IET Generation, Transmission & Distribution

Special Issue: Advanced Data-Analytics for Power System Operation, Control and Enhanced Situational Awareness

Knowledge-based artificial neural network for power transformer protection

ISSN 1751-8687
Received on 26th March 2020
Revised 28th July 2020
Accepted on 6th August 2020
E-First on 15th September 2020
doi: 10.1049/iet-gtd.2020.0542
www.ietdl.org

Zongbo Li[1], Zaibin Jiao[1] ✉, Anyang He[1]

[1]*School of Electrical Engineering, Xi'an Jiaotong University, Xianning West Road No. 28, Xi'an 71000, People's Republic of China*

✉ *E-mail: jiaozaibin@mail.xjtu.edu.cn*

Abstract: Data-driven and artificial intelligence based transformer protection has attracted increasing attention but not been widely applied in the power system owing to the poor generalisation ability. In this study, a feature transferring method is proposed for a knowledge-based artificial neural network (ANN) to develop a transformer protection with an improved generalisation ability. Normally, power experts can reliably identify the running states based on the professional knowledge of only focusing on the unsaturated parts of equivalent magnetisation curve (voltage of magnetising branch-differential current). In order to imitate the power experts, the images of equivalent magnetisation curves whose saturated parts are removed are defined as source domain and the original samples are target domain. An ANN named as S:ANN is firstly trained through the source domain where the extracted features are equivalent to the professional knowledge. Then another ANN with the same structure as S:ANN is trained through the target domain and named as T:ANN. It is specially designed for T:ANN that adaptive layers are employed between S:ANN and T:ANN to reduce the feature differences. Finally, simulations and experiments reveal that the knowledge-based ANN namely the determined T:ANN shows a better generalisation ability through paying more attention to the unsaturated parts.

1 Introduction

Power data has been expanding continually along with the rapid development of smart grid in recent years. The advancement of artificial intelligence (AI) is also unarguably accelerating. Therefore, the research methods based on data-driven and AI become hotspots in the planning, operation, maintenance, and marketing of power grid. Meanwhile, the protection based on data driven and AI has also attracted rising attention [1–10].

Power transformer, as one of the critical equipment in the smart grid, plays an important role in energy conversion, power transmission, and grid recovery. It requires better reliability of the configured protection. How to reliably identify internal fault and inrush current is still the core issue of transformer protection. Generally, the single feature-based transformer algorithms [11–25], including second harmonic restraint, dead angle principle and waveform symmetry, cannot meet the reliability requirements in the complicated power grid. According to the research idea of multi-feature fusion, many protection algorithms integrating data driven and AI have been published for transformer protection. At present, the differential current is still the main research object, for instance: (i) it is directly used as an input to train machine learning algorithms, such as decision tree [26, 27], random forest [28], artificial neural network (ANN) [29–33], probabilistic neural network [34–36], radial basis neural network [37], and so on; (ii) the features that are extracted from the differential current by the tools, such as wavelet transform [38–44], Clarke transform [45], principal component analysis [46], and so on, are used as the inputs of machine learning algorithms; (iii) the running states are identified through pattern recognition methods such as fuzzy theory [32, 47, 48]; (iv) according to the theory of image recognition, mathematical morphology [49, 50] is used for identifying the running states, in addition, deep learning algorithms such as convolutional neural networks (CNNs) [51, 52] have also received attention in recent years. Besides the methods mentioned above, some scholars put forward the concept of equivalent magnetisation curve [53] whose several geometric features are extracted to train an support vector machine (SVM) for the identification of running states. However, it is difficult to apply the multiple feature-based methods in the engineering field owing to the following issues: (i) the reliability cannot be ensured due to the inconsistency of training process with professional knowledge; (ii) the recorded data of internal faults and transformer energisations is insufficient for the training process of machine learning algorithm; (iii) the running environment of on-site transformer is so complicated that it is impossible for the training samples to cover all the scenarios. Therefore, the generalisation ability of the determined model is poor for the mere application of a machine learning algorithm.

Integrating data and professional knowledge, this paper proposes a special feature transferring method to develop a knowledge-based ANN for reliable transformer protection. To be specific, the images of equivalent magnetisation curve whose saturated parts have been removed are used to train an ANN called S:ANN. S:ANN is equivalent to a power expert who only focuses on the unsaturated parts. Then another ANN called T:ANN is trained through the original images, where the adaption layers are employed in the hidden layers to reduce the feature differences with S:ANN. The determined T:ANN is the knowledge-based ANN that has a good generalisation ability based on the interaction of data and professional knowledge.

This paper is organised as follows: Section 2 firstly illustrates the related work based on equivalent magnetisation curve, then states the definition of professional knowledge, finally puts forward the features transferring method for the knowledge-based ANN. Section 3 presents the specific implementation method of transformer protection. Section 4 shows the better performance of knowledge-based ANN utilising large amounts of simulations and experiments and makes a comparison with the second harmonic restraint. Section 5 discusses the adaptability of knowledge-based ANN to several typical scenarios and future work. Section 6 is the conclusion.

2 Proposed interaction method of data and professional knowledge

2.1 Theoretical basis and related works

Equivalent magnetisation curve that describes the relationship between the voltage of magnetising branch and differential current

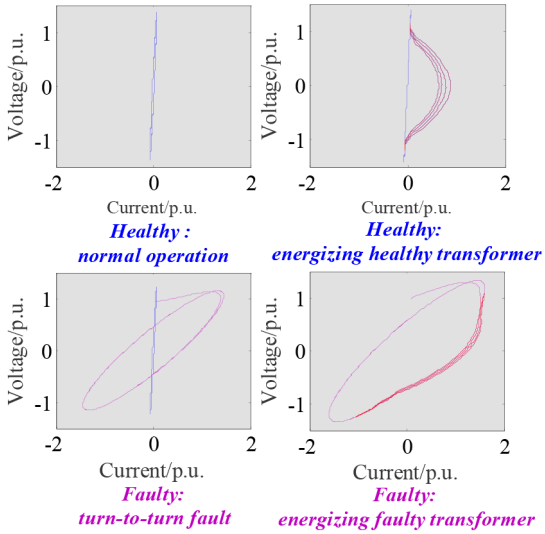

Fig. 1 *Equivalent magnetisation curve*

demonstrates the running states of the transformer, such as normal operation/external fault, internal fault, energising healthy transformer and energising faulty transformer. As shown in Fig. 1, the unsaturated parts coloured in blue and purple can reliably indicate whether the transformer is healthy; besides, the parts coloured in red parts depict the saturation characteristics of iron-core. At present, several works have been completed based on equivalent magnetisation curve for transformer protection, for instance, three geometric features extracted from the curves are used as input to an SVM, or the images of the curves are used to train a CNN. However, the determined SVM and CNN behave unstably because they fail to mine the essential features of the equivalent magnetisation curve. In fact, the simulation samples have the same essential features as the recorded ones.

Generally, power experts can reliably identify the running states through the professional knowledge which is defined as only focusing on the unsaturated parts of equivalent magnetisation curves in this paper. The purpose of this paper is to improve the generalisation ability by means of mathematically embedding this professional knowledge into the training process of ANN. Combining data and professional knowledge, the proposed knowledge-based ANN will adaptively focus on the unsaturated part of the equivalent magnetisation curve even with the limited training samples.

2.2 Proposed features transferring method

No matter whether the iron core is saturated, power experts will only focus on the unsaturated features Exp:F.

