Efficient CNN-XGBoost technique for classification of power transformer internal faults against various abnormal conditions

Maulik Raichura\textsuperscript{1} | Nilesh Chothani\textsuperscript{2} | Dharmesh Patel\textsuperscript{3}

\textsuperscript{1} Electrical Engineering Department, Gujarat Technological University (GTU), Ahmedabad, Gujarat, India
\textsuperscript{2} Electrical Engineering Department, Adani Institute of Infrastructure Engineering, Gujarat Technological University (GTU), Ahmedabad, Gujarat, India
\textsuperscript{3} Electrical Engineering College, Bharuch, Gujarat, India

\textbf{Abstract}

To increase the classification accuracy of a protection scheme for power transformer, an effective convolution neural network (CNN) extreme gradient boosting (XGBoost) combination is proposed in this work. Data generated from various test cases are fed to one-dimensional CNN for high-level feature extraction. After that, an efficient classifier tool XGBoost is used to properly discriminate different transformer internal faults against outside abnormalities. A portion of an Indian power system is considered and simulated in PSCAD software using the multi-run feature to collect a large number of data for various fault/abnormal situations. The generated data are used in MATLAB software where the proposed algorithm is programmed. A high-performance CPU is used for training and testing purpose of the projected artificial intelligent technique. The obtained results for classification accuracy as well as discrimination time shows that the proposed scheme is competent enough to properly discriminate transformer operational conditions. Further, the combined CNN-XGBoost technique is compared with existing relevance vector machine and hierarchical ensemble of extreme learning machine classifier techniques. Moreover, a hardware experiment is performed in a laboratory prototype of 50 kVA, 440/220 V transformer to verify the authenticity of the developed protective scheme. After analyzing a variety of experiments, the authors note that the presented method provides promising classification accuracy within a short time period.

\section{Introduction}

Having a sound protective scheme for a transformer is a burning desire for protection engineers because a transformer is considered the heart of the power system which pumps power from one place to another in the entire power grid. Hence, the protection of this precious asset is most essential to keep the power system alive. The protection scheme should be capable enough to detect the internal fault and isolate the transformer from a healthy system to reduce the damage to the transformer as well as to the surrounding environment. To execute the detection of only internal faults from various abnormal conditions, various researchers have already proposed different techniques. Artificial intelligence (AI), machine learning, filtration-based techniques as well as combinations of these techniques are the common platform for researchers. Due to advancements in technology, AI-based classification and wave shape recognition techniques have been trending among researchers.

Relevance vector machine (RVM)-based protective technique was utilized for the power transformer by Chothani et al. \cite{1}. In their paper, RVM along with support vector machine (SVM) \cite{2} as well as probabilistic neural network (PNN) \cite{3} based classifier methods are presented. The assessment shows that RVM performs better for classification accuracy than other considered techniques. However, during higher training/testing data size, RVM-based method takes a longer computational time than usual. Another protective technique, hierarchical ensemble of extreme learning machine (HE-ELM) has been presented in \cite{4}. Wang et al. \cite{5} and Dogaru et al. \cite{6} compared the ELM with SVM (i.e. SVM) and presented the ELM with SVM (i.e. SVM). According to that paper, it is found that the performance of ELM in terms of discrimination accuracy and learning is appreciable than SVM. Though ELM
performs better than SVM, it suffers from over-fitting, high computational time and requirement of feature extraction. Artificial neural network based transformer protection was presented by Balaga et al. [7]. The method is tedious and time-consuming as it has to undergo seven diverse steps each time throughout its training session.

Other than machine learning based classifier techniques, there are several techniques anticipated by assorted researchers to protect the transformer during fault conditions. Shah et al. [8] proposed an adaptive protection scheme that takes care of the transformer during a tap changing procedure. This scheme can defend 96% of transformer winding during the YY transformer connection. Moreover, if high-impedance fault occurs then the percentage winding protected is further reduced. Bagheri et al. [9] presented the effect of various mechanical defects and electrical abnormalities on the performance of the transformer differential protective technique. The estimation of change in various parameters like capacitance (C), inductance (L), etc., and based on that, mechanical defects of the winding can be identified. However, the algorithm takes action if the differential current exceeds the preset threshold but it is not always necessary that the differential current exceeds a predefined threshold. During such conditions, the proposed algorithm fails to protect the transformer.

Online condition monitoring of the transformer has been proposed by Ballal et al. [10]. However, only condition monitoring of the transformer is not enough, along with it a unit protection scheme has to be proposed. Hence, to provide complete protection to the transformer, a sound protection scheme along with online condition monitoring is required. Moreover, Ghanbari et al. [11] suggested bridge type fault current limiter at the transformer neutral to limit fault current and at the same time it retains the sensitivity of the protection scheme. However, this scheme may false operate in case of fault occurs close to the neutral point of the transformer winding; this is due to a low fault current magnitude.

Further, Dukic et al. [12] presented a technique to identify transformer faults with the use of M-robust evaluation of sound signals captured with the help of a microphone. This scheme completely depends on the microphone hardware tool; if this device gets defected then the whole scheme fails to protect the transformer. Here, if the fault occurs outside the transformer then also it may generate the fault vibrations, which may mislead the protective scheme. Hooshyar et al. [13] explained a fault identification technique with the help of the symmetry assessment window. This method blocks the relay signal if the current transformer (CT) saturates during an external fault. However, this methodology is not able to correct the saturated waveform, and hence if the external and internal fault occurs simultaneously then it will result in a very dangerous condition for the protection system. Abdoo [14] used variation mode decomposition for detection of CT saturation. The required constraint of the variable length window might not be satisfied in a noisy environment. Chothani et al. [15] presented an algorithm to sense the CT saturation condition and the compensation of the saturated waveforms using Newton’s backward difference formulae. However, if the amount of fault current is more than the preset threshold during the compensation, then it cannot measure the actual magnitude of the current.

