the ADBN model.

### 4.4 Influence of proportion of training samples

To investigate the effect of the proportion of training samples, the ratio of training samples to test samples was set to 50/50%, 60/40%, 70/30%, and 80/20%, respectively. The testing sample sizes remained unchanged. Using $T$ as the characteristic vector, the obtained recognition accuracy is shown in Table 5. It can be seen that the recognition accuracy of ADBN model slightly increased as the proportion of training samples increased. When the proportion of training samples is only 50%, the recognition accuracy can still reach to 93.7%. The results show that the proposed method can be applied to different sample proportions.

### 4.5 Influence of sampling rates

In order to study the effect of the sampling rate, the recognition accuracy for different sampling frequencies is investigated (including 200 kHz, 300 kHz, 500 kHz, 1 MHz, and 2 MHz). The transient travelling-wave signal is generated instantaneously by the fault, and the travelling-wave spectrum is mainly in the range of 10–100 kHz. In some cases, the signal frequency may be higher [27]. To accurately capture the high-frequency transients, the

#### Table 2 Results of different models and characteristic vectors

<table>
<thead>
<tr>
<th>Model</th>
<th>$I$</th>
<th>$O$</th>
<th>$P$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>74.3</td>
<td>65.3</td>
<td>80.6</td>
<td>79.5</td>
</tr>
<tr>
<td>BPNN</td>
<td>72.5</td>
<td>60.7</td>
<td>77.4</td>
<td>76.8</td>
</tr>
<tr>
<td>CNN</td>
<td>73.8</td>
<td>64.1</td>
<td>81.3</td>
<td>80.6</td>
</tr>
<tr>
<td>DBN</td>
<td>78.5</td>
<td>70.2</td>
<td>84.7</td>
<td>84.2</td>
</tr>
<tr>
<td>ADBN</td>
<td>81.4</td>
<td>72.2</td>
<td>87.5</td>
<td>86.4</td>
</tr>
</tbody>
</table>

#### Table 3 Training times (s) corresponding to different models and vectors

<table>
<thead>
<tr>
<th>Model</th>
<th>$I$</th>
<th>$O$</th>
<th>$P$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>168</td>
<td>108</td>
<td>163</td>
<td>241</td>
</tr>
<tr>
<td>BPNN</td>
<td>181</td>
<td>117</td>
<td>174</td>
<td>255</td>
</tr>
<tr>
<td>CNN</td>
<td>159</td>
<td>96</td>
<td>152</td>
<td>223</td>
</tr>
<tr>
<td>DBN</td>
<td>172</td>
<td>113</td>
<td>165</td>
<td>248</td>
</tr>
<tr>
<td>ADBN</td>
<td>145</td>
<td>79</td>
<td>137</td>
<td>206</td>
</tr>
</tbody>
</table>
required sampling frequency should be much higher than the highest frequency of the signals. Since some selected characteristics ($T_3$ and $T_6$) are associated with the frequency, the recognition results are related to the sampling rate. With the input vector $T$, the training and testing accuracy rate of the ADBN model are shown in Table 6. The recognition accuracy of ADBN model slightly increased as the sampling rate increased. As the sampling rate increased, the characteristic information extracted from the model can be more representative of the original travelling-wave signals has reached the MHz level. The results are shown in Table 6. The recognition accuracy of ADBN model can still reach to 91.9%. At present, the sampling frequency for classification model used by these methods with our proposed model. In this way, the existing methods can be compared intuitively with the proposed methods. Some studies have focused on a method to directly classify and identify the fault root cause (such as the methods presented in [2–4]). Thus, we can directly compare these methods with our methods.

In [2], the study uses the CN2 induction algorithm to determine rules to classify four causes of faults (lightning, tree, cable and animal). The CN2 rule induction algorithm contributes to induce an ordered list of classification rules from a set of classified observations. In [3], the calculation formula for extracting characteristic parameters according to fault waveform data was proposed, the recognition logic was established by taking multi-parameters fusion as a basis, and then automatic recognition of five types of faults (animal, equipment, lightning, tree and vehicle) caused by different factors was realised. In [4], the classification characteristics based on the waveform and external environmental characteristics have been identified to develop single-nearest-neighbour (1-NN) classifiers with F-measure. The method can be adopted to identify the underlying cause (lightning, tree, pollution, animal and others) of transmission-line faults. The comparison results are shown in Table 7. It can be seen that the proposed method has certain advantages in identifying the number of fault types and recognition accuracy.