Comparatively, the saturated features will be inevitably introduced in the training process of the common ANN. As can be concluded that the features ANN:F extracted by ANN will exhibit different distributions from Exp:F. Consequently, the common ANN has a poor generalisation ability. Aiming at developing a knowledge-based ANN only focusing on the unsaturated parts, a features transferring method is proposed to reduce the feature differences between ANN:F and Exp:F, namely d(ANN:F, Exp:F) $\rightarrow$0.

2.2.1 Features transferring: A data window of 13 ms is adopted to deal with a binary classification task: healthy transformer labelled with '1', including energising healthy transformer, normal operation/external fault; faulty transformer labelled with '2', including energising faulty transformer, internal fault. The images of equivalent magnetisation curve whose saturation parts are removed manually are defined as source domain S, namely the unsaturated parts; correspondingly, the original samples containing both saturated and unsaturated parts are target domain T. The training process of knowledge-based ANN through the source and target domains has been shown in Fig. 2. To be specific firstly, an ANN defined as S:ANN is trained with the source domain S. The loss function E_{S} is

$$E_{\mathrm{S}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{1}$$

where y_i and $\hat{y}_i$ are the label and actual output, respectively; and N is the number of training samples. An optimal S:ANN is determined with a higher accuracy and smaller classification loss and the hidden layers' features Exp:F are saved. Define the features of lth hidden layer as Exp:F$_{\boldsymbol{l}}$.

Secondly, another ANN with the same structure as S:ANN is trained with target domain T, defined as T:ANN. Suppose that the features of lth hidden layer is ANN:F$_{\boldsymbol{l}}$. To reduce the feature differences d(ANN:F$_{\boldsymbol{l}}$, Exp:F$_{\boldsymbol{l}}$), an adaption layer is embedded between the hidden layers of T:ANN and S:ANN. Combining the adaption and classification losses, the loss function E_{T} of T:ANN is given:

$$E_{\mathrm{T}} = \delta_1\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \delta_2\sum_{i=1}^{N}\sum_{l=1}^{L} d(\mathrm{ANN{:}F}_l, \mathrm{Exp{:}F}_l) \tag{2}$$

where δ_i denotes the weight of classification loss or adaption loss. With the objective function of (2), T:ANN will pay more attention to the unsaturated parts of target domain T when the feature differences d(ANN:F, Exp:F) are as small as possible. Finally, the determined T:ANN is the knowledge-based ANN mentioned above, which will be used to construct reliable transformer protection.

2.3 Feature differences definition

From the definitions, any sample in the source domain must be a part of the corresponding sample in the target domain. T:ANN will only focus on the unsaturated parts when ANN:F is equal to Exp:F. Therefore, d(ANN:F$_{\boldsymbol{l}}$, Exp:F$_{\boldsymbol{l}}$) is formulated by Euclidean distance. Let

$$\mathrm{Exp{:}F}_l = \left[f_{\mathrm{S},il1}, f_{\mathrm{S},il2}, \ldots, f_{\mathrm{S},ilj}, \ldots, f_{\mathrm{S},iln}\right]^{\mathrm{T}},$$

$$\mathrm{ANN{:}F}_l = \left[f_{\mathrm{T},il1}, f_{\mathrm{T},il2}, \ldots, f_{\mathrm{T},ilj}, \ldots, f_{\mathrm{T},iln}\right]^{\mathrm{T}},$$

where n is the number of neurons, i denotes the ith sample, and f is the output of a neuron. The feature differences between ANN:F$_{\boldsymbol{l}}$ and Exp:F$_{\boldsymbol{l}}$ can be expressed as the following formula:

$$d(\text{ANN:F}_l, \text{Exp:F}_l) = \frac{1}{n}\sum_{j=1}^{n}\left(f_{\text{T},ilj} - f_{\text{T},ilj}\right)^2. \tag{3}$$

Therefore, the loss function of T:ANN can be reformulated:

$$E_{\text{T}} = \delta_1\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \delta_2\sum_{i=1}^{N}\sum_{l=1}^{L}\sum_{j=1}^{n}\left(f_{\text{T},ilj} - f_{\text{S},ilj}\right)^2. \tag{4}$$

The loss function (4) will be optimised through gradient descent algorithm.

2.3.1 Optimisation process of T: ANN: Take T:ANN with double hidden-layer as an example to illustrate the optimisation process of the loss function (4). As shown in Fig. 3, the inputs of neurons 1, 2 and 3 are a_{i11}, a_{i21} and a_{i31}, respectively, then

$$\begin{aligned} E &= E_1 + E_2 \\ &= \delta_1\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \delta_2\sum_{i=1}^{N}\sum_{l=1}^{2}\sum_{j=1}^{n}(f_{\text{T},ilj} - f_{\text{S},ilj})^2. \end{aligned} \tag{5}$$

The parameters' optimisations of the output layer are only affected by the classification loss, like w_3

$$\begin{aligned} w_3^{(t)} &= w_3^{(t-1)} - \frac{\delta_1}{N}\sum_{i=1}^{N}\frac{\partial E}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial a_{i31}}\frac{\partial a_{i31}}{\partial w_3^{(t-1)}} \\ &= w_3^{(t-1)} - \frac{2\delta_1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)g'(a_{i31})f_{\text{T},i21} \end{aligned} \tag{6}$$

where t denotes the iterative times and $g(x)$ denotes the activation function.

Based on the chain rule, the parameters optimisations of the second hidden layer will be affected by classification loss and adaption loss, like w_2

$$\begin{aligned} w_2^{(t)} &= w_2^{(t-1)} - \frac{\delta_1}{N}\Delta_1 - \frac{\delta_2}{N}\Delta_2 \\ \Delta_1 &= \sum_{i=1}^{N}\frac{\partial E_1}{\partial \hat{y}_i}\frac{\partial \hat{y}_i}{\partial a_{i31}}\frac{\partial a_{i31}}{\partial f_{\text{T},i21}}\frac{\partial f_{\text{T},i21}}{\partial a_{i21}}\frac{\partial a_{i21}}{w_2^{(t-1)}} \\ &= 2w_3^{(t-1)}\sum_{i=1}^{N}(\hat{y}_i - y_i)g'(a_{i31})g'(a_{i21})f_{\text{T},i11} \\ \Delta_2 &= \sum_{i=1}^{N}\frac{\partial E_2}{\partial f_{\text{T},i21}}\frac{\partial f_{\text{T},i21}}{\partial a_{i21}}\frac{\partial a_{i21}}{w_2^{(t-1)}} \\ &= 2\sum_{i=1}^{N}(f_{\text{T},i21} - f_{\text{S},i21})g'(a_{i21})f_{\text{T},i11}. \end{aligned} \tag{7}$$