Turn-to-turn incipient fault detection for power transformers utilizing flux has been described by Mostafaei and Hagjho in [16]. This scheme does not apply to the existing transformers as the search coils should be wound around the core at the time of manufacturing and the method is not suitable for the case when the transformer is grounded through impedance which decreases the sensitivity of the scheme. Dashti et al. [17] proposed a discrimination technique for the transformer’s high inrush currents from the fault currents. However, this technique fails to detect mild inrush current as identified by the conventional second harmonic detection method. Fast discrimination through a superimposed component comparison for identification of transformer inside fault has been explained by Lin et al. [18]. This protection scheme requires additional potential transformers to measure voltages that augment the protection cost and complexity of protection schemes. Also, this scheme applies only to the initial conditions or sudden changes in the voltage and current and cannot provide perfect protection to the case like incipient fault conditions where the changes are slight. Oliveira et al. [19] described an adaptive differential protective method for the transformer by analyzing transient signals. The overall accuracy for fault discrimination is 97.11% using Daubechies mother wavelet and the accuracy will decrease with the increases in the fault resistance.

After analyzing various filtering and AI-based techniques, it is accomplished that if large training data set is available then the neural network of AI technique can be trained soundly. Earlier, due to lack of technology, a vast amount of training data set can become hurdles in the fast execution of the algorithm as well as cause storage problem. However, nowadays because of advancement and up-gradation in technology, a large data set can be stored and retrieved in an acceptable time frame. Hence, researchers can use a large data set to train a particular algorithm. Convolution neural network (CNN) is the most trending technique of AI to systematically train the neural network with a large data set. In this paper, the high-level feature extraction facility of the CNN and superior classification ability of extreme gradient boosting (XGBoost) are utilized. The combination of these two outstanding AI tools is utilized to achieve the foremost goal of discrimination of internal fault of the power transformer from other abnormalities like inrush, external fault along with CT saturation and/or without CT saturation, over-fluxing, cross country fault, etc.

The entire paper is organised as follows: the first section reports a brief description of historical work done by various researchers in the particular field and its limitations. The second section describes system modelling along with various data generation. The third section presents the proposed technique with the developed algorithm. The fourth section involves the parameter configuration of the proposed technique with software result analysis. Further, the fifth section demonstrates a comparison of the CNN-XGBoost technique with the RVM and HE-ELM based technique. In the sixth section, hardware-based authentication is
discussed. Finally, the whole work is concluded in the seventh section.

2 | COMBINED CNN-XGBOOST TECHNIQUE

Having numerous classification schemes for inrush and fault detection techniques, every technique has its limitations. So, it is necessary to define the most efficient, easy to use, less time consuming and reliable technique to discriminate various external abnormalities against transformer internal fault conditions. Classifier techniques have their most efficient pattern recognition methodology with maximum accuracy; however, training time as well as testing time are the main constraints of it.

The combination of the CNN with the XGBoost method [22] is deliberately featured here. The presented technique serves as an effective classifier for the transformer internal faults and outside abnormalities. This combination takes advantage of effective feature extraction by CNN and the obtained features are utilized for the classification competency of XGBoost. The output obtained from CNN’s fully connected layer (last layer) is given as an input to XGBoost classifier and this arrangement results in an outstanding classifier technique for transformer internal fault identification purposes.

2.1 | Convolutional neural network

CNN can be directly used as a classifier to classify internal fault against external abnormalities when a large set of two-dimensional (2D) data is available. Afrasiabi [23] proposed CNN-based transformer fault classification technique, but at the same time it is required to convert the tabular data of transformer fault (1D) to 2D data. Moreover, that scheme of accelerated CNN [23] is tested and trained with the help of artificial data. Thus, the obtained result may not reflect perfect accuracy due to the presence of artificial data.

After examining thorough literature, it is found that the CNN is the most advanced feature extractor technique. Cun [24] proposed CNN initially for the identification of handwritten digital images. Salient features of CNN are respective fields, weight sharing and sub-sampling (pooling) which sequentially lower the complexity of the network parameters and its structure. The term respective field works as a filter that obtains significant features from the input data set. Further, the weight sharing facility shrinks the required number of parameters to be trained. Moreover, the key problem of over-fitting conditions of the learning machine can be overcome with the help of the pooling feature of CNN. Hence, all the features of CNN make the technique more suitable to use particularly for the application of feature extraction in this work.

The structure of CNN is organized as shown in Figure 1. Initially, the input layer is followed by alternating convolutional layer and pooling layer (sub-sampling layer), after that at last fully connected output layer provides the required output. The network adjusts all the kernels with the help of the back-propagation method laid on a stochastic gradient descent algorithm, which reduces the space among output and training labels. The stated convolution layer derives the features with the use of respective fields. After that the pooling layer pools out the required features and acts as a secondary extraction which further reduces the matrix dimensions. An important thing to be pointed out here is that the dual feature filtration technique as mentioned above is capable to tackle highly distorted inputs; the distortion is filtered out in convolutional and pooling layers [25].

However, this paper deals with 1D data for transformer fault classification purposes. Hence, a 1D-CNN is utilized for feature extraction purposes. As per literature, the 1D-CNN was first proposed by Kiranyaz et al. [26–28]. The main reason behind using 1D-CNN here is that it eliminates the need for pre-processing of training as well as testing data that was performed in [23]. In 1D-CNN, the matrix operations are replaced by reverse conv1D instead of lateral rotation (rot180) and 2D convolution (conv2D) in conventional 2D CNN [27]. Further, the constraints for kernel size (K) and the sub-sampling (s) are now scalars quantities for 1D-CNN. Moreover, the fully connected layer and back propagation (BP) formula remain the same as that of 2D structure. 1D-CNN has already been used for motor fault classification in [29].

In 1D-CNN, the one dimension forward propagation (1D-FP) can be termed as follows:

$$x_k^l = b_k^l + \sum_{i=1}^{N_l-1} \text{conv1D} \left( w_{ik}^{l-1}, t_{i}^{l-1} \right) \quad (1)$$

where $x_k^l$ denotes input; $b_k^l$ is depicted as the bias of $k$th neuron, at layer $l$; $w_{ik}^{l-1}$ is termed as the kernel from $i$th neuron, at layer $l - 1$ to $k$th neuron at layer $l$; $t_{i}^{l-1}$ is denoted as an output of the $i$th neuron, at layer $l - 1$.