### 4.7 Adaptability analysis of model

To verify the adaptability of the ADBN model, this paper selected 15 fault cases beyond the sample set. The trained network in Section 4.1 was used to identify the fault cause, and the results are listed in Table 8. The trained ADBN model correctly diagnosed the fault causes of OTLs. Although the cases were not within the sample set, the ADBN model still accurately identified the fault causes with different voltage levels and line lengths. This further proves the strong generalisation ability of the model. In addition, if new fault cases are added to the original sample set, the diagnostic effect of the model will be further improved.

### 5 Conclusions

This paper presents a fault-cause identification method for OTLs based on an ADBN model and time-frequency characteristics of the travelling wave. The method is essentially a data-driven approach, which requires a relatively low sampling frequency and has high adaptability to different sampling rates and fault types. The proposed method has certain advantages in identifying the number of fault types and recognition accuracy. The method can be adopted to identify the underlying cause (lightning, tree, pollution, animal and others) of transmission-line faults. The comparison results are shown in Table 7. It can be seen that the proposed method has certain advantages in identifying the number of fault types and recognition accuracy.

### 4.6 Comparison with different methods

In this section, the proposed method is compared with other representative methods. At present, the existing methods mainly classify the faults from the perspective of line selection (such as the methods in [8, 10, 12]). Thus, we did not directly compare these methods with the proposed method. Instead, we replaced the classification model used by these methods with our proposed model. In this way, the existing methods can be compared intuitively with the proposed methods. Some studies have focused on a method to directly classify and identify the fault root cause (such as the methods presented in [2–4]). Thus, we can directly compare these methods with our methods.

In [2], the study uses the CN2 induction algorithm to determine rules to classify four causes of faults (lightning, tree, cable and animal). The CN2 rule induction algorithm contributes to induce an ordered list of classification rules from a set of classified observations. In [3], the calculation formula for extracting characteristic parameters according to fault waveform data was proposed, the recognition logic was established by taking multi-parameters fusion as a basis, and then automatic recognition of five types of faults (animal, equipment, lightning, tree and vehicle) caused by different factors was realised. In [4], the classification characteristics based on the waveform and external environmental characteristics have been identified to develop single-nearest-neighbour (1-NN) classifiers with F-measure. The method can be adopted to identify the underlying cause (lightning, tree, pollution, animal and others) of transmission-line faults. The comparison results are shown in Table 7. It can be seen that the proposed method has certain advantages in identifying the number of fault types and recognition accuracy.
method. In the deep learning process, the ADBN model extracts the relevant information contained in the characteristic vector and represents the correlation between the time–frequency parameter and the fault cause. Through the feature learning of multiple faults, the characteristic expression of fault causes is extracted.

Based on the characteristics of different fault causes and real-world fault data, various time–frequency parameters of the travelling wave are used as the input of the model, and the labels corresponding to different fault causes are selected as the output. The selection of travelling-wave features is related to the identification results. In the paper, different travelling-wave features are proposed. The study concluded that the combined use of time–frequency features can help to improve the accuracy rate.

The selection of ADBN parameters is also significant. This is due to the high impact of different ADBN parameters on the identification accuracy. In this study, the most suitable parameters are determined based on the comparison of multiple experiments. According to case studies, the identification accuracy based on the ADBN model with time–frequency parameters is higher than that of other traditional models. In addition, the increase in the sample set helps to optimise the model parameters. In the future, more fault cases will be expanded to the original fault sample set, which will allow the ADBN model to more effectively represent the internal relationship between the characteristic parameter and the fault cause.

6 Acknowledgments

This work has been supported by the National Natural Science Foundation of China under grants nos. 51307109 and 51477100, as well as the science and technology project of State Grid Corporation in China.

7 References

[27] Ge, Y.: ‘New types of protective relaying and fault location theory and techniques’ (Xian Jiaotong University Press, Xian, China, 2007).