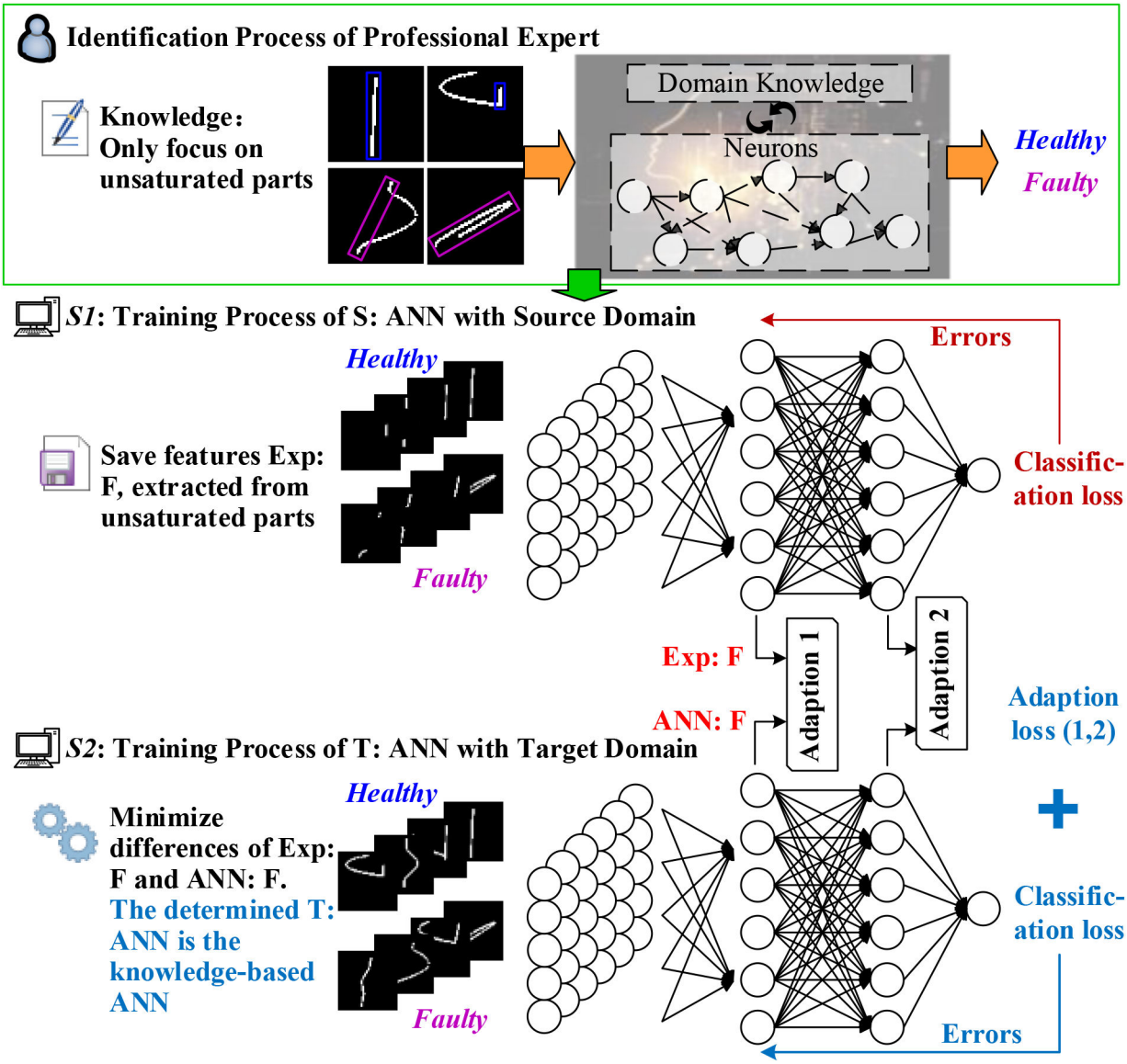

Fig. 2 *Training process of ANN*

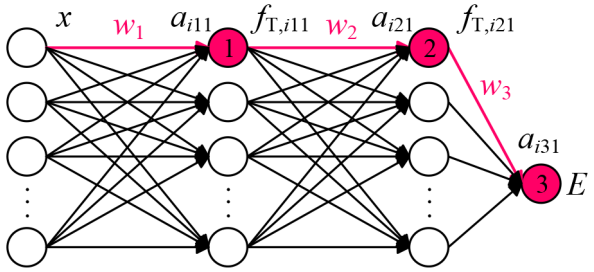

Fig. 3 *Example of T: ANN for optimisation illustration*

Similar as (7), the parameters optimisations of the first hidden layer will be affected by classification loss, and first and second adaption losses.

3 Transformer protection algorithm

3.1 Start-up criterion

To start reliably and rapidly, both voltage of magnetising branch and differential current are considered in start-up criterion:

$$|e(k+M)-e(k)| > e_{\text{set}} \tag{8}$$

where e denotes the voltage of magnetising branch, u, or the differential current, i; e_{set} is the threshold u_{set} or i_{set}; M is the samples number in one cycle; k denotes the kth sampling point. As the knowledge-based ANN has sufficiently considered the normal operation/external fault, the threshold can be as small as possible regardless of the sensitivity.

3.2 Operation criterion

Equivalent magnetisation curve of each phase is used as the input to the knowledge-based ANN once the start-up criterion (8) is met.

3.2.1 Data processing: Data processing involves: (i) data conversion from discrete sampling data to images; and (ii) dimension reduction of the images.

Data conversion: Equivalent magnetisation curve in the data window of 13 ms is normalised to be limited within a fixed range:

$$u \in [-1, 1],\ i \in [-1, 1].$$

It is notable that the features of the curves cannot be changed during the normalisation process. Therefore, the voltage and current are simultaneously normalised through the maximum and minimum values of the sequence $[u; i]$:

$$e' = 2 \times \frac{e - e_{\min}}{e_{\max} - e_{\min}} - 1 \tag{9}$$

where $e_{\max}$ and $e_{\min}$ are the maximum and minimum values of $[u; i]$, respectively; and e' is the normalised voltage or current. Finally, the $G \times G$ images are obtained.

Dimension reduction: As shown in Fig. 4, owing to harmonics or non-periodic components, some pixel points even the whole curve will suffer translation but the essential features keep invariant. Correspondingly, the performance improvement of T:ANN is limited when the training samples are insufficient.

Max-pooling of CNN [34, 35] is employed to reduce image dimension for translation invariance. Suppose the filter size and step size are $p \times p$, s, respectively. Then image size after dimension reduction is $g \times g$:

$$g = \frac{G - p}{s} + 1. \tag{10}$$

Fig. 4 details the positive effects of max-pooling on translation invariance when both p and s are 3. It is helpful to improve the generalisation ability of S:ANN and T:ANN.

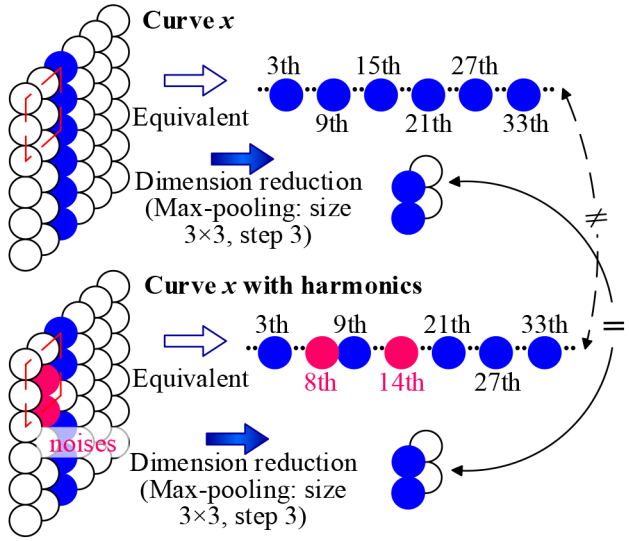

Fig. 4 *Influences of harmonics and a periodic component on curves and process of dimension reduction*