Now after FP, the BP of the error is initiated from the fully connected output layer. Let us take layer $l = 1$ for the case of the input layer and $l = L$ in the case of the output layer. The mean square error (MSE) for the output layer, that is, in $L$ layer, for the input vector $\hat{y}$, is written as

$$E_p = \text{MSE} \left( \hat{y}_l, y^{l}_{1}, ..., y^{l}_{N_l} \right) = \sum_{i=1}^{N_l} \left( y_i^l - \hat{y}_i \right)^2. \quad (2)$$
Here, \( E \) is the MSE, \( N \) is the number of classes, a \( t_i^p \) is the target vector and \( [x_i^p,\ldots,x_i^p'] \) is the output vector.

Moving further, by finding a derivative of this error concerning the individual weight of \( k \)th neuron and its bias \( b_k \), the error can be minimized with the help of gradient descent method. Once the error is calculated, then update the corresponding weight and biases with the help of the gradient descent method. The formula of \( \Delta_k^l \) is utilized for updating the process of weights as well as biases

\[
\frac{\partial E}{\partial w_{ik}} = \Delta_k^l y_i^l - \Delta_k^l y_i^l \quad \text{and} \quad \frac{\partial E}{\partial b_k} = \Delta_k^l .
\]

The regular (scalar) BP is generally calculated as

\[
\frac{\partial E}{\partial z_i^l} = \Delta_k^l = \sum_{k=1}^{N_k} \frac{\partial E}{\partial x_i^{l+1}} \frac{\partial x_i^{l+1}}{\partial y_i^l} \frac{\partial y_i^l}{\partial z_i^l}.
\]

Here, \( f_i = (\delta j)^{-1} \)

As every element of \( s_j \) was gained by averaging ss of intermittent output \( y_i \). The delta error (\( \Delta_{i+1} \)) can be written as

\[
\Delta_{i+1}^l = \sum_{i=1}^{N_i} \text{conv1Dz} (\Delta_{i+1}^l, \text{rev}(w_{jk})) .
\]

Here, \( \text{rev}(\ldots) \) will reverse the array, similarly, \( \text{conv1Dz}(\ldots) \) performs full convolution in 1D. Then, the weights and bias sensitivities can be given as

\[
\frac{\partial E}{\partial w_{sk}} = \text{conv1D}(\Delta_{i+1}, \text{rev}(w_{jk})) \quad \text{and} \quad \frac{\partial E}{\partial b_k} = \sum_{i} \Delta_i^l (u_i).
\]

Hence, the weights and biases are updated as follows by estimating its sensitivities [28] by learning factor " as

\[
w_{ik}^{l-1} (t + 1) = w_{ik}^{l-1} (t) - \varepsilon \frac{\partial E}{\partial w_{ik}^{l-1}}
\]

\[
b_k^{l-1} (t + 1) = b_k^{l-1} (t) - \varepsilon \frac{\partial E}{\partial b_k^{l-1}}.
\]

The output of the CNN is now used as input for the XGBoost classifier.

### 2.2 eXtreme gradient boosting

XGBoost is an extended version of the gradient boosting machine learning ensemble method which is proposed by Guestrin et al. [30]. The gradient boosting method sequentially combines the decisions of weak classifiers and consequently becomes an effective learning machine of the ensemble decision tree method. Further, the XGBoost method reduces the computational complexity and calculation time of the gradient boosting method. XGBoost is an effective classifier method in various areas to obtain the state of the art result on challenges.

The mathematical expression for the XGBoost method is elaborated as discussed further. The given dataset with ‘n’ examples and ‘m’ features can be represented as \( D = \{(x_i, y_i)\} \quad |D| = n, x_i \in \mathbb{R}^m, y_i \in \mathbb{R}^m \), the output of the tree boosting model with \( k \) additive functions/trees can be defined as

\[
\hat{y}_i = \sum_{k=1}^{k} f_k(x_i), \quad f_k \in F
\]

where \( F = \{f(x) = \omega(x)\} \quad (\omega : \mathbb{R}^m \rightarrow T, \omega \in \mathbb{R}^T) \) is called space for classification and regression tree, whereas the term \( T \) represents several leaves of a particular tree. Here, \( f_k \) consists of \( g \) (structure part of a tree) and \( \omega \) (leaf weights part of the tree). For the classification purpose, a decision rule in the tree is used which is given by \( g \) and estimates the final prediction by summing up the score of respective leaves which is given by \( \omega \). Here, \( \omega \) represents the score of \( q \)th leaf.

The function \( f_k \) learns by minimizing the objective function:

\[
O = \sum_{i} I (\hat{y}_i, y_i) + \sum_{k} \Omega (f_k)
\]

where \( I \) of Equation (11) depicts training loss function or error estimation function. The term \( / \) measures the distance between the predicted term \( \hat{y}_i \) and actual term \( y_i \). The second term \( \Omega \) shows the tree model complexity penalty term

\[
\Omega (f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2.
\]

The additional regularization parameters available here help in smoothing the ultimate learning weights to prevent the model from conditions like over-fitting. Here in Equation (12), \( T \) is the number of leaves and \( \omega \) is the vector of the score on leaves. Now, if the objective function of a tree boosting model is as per Equation (11) then it can’t be optimized with the help of traditional optimization techniques in Euclidian space. The tree boosting model is trained in an additive manner. Hence, it can be said that the gradient tree boosting is an enhanced version of the tree boosting method. As mentioned above, if \( \hat{y}_i^{(j)} \) is a predictive term of \( j \)th instance, at \( j \)th iteration, then it is needed to add \( f_j \) in the previous predictive term, that is, \( \hat{y}_i^{(j-1)} \). Consequently, the
Objective function Equation (11) has been altered as

\[ O^{(i)} = \sum_{j=1}^{n} \left( y_j \hat{y}_j^{(i-1)} + f_j(x_j) \right) + \Omega (f_i) \].

(13)

Now, the XGBoost technique approximates Equation (13) by second-order Taylor expansion. The ultimate objective function at step ‘t’ can be given as

\[ O^{(i)} \approx \tilde{O}^{(i)} = \sum_{j=1}^{n} \left[ l \left( y_j, \hat{y}_j^{(i-1)} \right) + g_j f_j(x_j) + \frac{1}{2} b_j f_j^2(x_j) \right] + \Omega (f_i) \].