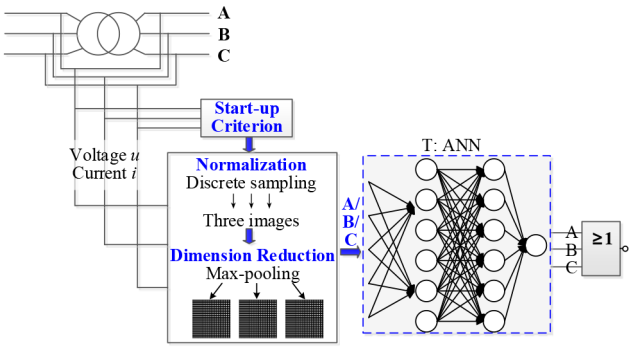

Fig. 5 *Transformer protection algorithm*

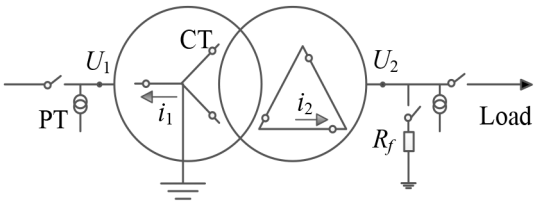

Fig. 6 *Simulation and experimental model*

3.2.2 Transformer protection algorithm: The proposed protection method is illustrated in Fig. 5. Once the start-up criterion is met, the equivalent magnetisation curve of each phase after data conversion and dimension reduction is used as input to the knowledge-based ANN to identify the running state. A tripping signal is issued when at least one phase is identified as an internal fault.

4 Case study

All the training samples are collected on PSCAD; and test samples include simulation and experimental samples. The simulation and experimental model is shown in Fig. 6.

4.1 Training and test samples collection

4.1.1 Training samples: Transformer ratio, sampling rate and other system parameters are shown in Table 1; and magnetic properties of iron core are detailed in Table 2. In addition, energisation time, fault occurrence time, fault turns etc. should be fully considered owing to the influences on equivalent magnetisation curves, as shown in Table 1.

Table 1 Parameters and scenarios of training samples

PSCAD simulation		Training samples	
		1st transformer	2nd transformer
parameters	connection	Y/Δ-11	Y/Δ-11
	ratio	230/11 kV	230/35 kV
	leakage resistance	0.1 p.u.	0.1 p.u.
	copper loss	0.005 p.u.	0.005 p.u.
	iron loss	5, 7, 10%	5, 7, 10%
	sampling frequency	20 kHz	20 kHz
operation conditions	NO	start-up time: 1.001, 1.002, …, 1.020 s	
	EF	occurrence time: 1.001, 1.002, …, 1.020 s	
		type: single phase-to-ground, two phase faults; two phase-to-ground faults.	
	EHT	occurrence time: 1.001, 1.002, …, 1.020 s	
	IF	occurrence time: 1.001, 1.002, …, 1.020 s; fault turns: 1.5, 2.0, 2.5, 3.0, 5.0%.	
	EFT	occurrence time: 1.001, 1.002, …, 1.020; fault turns: 1.5, 2.0, 2.5, 3.0, 5.0%.	
number	NO/EF	564	
	EHT	478	
	IF	760	
	EFT	760	

[a]NO: normal operation; EF: external fault; EHT: energising healthy transformer; IF: internal fault; EFT: energising faulty transformer.

Table 2 Magnetic properties of iron-core (p.u.)

Training: simulation				Test: simulation:	
1st transformer		2nd transformer			
Voltage	Current	Voltage	Current	Voltage	Current
0	0	0	0	0	0
0.1	9.2280×10^{-05}	1.0	0.0025	0.10	0.0000533
0.5	4.4895×10^{-04}	**1.1**	**0.005**	0.50	0.000259
1.0	1.04754×10^{-03}	1.2	0.02	1.00	0.000605
1.1	$\mathbf{7.18316\times10^{-03}}$	1.25	0.04	**1.10**	**0.00415**
1.2	9.95666×10^{-02}	1.28	0.1	1.20	0.0575
1.4	0.6285765	1.32	0.2	1.40	0.3629
1.6	1.352529	1.36	0.3	1.60	0.78088
1.8	2.179491	1.535	1	1.80	1.25834
1.9	2.6178888	3.7	10	1.90	1.51145

Bold values indicates the saturation point where the iron core begins to saturate. We find that we have reversed the order of 'Voltage' and 'Current' when filling in the values. In addition, the primary unit of 'Current' is '%' and we forgot to divide it by 100

Table 3 Parameters and scenarios of test samples in simulation system

PSCAD simulation		Test samples
parameters	Connection	Δ/Y/Y-11
	Ratio	35/220/500 kV
	leakage resistance	0.1 p.u.
	copper loss	0.005 p.u.
	iron loss	6%
	sampling frequency	2, 10 kHz
operation conditions	Same as Table 1	
number	NO/EF	225
	EHT	80
	IF	160
	EFT	160

[a]NO: normal operation; EF: external fault; EHT: energising healthy transformer; IF: internal fault; EFT: energising faulty transformer.

4.1.2 Test samples: Test samples are divided into two groups to verify the good generalisation ability of the knowledge-based ANN, namely the ones collected on PSCAD and the ones collected in dynamic model experiments.

(1) *Simulation samples:* The test samples on PSCAD adopt different system parameters from the training samples, as shown in Tables 2 and 3.

(2) *Experimental samples:* The experimental samples can partially verify the generalisation ability of the knowledge-based ANN to the on-site transformers. The experimental system and schematic diagram have been exhibited in Figs. 7 and 8, respectively. Table 4 shows the parameters and scenarios of experimental transformer. In order to accurately obtain the image of equivalent magnetisation curves, the Waveform Recorder with Hall sensor (DF1024), instead of the electromagnetic transformers, was used to construct the data acquisition system.

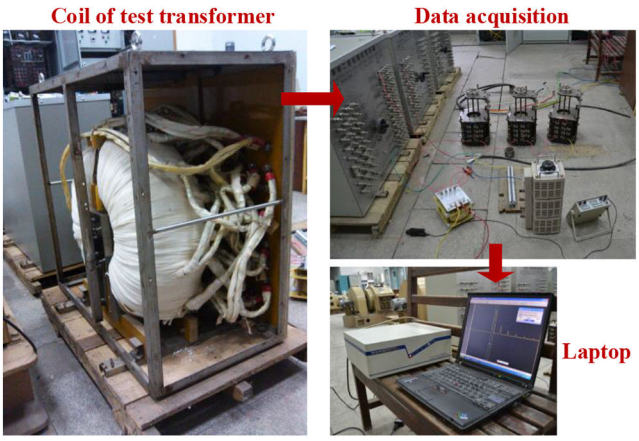

Fig. 7 *Partial details of the experimental system*

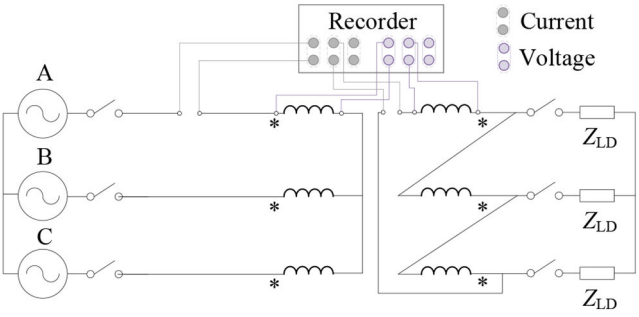

Fig. 8 *Schematic diagram of the experimental system*

Table 4 Parameters and scenarios of test samples in experimental system

Test samples (experiments)		
parameters	connection	Y/Δ-11
	singe phase voltage ratio	220/220 V
	single phase rated capacity	2 kVA
	single phase no-load current, loss	1.17%, 0.7%
	single phase short-circuit voltage, loss	11.2%, 1.0%
	sampling frequency	10 kHz
operation conditions	NO	load: 100, 75, 50, 25% of rated load; start-up time: random time
	EF	load: 100% of rated load; occurrence time: random time; type: single phase-to-ground, two phase faults; location: primary, secondary sides
	EHT	occurrence time: random time
	IF	load: 100% of rated load; occurrence time: random time; fault turns: primary side, 2.3, 4.5%; secondary side, 4.5%
	EFT	occurrence time: random time; fault turns: primary side, 2.3, 4.5, 9.1%; secondary side, 4.5, 9.1%.
number	NO/EF	139
	IF	60
	EHT	141
	EFT	22

[a]NO: normal operation; EF: external fault; EHT: energising healthy transformer; IF: internal fault; EFT: energising faulty transformer.