(14)

where \( g_j \) and \( b_j \) are the first as well as second-order gradient statistics on the loss function and \( \Omega (f_i) \) is given as per the Equation (12). The constants are eliminated to derive the simplified version of Equation (14)

\[ \tilde{O}^{(i)} = \sum_{j=1}^{T} \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left( \sum_{i \in I_j} b_i + \lambda \right) \omega_j^2 \right] + \gamma T \]

(15)

here the term \( I_j \) is denoted as \( I_j = \{ i | g_i(x_i) = j \} \) the instance of leaf \( j \).

For a fixed structure of a tree, \( q(x) \), the solution weight \( \omega_j^* \) of leaf \( j \) is given as

\[ \omega_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} b_i + \lambda} \]

(16)

From Equation (15) and Equation (16)

\[ \tilde{O} (q) = -\frac{1}{2} \sum_{j=1}^{T} \left( \sum_{i \in I_j} g_i \right)^2 \sum_{i \in I_j} b_i + \lambda + \gamma T \]

(17)

Equation (17) is considered as a scoring function for \( q(x) \) tree structure estimation to get optimal tree structure for classification. Generally, it is not possible to evaluate all tree structure \( q \). An effective algorithm [30] which initiated from a single leaf and iteratively append branches to the tree is utilized. Assuming \( I_L \) and \( I_R \) are the case sets of left as well as right nodes, respectively, after splitting and also \( I = I_L \Delta I_R \). The loss minimization equation after splitting is written as

\[ O_{split} = \frac{1}{2} \left[ \left( \frac{\sum_{i \in I_L} g_i}{\sum_{i \in I_L} b_i + \lambda} \right)^2 + \left( \frac{\sum_{i \in I_R} g_i}{\sum_{i \in I_R} b_i + \lambda} \right)^2 - \left( \frac{\sum_{i \in I} g_i}{\sum_{i \in I} b_i + \lambda} \right)^2 \right] - \gamma \]

(18)

XGBoost technique is a fast performing sequential gradient boosting classifier with highly accurate and promising results. Hence, this methodology is proposed here for the detection of transformer in-zone fault from various abnormalities. Figure 1 shows the representation of the structure of our anticipated combined CNN-XGBoost technique. The input in terms of 1D is directly given to CNN to extracts significant features from the given input data. The extracted features are then directly fed to XGBoost classifier to recognize that the transformer encounters any internal abnormality or not.

3 | POWER SYSTEM NETWORK

Figure 2 shows a line diagram of the seven-bus Indian power system. The Thevenin’s equivalent generators of different power and voltage levels are connected with four transformers as shown in Figure 2. Here, for the analysis, transformer-1 is considered as a principal transformer for all the test cases. That means the effect of all the peripheral equipment or devices is observed on transformer-1 itself. The transformer-1 is having 150-MVA capacity with a 13.8/220 kV voltage rating and is connected in a YΔ manner. The rated frequency of the entire power system is 50 Hz for all the equipment. Various generators, loads, reactors are connected as per the capacity of the network as seen in Figure 2. The deemed power network is simulated in PSCAD™ software to create tremendous data simultaneously using a multi-run block. Various system parameters are widely varied for huge data collection in PSCAD™ software. The developed algorithm is validated in MATLAB software. Set of CTs (CT1 and CT2) are connected to high voltage as well as on the low voltage side of the dedicated transformer sequentially to measure currents on both the sides of the transformer-1. In-zone faults are simulated on the transformer windings (at a different percentage of winding) as well as inside the CTs protective zone, as can be seen from Figure 2. Similarly, an undesired current path can be implicated outside CT protective zone to simulate the external fault of the considered power system.
### Table 1: Parameter variation and training as well as testing data segregation for a variety of inrush cases

| Considered variable | Initial inrush | Initial inrush (training data) | Sympathetic inrush | Sympathetic inrush (training data) | Recovery inrush | Recovery inrush (training data) |
|---------------------|----------------|--------------------------------|--------------------|------------------------------------|----------------|---------------------------------|
| Source impedance (SI) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) |
| CB switching instant | 6 (0°, 30°, 60°, 90°, 120°, 150°) | 5 (0°, 30°, 90°, 120°, 150°) | 5 (0°, 30°, 90°, 120°, 150°) | 5 (0°, 30°, 90°, 120°, 150°) | 5 (0°, 30°, 90°, 120°, 150°) | 5 (0°, 30°, 90°, 120°, 150°) |
| Load angle (δ) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) |
| Residual flux | 6 (0%, 10%, 25%, 45%, 60%, 75%) | 5 (0%, 10%, 45%, 60%, 75%) | – | – | – | – |
| Total | 540 | Training data = 375, hence testing data = 165 | Training data = 75, hence testing data = 15 | 90 | Training data = 75, hence testing data = 15 |

Total inrush data generated = 720 (540 + 90 + 90)
Training data for inrush cases = 525 (375 + 75 + 75)
Testing data for inrush cases = 195 (165 + 15 + 15)

### 3.1 Data collections

As shown in Table 1, 540 data are provoked for the initial inrush case, similarly, for sympathetic inrush condition, 90 data are fetched and the same numbers of data (i.e. 90) are created for recovery inrush condition by altering various parameters of the considered system. For initial inrush condition source impedance, residual flux, fault inception angle and load angle are varied. For sympathetic inrush condition, from Figure 2 it can be seen that another transformer named ‘transformer-2’ is connected in parallel to the transformer-1. To observe the effect of sympathetic inrush condition on the considered transformer-1, transformer-2 is energized. Hence, as a whole, 720 inrush data are generated by varying different parameters including source impedance, switching instant, residual flux and load angles. Among the total 720 data, 525 data are picked up for training and the leftover 195 data are kept for validation purposes.

As per Table 2, various faults are simulated by changing system parameters for turn to turn fault, inter-winding fault and intra winding fault within the transformer protection zone. As far as inter-winding fault of the transformer is concerned, there are ten types of inter-winding faults possible in the transformer as seen in Table 2. To understand this thoroughly, let us have a look at Figure 3. There are three L-g, three L-L, three L-L-g and one symmetrical, that is, L-L-L fault (total ten faults) possible on any one winding of a three-phase power transformer. Hence, there may be a possibility of such ten faults on primary winding and ten on the secondary winding of the transformer. If the fault occurs between two turns of the same winding, then it is a turn-to-turn fault as shown in Figure 3. In Figure 3, a turn-to-turn fault is illustrated which takes place between two turns of the R phase (primary side).
Moreover, intra-winding fault is a type of fault which takes place between the same phase of winding but of opposite sides. The intra-winding fault can be more understandable with the help of an example that is shown in Figure 3. In Figure 3, windings of red phases of both primary and secondary sides get short-circuited as they are wounded on a common limb of the transformer core. There are three possible intra-winding faults such as R-r fault, Y-y fault and B-b fault.

Around 3240 data are created for the case of turn to turn fault, 11,340 data are collected for primary to secondary intra-winding fault. For internal inter-winding fault, 75,600 data are provoked. Collectively, total of 90,180 data are fetched for in-zone fault conditions alone, from which 54,000 data are accessed for training and the last 36,180 data are exploited for testing purposes.

Similarly, as shown in Table 3, various external faults have been simulated on both 13.8 kV bus-1 and 220 kV bus-2 as well as on transmission lines. Total 48,600 data generated for external fault conditions from which 33,750 and 14,850 data are segregated as shown in Table 3 for training and testing purposes, respectively. The detailed separation of the data can be seen from Table 3 for various parameter variations.

Table 4 shows the data collection for over-flux and cross country fault [20] cases. The term, cross country fault can be described as two simultaneous faults occurring on two diverse locations in the dedicated power system [21]. As shown in Table 4, for the generation of over-fluxing conditions, variation in voltage, frequency, source switching angle, residual flux and percentage of load rejection are considered. It is true that the flux depends on $V/f$ ratio of the transformer ($\Phi/\alpha(V/f)$). So, over-fluxing is a phenomenon of the transformer when the voltage of the transformer is increasing or the frequency is decreasing. Moreover, the variation of other parameters like source switching angle, residual flux and percentage of load rejection also impact on the over-fluxing condition of the transformer. On the other hand, for cross country fault types of fault, fault location and fault resistances are varied. A total of 2268 data are generated for this miscellaneous condition. Among these 2268 total cases, 1212 cases are taken for training and the leftover 1056 cases are considered for testing purposes.

Table 5 shows the consolidated Intrush, miscellaneous and fault data simulated in considered power systems for validation of the projected algorithm. For in-zone fault case, 54,000 data are used for training purposes out of 90,180 total internal fault data and residual 36,180 data are consumed for the testing phase. From 48,600 total external fault data, 33,750 cases are utilized in training and the left behind 14,850 data are employed for testing purposes. Moreover, for intrush condition, 525 data are utilized for training from 720 total intrush data and another leftover 195 data are utilized for the testing case. Also from a total of 2268 data of miscellaneous condition, 1212 cases are taken for training and the leftover 1056 cases are considered for testing purposes. Overall, from the consolidated 1,41,768 data, 89,487 data are utilized for training and the remaining 52,281 data are availed for testing purposes.

Moreover, intra-winding fault is a type of fault which takes place between the same phase of winding but of opposite sides. The intra-winding fault can be more understandable with the help of an example that is shown in Figure 3. In Figure 3, windings of red phases of both primary and secondary sides get short-circuited as they are wounded on a common limb of the transformer core. There are three possible intra-winding faults such as R-r fault, Y-y fault and B-b fault.

Around 3240 data are created for the case of turn to turn fault, 11,340 data are collected for primary to secondary intra-winding fault. For internal inter-winding fault, 75,600 data are provoked. Collectively, total of 90,180 data are fetched for in-zone fault conditions alone, from which 54,000 data are accessed for training and the last 36,180 data are exploited for testing purposes.

Similarly, as shown in Table 3, various external faults have been simulated on both 13.8 kV bus-1 and 220 kV bus-2 as well as on transmission lines. Total 48,600 data generated for external fault conditions from which 33,750 and 14,850 data are segregated as shown in Table 3 for training and testing purposes, respectively. The detailed separation of the data can be seen from Table 3 for various parameter variations.

Table 4 shows the data collection for over-flux and cross country fault [20] cases. The term, cross country fault can be described as two simultaneous faults occurring on two diverse locations in the dedicated power system [21]. As shown in Table 4, for the generation of over-fluxing conditions, variation in voltage, frequency, source switching angle, residual flux and percentage of load rejection are considered. It is true that the flux depends on $V/f$ ratio of the transformer ($\Phi/\alpha(V/f)$).

So, over-fluxing is a phenomenon of the transformer when the voltage of the transformer is increasing or the frequency is decreasing. Moreover, the variation of other parameters like source switching angle, residual flux and percentage of load rejection also impact on the over-fluxing condition of the transformer. On the other hand, for cross country fault types of fault, fault location and fault resistances are varied. A total of 2268 data are generated for this miscellaneous condition. Among these 2268 total cases, 1212 cases are taken for training and the leftover 1056 cases are considered for testing purposes.

Table 5 shows the consolidated Intrush, miscellaneous and fault data simulated in considered power systems for validation of the projected algorithm. For in-zone fault case, 54,000 data are used for training purposes out of 90,180 total internal fault data and residual 36,180 data are consumed for the testing phase. From 48,600 total external fault data, 33,750 cases are utilized in training and the left behind 14,850 data are employed for testing purposes. Moreover, for intrush condition, 525 data are utilized for training from 720 total intrush data and another leftover 195 data are utilized for the testing case. Also from a total of 2268 data of miscellaneous condition, 1212 cases are taken for training and the remaining 1056 facts are considered for testing purposes. Overall, from the consolidated 1,41,768 data, 89,487 data are utilized for training and the remaining 52,281 data are availed for testing purposes.
### TABLE 3  Parameter variation and training as well as testing data separation for a variety of external faults