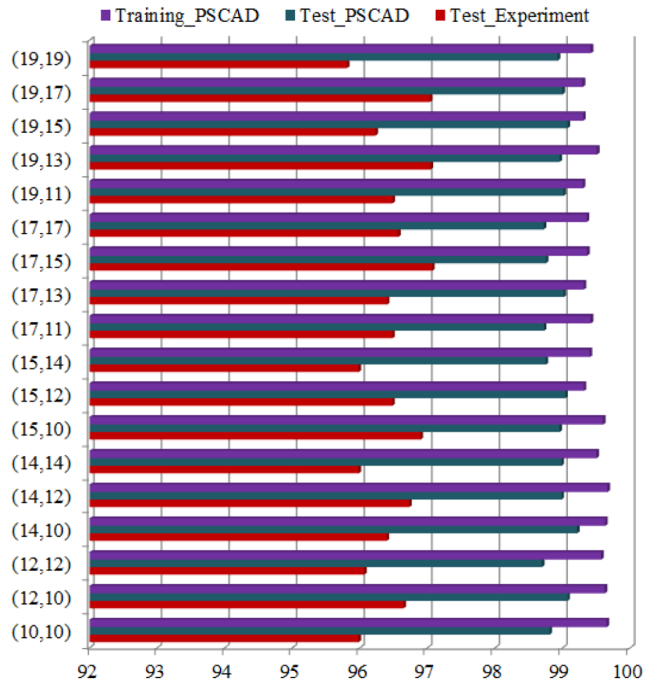

Fig. 9 *Average accuracy (%) of S: ANN*

The training and test samples are transformed into greyscale images by saturated parts removal, data conversion and dimension reduction successively, named as source domain S. The greyscale images which are obtained only by data conversion and dimension reduction are target domain T. The samples in S and T are 16×16 images according to the formula (10) where G, p and s are 50, 3, and 3, respectively.

4.2 Training process of knowledge-based ANN

4.2.1 S:ANN: Train S:ANN utilising source domain S. Accuracy, training and test time, vanishing gradient problem etc. are considered comprehensively for an optimal S:ANN. S:ANN adopts a double hidden layer with 10–19 neurons. Fig. 9 shows the average accuracies of alternative S:ANNs after random initialisation is employed repeatedly, where (x, y) denote the neurons of first and second hidden layers. From Fig. 9, when the neurons of hidden layers are (19, 17), (19, 13) or (15, 10), S:ANN performs better with accuracies of over 98.5% and nearly 97% for simulation and experimental samples, respectively.

S:ANN which has the highest accuracy for the experimental samples among the structures with neurons of (19, 17), (19, 13) or (15, 10) is used to extract the features Exp:F. Finally, the neurons of input, first hidden, second hidden and output layers are 256, 15, 10 and 1, respectively. As shown in Fig. 10, the training and test accuracies increase gradually as the classification loss decreases. The accuracy of experimental samples is 98.24%; and the corresponding accuracies of simulation samples are 99.18 and 98.95% for training and test, respectively.

In the next step, Exp:F is used to employ the adaption layers for training the T:ANN.

4.2.2 T:ANN: Train T:ANN utilising the recorded features Exp:F and target domain T. Random initialisation of T:ANN with the same structure as S:ANN is employed repeatedly and network parameters is optimised with the objective of minimising loss function (5). When δ_1 and δ_2 are 0.001 and 0.0002, respectively, T:ANN has a more reliable performance. Average accuracies of the target domain and highest accuracy of experimental samples have been shown in Fig. 11. The performances of common ANN are also added for the comparison with T:ANN.

According to Fig. 11, the common ANN performs well if adopting the neurons of (19, 17), with average accuracies of 99.15, 98.63 and 90.96% for training and test samples and highest accuracy of 91.96% for experimental samples. Comparatively,

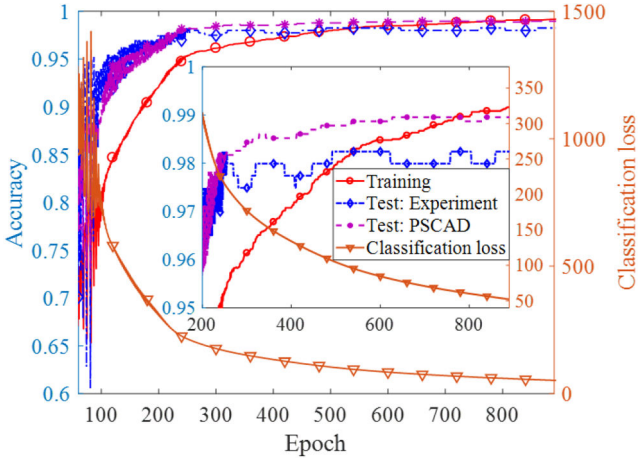

Fig. 10 *Classification loss and accuracies*

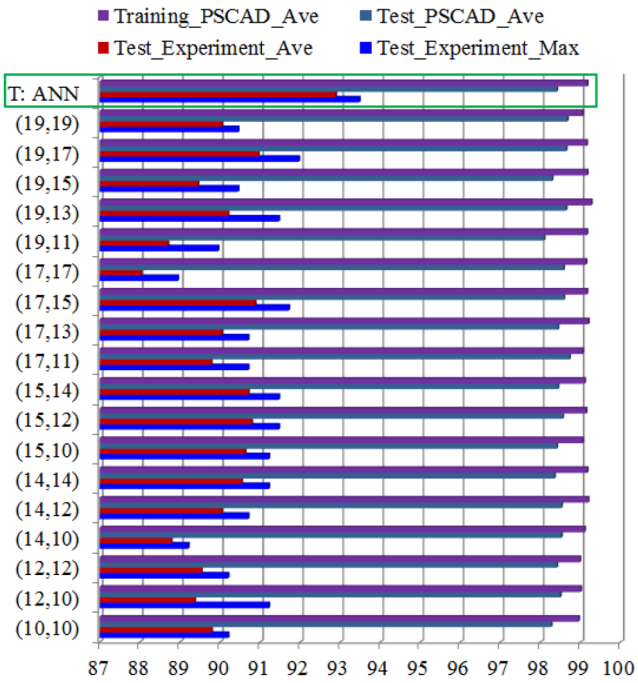

Fig. 11 *Comparison of T: ANN and common ANN*

T:ANN has a better generalisation ability owing to the combination of data and professional knowledge, with accuracies of 99.16, 98.40, 92.88 and 93.47%. Then T:ANN with the highest accuracy for the experimental samples is selected as knowledge-based ANN. Its losses and accuracies have been shown in Fig. 12. As adaption and classification losses decrease, accuracies of training and test samples increase gradually. Finally, accuracies of simulation samples are 99.06, 98.17%; the improved accuracy of experimental samples, 93.47%, verify the good generalisation ability of the knowledge-based ANN.