| Considered Parameter | Fault on the 13.8 kV & 220 kV buses (No. of training data) | Fault on the 13.8 kV & 220 kV Buses (No. of training data) | Fault occurred on 220 kV, 150 km line (No. of training data) | Fault occurred on 220 kV, 150 km line (No. of training data) |
|----------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------------------------------|
| Fault type (Ftype)   | 40 (10 ('L-G', 'LL-G', 'LL', 'LLL') * 2 (with and without CT saturation)) | 40 (10 ('L-G', 'LL-G', 'LL', 'LLL') * 2 (with and without CT saturation)) | 10 ('L-G', 'LL-G', 'LL', 'LLL') | 10 ('L-G', 'LL-G', 'LL', 'LLL') |
| Source impedance (SI) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) | 3 (80%, 100%, 120%) |
| Fault location (FL) | 1 | 1 | 5 | 5 |
| Fault inception angle (FIA) | 6 (0°, 30°, 60°, 90°, 120°, 150°) | 6 (0°, 30°, 60°, 90°, 120°, 150°) | 6 (0°, 30°, 90°, 120°, 150°) | 6 (0°, 30°, 90°, 120°, 150°) |
| Fault resistance (Rf) | 6 (0 Ω, 5 Ω, 10 Ω, 15 Ω, 25 Ω) | 6 (0 Ω, 5 Ω, 10 Ω, 15 Ω, 25 Ω) | 5 (0 Ω, 5 Ω, 10 Ω, 15 Ω, 25 Ω) | 5 (0 Ω, 5 Ω, 10 Ω, 15 Ω, 25 Ω) |
| Load angle (δ) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) |
| Total | 21,600 | 21,600 | 27,000 | 27,000 |

Total data generated for external fault condition = 48,600 (21,600 + 27,000)

Training data considered for external fault condition = 33,750 (15,000+18,750)

Testing data considered for external fault condition = 14,850 (6600+8250)

### TABLE 4  Training as well as testing data separation for over-flux and cross country fault (miscellaneous conditions)

| Considered parameter | Over-fluxing condition (training data) | Over-fluxing condition (training data) | Cross country fault/simultaneous fault (training data) | Cross country fault/simultaneous fault (training data) |
|----------------------|----------------------------------------|----------------------------------------|--------------------------------------------------------|--------------------------------------------------------|
| Percentage increase in primary side voltage | 5 (25%, 30%, 35%, 40%, 45%) | 4 (25%, 30%, 40%, 45%) | -- | -- |
| Frequency (Hz) | 5 (48.5, 49, 49.5, 50, 50.5) | 5 (48.5, 49, 49.5, 50, 50.5) | -- | -- |
| Source switching Angel | 6 (0°, 30°, 60°, 90°, 120°, 150°) | 5 (0°, 30°, 90°, 120°, 150°) | 5 (0°, 5°, 10°, 15°, 20°) | 5 (0°, 5°, 10°, 15°, 20°) |
| Residual flux | 5 (0%, 10%, 25%, 45%, 60%) | 4 (0%, 10%, 45%, 60%) | 6 (0Ω, 5Ω, 10Ω, 15Ω, 20Ω, 25Ω) | 6 (0Ω, 5Ω, 10Ω, 15Ω, 20Ω, 25Ω) |
| Percentage of load rejection | 3 (100%, 80%, 70%) | 3 (100%, 80%, 70%) | 1 (L-G) | 1 (L-G) |
| Fault type (Ftype) | -- | -- | 3 ('Internal to Internal', 'Internal to External', 'External to External') | 3 ('Internal to Internal', 'Internal to External', 'External to External') |
| Fault location | -- | -- | 6 | 4 |
| Fault resistance (Rf) | -- | -- | (0Ω, 5Ω, 10Ω, 15Ω, 20Ω, 25Ω) | (0Ω, 5Ω, 15Ω, 20Ω, 25Ω) |
| Total | 2250 | Training data = 1200 Hence, testing data = 1050 | 18 | Training data = 12, hence, testing data = 6 |

Total data generated for miscellaneous condition = 2268 (2250+18)  
Training data considered for miscellaneous condition = 1212 (1200+12)  
Testing data considered for miscellaneous condition = 1056 (1050+6)
TABLE 5  Total training as well as testing data collection and its segregation for a variety of cases for simulation

| Considered case       | No. of training data | No. of testing data | Total data  |
|-----------------------|----------------------|---------------------|-------------|
| Internal fault        | 54,000               | 36,180              | 90,180      |
| External fault        | 33,750               | 14,850              | 48,600      |
| Inrush                | 525                  | 195                 | 720         |
| Miscellaneous conditions | 1212               | 1056                | 2268        |
| Total                 | 89,487               | 52,281              | 1,41,768    |

FIGURE 4  Flow chart of the proposed CNN-XGBoost technique. CNN, Convolution neural network; XGBoost, Extreme Gradient Boosting

4  | FLOWCHART OF THE PROJECTED PROTECTION TECHNIQUE FOR INRUSH AND FAULT CONDITION IDENTIFICATION

Figure 4 shows a fundamental flow chart of the projected CNN-XGBoost scheme. Initially, the required data are acquired/gathered from the real field of the simulated system. The abnormal condition detection algorithm [31], [32] identifies the occurrence of an abnormal condition. If the abnormal condition is sensed, then the algorithm moves one step forward. The collected data are then sampled in the required number of samples based on the Nyquist criteria. The moving window captures the waveform of the selected quantities. After collecting the data by sampling, each sample is extracted to fetch the desired statistical data. The feature extraction is performed with the help of CNN. The already trained model of XGBoost screening out the disturbance data based on the waveform pattern. If XGBoost detects the disturbance as an internal fault, then it immediately sends a trip signal ('1') to the circuit isolator otherwise the system remains stable ('0') and the algorithm goes to the next set of samples.

5  | PARAMETER CONFIGURATION OF THE PROPOSED TECHNIQUE AND SOFTWARE RESULT ANALYSIS

5.1  | Parameter configuration

Before moving further for result analysis, let’s talk about parameter configuration for the CNN and XGBoost. The dedicated CNN framework is comprised of three convolutional layers and two fully connected layers. The convolutional layer is the most important layer to extort structural features. Each convolutional layer is chased by a MaxPooling layer and triggered by a Rectified Linear Unit (ReLU) function. MaxPooling is efficient to lower down the dimensions of the features and consequently speed up the calculations. The fully connected layer is chased by a Dropout operation and is triggered by a ReLU function. The Dropout operation can avoid over-fitting of the network effectively. As mentioned, the convolutional layer takes out the relatively specific features, while the fully connected layer extorts relatively abstractive features. These combined extracted features are used to extort effective features.