4.3 Performances on turn to ground faults and lead wire fault

To further verify the generalisation ability of the proposed transformer protection, the knowledge-based ANN is tested through the turn to ground faults and lead wire faults that are not involved in the training process. The operation conditions and transformer parameters of turn to ground faults and lead wire faults have been shown in Table 5. The magnetic properties of iron-core are same as that of the first transformer in Table 2.

Fig. 13 exhibits the input images of two phases for several faults occurring at 1.001 s, where they are the turn to ground faults occurring at the positions of 5.0, 10, 50, 90, 95, 97, 98.5% and the lead wire faults with the transition resistance of 0, 0.3, 0.5, 1, 2 Ω from left to right and from top to bottom. Although these faults are not involved in the training process, the images are similar to that

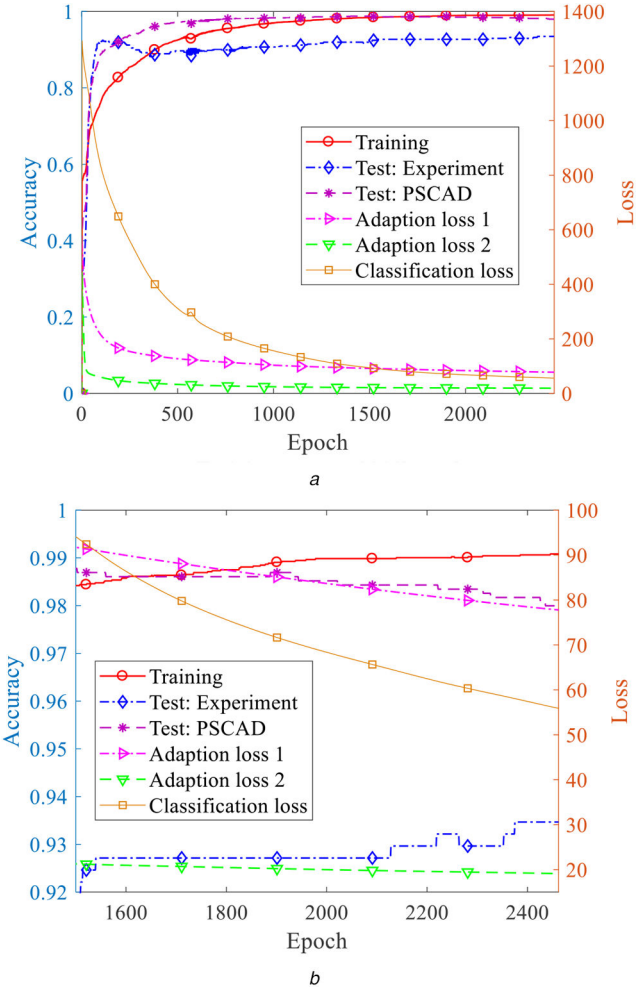

Fig. 12 *Losses and accuracies*
(a) Training process of 2462 epochs, ***(b)*** Training process of 1500–2462 epochs

of turn to turn faults from the perspective of the curve shapes. Therefore, the knowledge-based ANN performs well on the turn to ground faults and lead wire faults with the accuracies of 100% (180/180) and 97% (97/100).

4.4 Comparison with second harmonic restraint

This paper makes a comparison between the proposed method and second harmonic restraint through the experimental samples. The results have been shown in Table 6. For the samples of internal faults, energising faulty transformers and energising healthy transformers, the second harmonic restraint has a worse accuracy of 90.49%, compared with 93.15% of the proposed method. In addition, second harmonic restraint is not suitable for identifying normal operation/external fault. Once the differential protection starts by mistake, the lower second harmonic will lead to a malfunction in the system. However, the proposed method can reliably identify normal operation/external fault with the accuracy of 100% because such scenarios have been sufficiently considered in the training process of ANNs. In contrast, the accuracy of second harmonic restraint is only 24.44%.

5 Discussion

This section will discuss (i) adaptability of T:ANN to CT configuration, CT saturation and remanence; (ii) future work based on the proposed features transferring method.

5.1 Adaptability analysis

5.1.1 CT configuration: When CTs are installed outside delta winding, the 'voltage of magnetising branch-differential current' curves still have obvious differences among the transformer

Table 5 Transformer parameters of turn to ground fault and lead wire fault

PSCAD simulation		
parameters	connection	Y/Δ-11
	ratio	230/11 kV
	leakage resistance	0.1 p.u.
	copper loss	0.005 p.u.
	iron loss	5%
	sampling frequency	20 kHz
operation conditions (phase B)	TTGF	occurrence time: 1.001, 1.002, …, 1.020 s
		fault positions: 1.5, 3.0, 5.0, 10, 50, 90, 95, 97, 98.5%
		transition resistance: 0 Ω
	LWF	occurrence time: 1.001, 1.002, …, 1.020 s
		transition resistance: 0, 0.3, 0.5, 1, 2 Ω
number	TTGF	180
	LWF	100

[a]TTGF: turn to ground fault; LWF: lead wire fault.

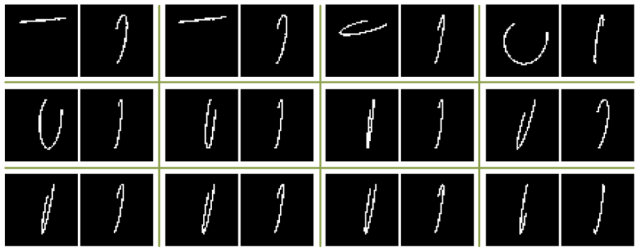

Fig. 13 *Input images of turn to ground faults and lead wire faults*

Table 6 Comparison between proposed method and second harmonic restraint

Performance comparison	Second-harmonic restraint	Proposed algorithm, %
IF/EFT 116	90.49	93.15
EHT 147		
NO/EF 135	—	100

[a]NO: normal operation; EF: external fault; EHT: energising healthy transformer; IF: internal fault; EFT: energising faulty transformer.

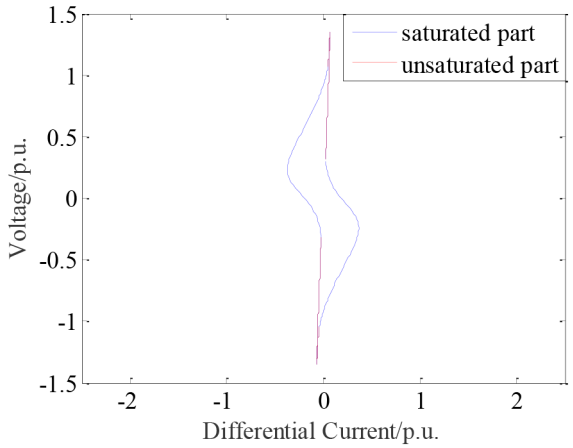

Fig. 14 *CT steady saturation*

scenarios but are slightly different from equivalent magnetisation curves in this paper. To be specific, when an internal fault occurs, at least two phases will exhibit fault features; for energising healthy transformer, differential current of at most one phase is symmetry inrush current. Therefore, the proposed transformer protection is still applicable. However, the knowledge-based ANN needs to be re-trained with new training samples and tripping will be issued at least two phases are identified as faults.

5.1.2 CT saturation: CT saturation includes steady saturation and transient saturation.

(1) *Transient saturation:* In one cycle, differential current under CT transient saturation exhibits the features of unilateral saturation. Therefore, the equivalent magnetisation curve is similar to that of energising transformer. The proposed protection method is also effective in CT transient saturation and it is unnecessary to re-train the knowledge-based ANN.

(2) *Steady saturation:* Equivalent magnetisation curve under CT steady saturation shows the bilateral features, as shown in Fig. 14. However, in one cycle, it still contains unsaturated parts that can reliably reflect whether the transformer is healthy. Therefore, the proposed protection method is applicable but re-training the knowledge-based ANN is necessary with the supplement of CT transient saturation Fig. 14.

5.1.3 Remanence: Remanence only affects the duration and occurrence time of iron-core saturation. In one cycle, the equivalent magnetisation curve still contains sufficient unsaturated parts. Therefore, the proposed protection method is applicable but the knowledge-based ANN should be re-trained with a longer data window to enhance the robustness to remanence.

5.2 Future work

CNN, as a deep learning algorithm, performs better than ANN in capturing spatial features of equivalent magnetisation curves. Theoretically, the performance of transformer protection will be further improved if a knowledge-based CNN can be developed with the proposed transferring method in this paper.

However, CNN has a higher requirement on the scale of training samples; moreover, the response speed of CNN with a large-scale network also hinder its application in transformer

protection. Therefore, developing a knowledge-based CNN with a rapid response will be the emphasis of future work.

6 Conclusion

This paper develops a knowledge-based ANN through the proposed features transferring method. This knowledge-based ANN which integrates data and professional knowledge is used to construct reliable transformer protection. To be specific, the images of equivalent magnetisation curve whose saturation parts are removed are defined as the source domain and the original samples are target domain. Source and target domains are used to train two ANNs with the same structure, respectively. The adaption layers are employed between the hidden layers of these two ANNs to reduce the feature differences. Through this features transferring method, ANN which is trained with target domain will adaptively focus on the essential features of equivalent magnetisation curves. Finally, the determined ANN is selected as the knowledge-based ANN, which is used to construct transformer protection.

The results of simulation and experimental samples show that the knowledge-based ANN has a good generalisation ability with the accuracies of 99.06, 98.17 and 93.47% for training samples, test samples on PSCAD and test samples in dynamic model experiments. As can be verified that the proposed features transferring method and the knowledge-based ANN have certain research values in the data-driven based power system problems.

7 Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 51877167).

8 References

[1] Wang, Y., Lu, H., Xiao, X., *et al.*: 'Cable incipient fault identification using restricted Boltzmann machine and stacked autoencoder', *IET Gener. Transm. Distrib.*, 2020, **14**, (7), pp. 1242–1250

[2] Bukhari, S.B.A., Kim, C., Mehmood, K.K., *et al.*: 'Convolutional neural network-based intelligent protection strategy for microgrids', *IET Gener. Transm. Distrib.*, 2020, **14**, (7), pp. 1177–1185

[3] Petrović, I., Nikolovski, S., Baghaee, H.R., *et al.*: 'Determining impact of lightning strike location on failures in transmission network elements using fuzzy decision-making', *IEEE Syst. J.*, 2020, **14**, (2), pp. 2665–2675

[4] Mishra, M., Rout, P.K.: 'Detection and classification of micro-grid faults based on HHT and machine learning techniques', *IET Gener. Transm. Distrib.*, 2018, **12**, (2), pp. 388–397

[5] Silva, S., Costa, P., Gouvea, M, *et al.*: 'High impedance fault detection in power distribution system using wavelet transform and evolving neural network', *Electr. Power Syst. Res.*, 2018, **154**, pp. 474–483

[6] Abdullah, A.M., Butler-Purryb, K.: 'Secure transmission line distance protection during wide area cascading events using artificial intelligence', *Electr. Power Syst. Res.*, 2019, **175**, p. 105914

[7] Gashteroodkhani, O.A., Majidi, M., Etezadi-Amoli, M.: 'A combined deep belief network and time-time transform based intelligent protection scheme for microgrids', *Electr. Power Syst. Res.*, 2020, **182**, p. 106239

[8] Manohara, M., Koleya, E., Ghosha, S., *et al.*: 'Spatio-temporal information based protection scheme for PV integrated microgrid under solar irradiance intermittency using deep convolutional neural network', *Int. J. Electr. Power Energy Syst.*, 2020, **116**, p. 105576

[9] Farshad, M.: 'Detection and classification of internal faults in bipolar HVDC transmission lines based on K-means data description method', *Int. J. Electr. Power Energy Syst.*, 2019, **104**, pp. 615–625

[10] Morales, J., Orduñab, E., Villarroelb, H., *et al.*: 'High-speed directional protection without voltage sensors for distribution feeders with distributed generation integration based on the correlation of signals and machine learning', *Electr. Power Syst. Res.*, 2020, **184**, p. 106295

[11] Lin, X.N., Huang, J.G., Zeng, L.J., *et al.*: 'Analysis of electromagnetic transient and adaptability of second-harmonic restraint based differential protection of UHV power transformer', *IEEE Trans. Power Deliv.*, 2010, **25**, (4), pp. 2299–2307

[12] Lin, X.N., Liu, P., Malik, O.P.: 'Studies for identification of the inrush based on improved correlation algorithm', *IEEE Trans. Power Deliv.*, 2002, **17**, (4), pp. 901–907

[13] Guillen, D., Esponda, H., Vazquez, E., *et al.*: 'Algorithm for transformer differential protection based on wavelet correlation modes', *IET Gener. Transm. Distrib.*, 2016, **10**, (12), pp. 2871–2879

[14] He, B.T., Zhang, X.S., Bo, Z.Q.: 'A new method to identify inrush current based on error estimation', *IEEE Trans. Power Deliv.*, 2006, **21**, (3), pp. 1163–1168

[15] Phadke, A.G., Thorp, J.S.: 'A new computer-based flux-restrained current-differential relay for power transformer protection', *IEEE Trans. Power Appar. Syst.*, 1983, **102**, (11), pp. 3624–3629

[16] Ge, B.M., De Almeida, A.T., Zheng, Q.L., *et al.*: 'An equivalent instantaneous inductance-based technique for discrimination between inrush current and internal faults in power transformers', *IEEE Trans. Power Deliv.*, 2005, **20**, (4), pp. 2473–2482

[17] Ma, J., Wang, Z.P., Yang, Q.X., *et al.*: 'A two terminal network-based method for discrimination between internal faults and inrush currents', *IEEE Trans. Power Deliv.*, 2010, **25**, (3), pp. 1599–1605

[18] Medeiros, R.P., Costa, F.B.: 'A wavelet-based transformer differential protection with differential current transformer saturation and cross-country fault detection', *IEEE Trans. Power Deliv.*, 2018, **33**, (2), pp. 789–799

[19] Roy, A., Singh, D., Misra, R.K., *et al.*: 'Differential protection scheme for power transformers using matched wavelets', *IET Gener. Transm. Distrib.*, 2019, **13**, (12), pp. 2423–2437

[20] Haghjoo, F., Mostafaei, M., Mohammadi, H.: 'A new leakage flux-based technique for turn-to-turn fault protection and faulty region identification in transformers', *IEEE Trans. Power Deliv.*, 2018, **33**, (2), pp. 671–679

[21] Farzin, N., Vakilian, M., Hajipour, E.: 'Transformer turn-to-turn fault protection based on fault-related incremental currents', *IEEE Trans. Power Deliv.*, 2019, **34**, (2), pp. 700–709