The 1D CNN parameterizations are considered as [60; 40; 40] neurons make three convolution layers. The fully connected layer is comprised of 20 neurons. The size of the output (full connected) layer is 2, which is based on types of classes (Internal Fault, Not Internal Fault) and only 1 input neuron for the input signal of 240 (1D) current samples. Two parameters of the 1D CNN are emphasized to optimize the system performance which is kernel size ($K$) as well as the subsampling factor ($ss$). These two parameters are valued as 9 and 4, sequentially. Here, the $K$ and $ss$ are decided based on the trial-and-error method [27]. However, if required, one can go for parameter optimization methods such as grey wolf optimizer [33]. In this case, there is no need to make the system more complex by adopting a parameter optimization method as the considered system is not so complex.

Moreover, the least train classification error is set at 0.5 percentage and the highest number of BP iterations is limited to 100. If any above-mentioned criteria are satisfied, the BP training brings to an end. The learning factor ($\epsilon$) in the beginning is valued at 0.001. From the afterward iteration, if the MSE is further lowered down, the learning factor ($\epsilon$) is augmented by 5%; otherwise, it is reduced by 30% [29].

Moreover, in the case of the XGBoost model, four important parameters are given full concentration, that is, a number of iterations, max delta step, gamma, max.depth [34]. The number of iterations and max.depth are utilized to manage the number of model iterations and ceiling depth limited for a tree, respectively. The loss reduction requirement is minimized by the
### TABLE 6
Classification accuracy of the proposed scheme for different abnormalities

| Sr. No. | Types of stimulus | Type of faults/abnormalities | Number of test cases | TC    | FC    | Accuracy (%) |
|---------|-------------------|------------------------------|----------------------|-------|-------|--------------|
| 1       | All the types of in-zone faults | Turn to turn faults | 990 | 986    | 4     | 99.60        |
|         |                   | Primary to secondary faults winding | 4590 | 4588   | 2     | 99.96        |
|         |                   | Internal winding faults | 30,600 | 30,588 | 12    | 99.96        |
| 2       | Faults outside the protective zone | Faults on 13.8 kV & 220 kV Bus (with CT saturation as well as without CT saturation) | 6600 | 6596   | 4     | 99.94        |
| 3       | Inrush conditions | Initial inrush | 165 | 165    | 0     | 100.00       |
|         |                   | Sympathetic inrush | 15 | 15     | 0     | 100.00       |
|         |                   | Recovery inrush | 15 | 15     | 0     | 100.00       |
| 4       | Miscellaneous conditions | Over-flux condition | 1050 | 1050   | 0     | 100.00       |
|         |                   | Cross country fault | 6 | 5      | 01    | 83.33        |
| Total data |                       |                            | 52,281 | 52,253 | 28    | 99.95        |

Abbreviations: TC, true classified; FC, false classified.

gamma parameter which makes additional separation of a leaf node of the particular tree. However, the positive value of the Max delta parameter helps in logistic regression when the class of data is extremely disparity. It is found that when the numbers of iterations are reached to 300, the considered model gives the best performance on the considered datasets. During less iteration, the properties of the datasets are not fully explored. On the other hand, for a higher number of iterations, the properties of the datasets are exaggerated, and consequently, the training set is not fully valued. Further, a reasonable value for max.depth should be chosen. The lower value of max.depth results in a lower rate of classification accuracy because of insufficient tree depth of the XGBoost model. Conversely, when max.depth is settled too high, the risk of over-fitting the training data is increased. Hence, after analyzing a moderate value of max.depth is settled to 5. Also, it is noted that with an increase in the magnitude of gamma, the downward trend in accuracy is observed, and hence the optimum value of gamma to be selected is 0. Further for max delta step, it is observed that if the max delta step is chosen 1 then the XGBoost model gives the best performance.

#### 5.2 Software result analysis

After collecting a tremendous data set of inrush, fault and miscellaneous cases, the proposed algorithm is ready for its training and testing phase. The test cases which are classified truly are denoted true classified. Conversely, test cases that are identified wrongly are indicated false classified (FC). Performance check of the projected technique is performed on 52,281 test cases and the result is derived in terms of classification accuracy as can be seen from Table 6. It is to be observed from Table 6 that the projected technique discriminates in-zone fault of the transformer with a classification accuracy of more than 99.60%. Further, the proposed scheme gives a fault identification accuracy of 99.94% during faults outside the transformer protective zone, that is, on the adjoining transmission line and each of the bus (13.8 and 220 kV) under minor to severe CT saturation conditions. Moreover, based on current signal samples of magnetizing inrush (distinctive wave pattern), it provides 100% classification accuracy during all types of inrush conditions. For miscellaneous conditions, the presented scheme provides promising accuracy. Overall, the proposed scheme gives an identification accuracy of 99.95%. The obtained results verify the efficacy of the proposed technique for varying test cases including fault types, fault locations, cross country faults, over-fluxing situations and faults with high resistance and CT saturation conditions.

#### 6 COMPARISON OF THE ANTICIPATED SCHEME WITH OTHER PROTECTIVE SCHEMES

To prove the outstanding performance of the anticipated combined CNN-XGBoost technique, a fault classification accuracy based comparison is carried out with the conventional transformer protection techniques. Table 7 shows a comparison of classification accuracy between the proposed combined CNN-XGBoost scheme with RVM [1] and HE-ELM [4] based transformer protection schemes. It is a trivial observation as displayed in Table 7, the rectification accuracy of the suggested scheme is more than that of the present protective schemes (RVM and HE-ELM) in all types of test cases whether it is an only abnormality or a real internal fault. Overall classification accuracy of the CNN-XGBoost scheme is 99.95%, while the same for RVM is 99.47% and for HE-ELM it is 99.61%. Classification accuracy during various test cases can be analysed in detail from Table 7. Hence, from the above comparative analysis, one can say that the proposed CNN-XGBoost scheme is a better fault classifier tool than the existing protection techniques used for the transformer.