[22] Shah, A.M., Bhalja, B.R., Patel, R.M.: 'New protection scheme for power transformer based on superimposed differential current', *IET Gener. Transm. Distrib.*, 2018, **12**, (14), pp. 3587–3595

[23] Batista, Y.N., de Souza, H.E.P., de Assis dos Santos Neves, F., *et al.*: 'A GDSC-based technique to distinguish transformer magnetizing from fault currents', *IEEE Trans. Power Deliv.*, 2018, **33**, (2), pp. 589–599

[24] Murugan, S.K., Simon, S.P., Sundareswaran, K., *et al.*: 'An empirical Fourier transform-based power transformer differential protection', *IEEE Trans. Power Deliv.*, 2017, **32**, (1), pp. 209–218

[25] Sahebi, A., Samet, H.: 'Efficient method for discrimination between inrush current and internal faults in power transformers based on the non-saturation zone', *IET Gener. Transm. Distrib.*, 2017, **11**, (6), pp. 1486–1493

[26] Samantaray, S.R., Dash, P.K.: 'Decision tree based discrimination between inrush currents and internal faults in power transformer', *Int. J. Emerg. Electr. Power Syst.*, 2011, **33**, (4), pp. 1043–1048

[27] Ozgonenel, O., Karagol, S.: 'Power transformer protection based on decision tree approach', *IET Electr. Power Appl.*, 2014, **8**, (7), pp. 251–256

[28] Shah, A.M., Bhalja, B.R.: 'Fault discrimination scheme for power transformer using random forest technique', *IET Gener. Transm. Distrib.*, 2016, **10**, (6), pp. 1431–1439

[29] Asrami, M.A., Gorjikolaie, M.T., Razavi, S.M.: 'A novel intelligent protection system for power transformers considering possible electrical faults, inrush current, CT saturation and over-excitation', *Int. J. Emerg. Electr. Power Syst.*, 2015, **64**, pp. 1129–1140

[30] Balaga, H., Gupta, N., Vishwakarma, D.N.: 'GA trained parallel hidden layered ANN based differential protection of three phase power transformer', *Int. J. Emerg. Electr. Power Syst.*, 2015, **67**, pp. 286–297

[31] Segatto, E.C., Coury, D.V.: 'A differential relay for power transformers using intelligent tools', *IEEE Trans. Power Syst.*, 2006, **21**, (3), pp. 1154–1162

[32] Barbosa, D., Coury, D.V., Oleskovicz, M.: 'New approach for power transformer protection based on intelligent hybrid systems', *IET Gener. Transm. Distrib.*, 2012, **6**, (10), pp. 1009–1018

[33] Geethanjali, M., Raja Slochanal, S.M., Bhavani, R.: 'PSO trained ANN-based differential protection scheme for power transformers', *Neurocomputing*, 2008, **71**, (4–6), pp. 904–918

[34] Tripathy, M., Maheshwari, R.P., Verma, H.K.: 'Application of probabilistic neural network for differential relaying of power transformer', *IET Gener. Transm. Distrib.*, 2007, **1**, (2), pp. 218–222

[35] Tripathy, M., Maheshwari, R.P., Verma, H.K.: 'Probabilistic neural-network-based protection of power transformer', *IET Electr. Power Appl.*, 2007, **1**, (5), pp. 793–798

[36] Tripathy, M., Maheshwari, R.P., Verma, H.K.: 'Power transformer differential protection based on optimal probabilistic neural network', *IEEE Trans. Power Deliv.*, 2010, **25**, (1), pp. 102–112

[37] Tripathy, M., Maheshwari, R.P., Verma, H.K.: 'Radial basis probabilistic neural network for differential protection of power transformer', *IET Gener. Transm. Distrib.*, 2008, **2**, (1), pp. 43–52

[38] Salama, A.M., Abdel-Latif, K.M., Ismail, M.M., *et al.*: 'A new hybrid protection algorithm for protection of power transformer based on discrete wavelet transform and ANFIS inference systems', *Int. J. Emerg. Electr. Power Syst.*, 2018, **19**, (3), pp. 1–8

[39] Jazebi, S., Vahidi, B., Hosseinian, S.H., *et al.*: 'Magnetizing inrush current identification using wavelet based Gaussian mixture models', *Simul. Modelling Pract. Theory*, 2009, **17**, (6), pp. 991–1010

[40] Fei, S.W., Wang, M.J., Miao, Y.B., *et al.*: 'Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil', *Energy Convers. Manage.*, 2009, **50**, (6), pp. 1604–1609

[41] Shah, A.M., Bhalja, B.R.: 'Discrimination between internal faults and other disturbances in transformer using the support vector machine-based protection scheme', *IEEE Trans. Power Deliv.*, 2013, **28**, (3), pp. 1508–1515

[42] Raichura, M.B., Chothani, N.G., Patel, D.D.: 'Identification of internal fault against external abnormalities in power transformer using hierarchical ensemble extreme learning machine technique', *IET Sci. Meas. Technol.*, 2020, **14**, (1), pp. 111–121

[43] Jazebi, S., Vahidi, B., Hosseinian, S.H.: 'A novel discriminative approach based on hidden markov models and wavelet transform to transformer protection', *Simulation*, 2010, **86**, (2), pp. 93–107

[44] Jazebi, S., Vahidi, B., Jannati, M.: 'A novel application of wavelet based SVM to transient phenomena identification of power transformers', *Energy Convers. Manage.*, 2011, **52**, (2), pp. 1354–1363

[45] Barbosa, D., Netto, U.C., Coury, D.V., *et al.*: 'Power transformer differential protection based on clarke's transform and fuzzy systems', *IEEE Trans. Power Deliv.*, 2011, **26**, (2), pp. 1212–1220

[46] Tripathy, M.: 'Power transformer differential protection using neural network principal component analysis and radial basis function neural network', *Simul. Modelling Pract. Theory*, 2010, **18**, (5), pp. 600–611
[47] Bejmert, D., Rebizant, W., Schiel, L.: 'Transformer differential protection with fuzzy logic based inrush stabilization', *Int. J. Emerg. Electr. Power Syst.*, 2014, **63**, pp. 51–63
[48] Li, C.Y., Zhou, N.C., Liao, J.Q.*, et al.*: 'Multiscale multivariate fuzzy entropy-based technique to distinguish transformer magnetising from fault currents', *IET Gener. Transm. Distrib.*, 2019, **13**, (12), pp. 2319–2327
[49] Lu, Z., Tang, W.H., Ji, T.Y.*, et al.*: 'A morphological scheme for inrush identification in transformer protection', *IEEE Trans. Power Deliv.*, 2009, **24**, (2), pp. 560–568
[50] Wu, W.C., Ji, T.Y., Li, M.S.*, et al.*: 'Using mathematical morphology to discriminate between internal fault and inrush current of transformers', *IET Gener. Transm. Distrib.*, 2016, **10**, (1), pp. 73–80
[51] Afrasiabi, S., Afrasiabi, M., Parang, B.*, et al.*: 'Designing a composite deep learning based differential protection scheme of power transformers', *Appl. Soft Comput. J.*, 2020, **87**, p. 105975
[52] Afrasiabi, S., Afrasiabi, M., Parang, B.*, et al.*: 'Integration of accelerated deep neural network into power transformer differential protection', *IEEE Trans. Ind. Inf.*, 2020, **16**, (2), pp. 865–876
[53] Jiao, Z.B., Li, Z.B.: 'Novel magnetization hysteresis-based power-transformer protection algorithm', *IEEE Trans. Power Deliv.*, 2018, **33**, (5), pp. 2562–2570