Another aspect of comparison which can be considered is discrimination time (DT). The time required to classify the
TABLE 7  Comparison of the anticipated CNN-XGBoost technique with RVM and HE-ELM scheme

| Sr. No. | Type of fault cases/abnormal condition | Total test cases | TC | FC % | TC | FC % | TC | FC % |
|---------|---------------------------------------|-----------------|----|------|----|------|----|------|
| 1       | Internal faults                        | 36,180          | 36,162 | 18  | 99.95 | 35,990 | 190 | 99.47 | 36,047 | 133 | 99.63 |
| 2       | External faults                        | 14,850          | 14,841 | 9   | 99.94 | 14,784 | 66  | 99.56 | 14,791 | 59  | 99.60 |
| 3       | Inrush conditions                      | 195             | 195 | 0   | 100.00 | 195 | 0 | 100.00 | 195 | 0 | 100.00 |
| 4       | Miscellaneous conditions               | 1056            | 1055 | 1   | 99.91 | 1037 | 19  | 98.20 | 1042 | 14  | 98.67 |
| Total   |                                       | 52,281          | 52,253 | 28  | 99.95 | 52,006 | 275 | 99.47 | 52,075 | 206 | 99.61 |

Discrimination time (DT) 22 ms 27 ms 30 ms

Abbreviations: CNN, Convolution neural network; FC, false classified; HE-ELM, hierarchical ensemble of extreme learning machine; RVM, relevance vector machine; TC, true classified; XGBoost, extreme gradient boosting; %η, percentage classification accuracy.

FIGURE 5  Proposed hardware setup for validation of combined CNN-XGBoost classification scheme. CNN, Convolution neural network; XGBoost, Extreme Gradient Boosting

disturbance that whether it is a simple abnormality or a real internal fault is considered as DT. It has been examined from the comparison that the DT required by a combined CNN-XGBoost scheme to discriminate a disturbance is in the order of 22 ms while for RVM and HE-ELM, it is 27 and 30 ms, respectively. Hence, from the fault or abnormality identification time point of view, the anticipated algorithm is faster than the other considered AI-based fault classification methods.

7  HARDWARE SETUP FOR VARIOUS RESULT ANALYSES

Figure 5 depicts the hardware setup prepared for authentication of the projected protection scheme in a real-time scenario. The hardware prototype is developed in a laboratory environment using 50 kVA, three-phase 440/220V transformer. Various faults and abnormal conditions are created sequentially and sampled data of these events are recorded. The high-voltage winding of the transformer (440 V) is supplied from the main power of the State Electricity Board (utility supply). The low-voltage winding of the transformer (220 V) is connected with a separate generator available in the laboratory. A three-phase load is connected on a secondary side and heavy-duty contactors are utilized as circuit breakers in the line. Variable resistors and inductors are inserted on both sides of the transformer to imitate the consequences of a transmission line. In Figure 5, the control circuit is also displayed which controls the protective scheme of the dedicated transformer. Moreover, as can be seen from Figure 5, the considered transformer is multi-tapping at various voltage levels and hence inter-turn and internal fault conditions can be easily created in between the transformer windings. Various external as well as internal faults are generated using fault switches (which can be seen exactly below the transformer in Figure 5) and 12-A, 18-ohm variable resisters. CTs of apposite ratings are coupled on either side of the transformer to measure and analyze the current. IEEE advisory [35] gives direction to build the proposed hardware set-up. The CT second side currents from both sides are given to the 32-bit, 10-channel analog to digital converter (ADC-ADS1263) through signal conditioning unit (SCU). SCU is utilized to level down the current signal before furnished to the ADC. The hardware prototype comprised of SCU, ADC with peripheral devices and Intel core i7, 16-GB RAM, 4.2 GHz CPU with Windows 10 operating system. Various internal, as well as external faults such as L-G, LL, LL-G and LLL-G (at lower voltage scale) are performed using fault switches. The currents are acquired at 4 kHz sampling rate. After that, the sampled currents are stored in buffer memory before utilizing it for feature extraction and classification purpose.

Based on the hardware setup as explained in this section, Table 8 exhibits the data created for different test cases. Here, collected hardware-based data are added with the software data to maintain the large data requirement of CNN. The detailed training data, testing data, and total data for various cases are depicted in Table 8.

It can be observed from Table 9 that the classification accuracy achieved for hardware as well as software data remains almost the same (99.94%). It is to be noted from Table 9 that the proposed scheme provides better fault classification...
TABLE 8  Consolidated data for software and hardware analysis

| Cases               | Training data (software) | Training data (hardware) | Total training data | Testing data (software) | Testing data (hardware) | Total testing data | Total data for hardware & software,(Training+Testing) |
|---------------------|--------------------------|--------------------------|---------------------|-------------------------|-------------------------|-------------------|----------------------------------------------------|
| Internal fault      | 54,000                   | 22                       | 54,022              | 36,180                  | 15                      | 36,195            | 90,217                                             |
| External fault      | 33,750                   | 22                       | 33,772              | 14,850                  | 15                      | 14,865            | 48,637                                             |
| Inrush              | 525                      | 11                       | 536                 | 195                     | 6                       | 201               | 737                                                |
| Miscellaneous       | 1212                     | 8                        | 1220                | 1056                    | 5                       | 1061              | 2281                                               |
| Total               | 89,487                   | 63                       | 89,550              | 52,281                  | 41                      | 52,322            | 1,41,872                                           |

TABLE 9  Classification accuracy of the anticipated CNN-XGBoost technique for combined real-time data generated on hardware setup and software data

| Sr. No. | Type of abnormalities | Combined test cases (Hardware + Software) | Projected scheme (CNN-XGBoost) |
|---------|-----------------------|-------------------------------------------|-------------------------------|
| 1       | Internal faults       | 36,195                                    | TC 36,176 FC 19 Accuracy (%) 99.95 |
| 2       | External faults       | 14,865                                    | TC 14,855 FC 10 Accuracy (%) 99.93 |
| 3       | Inrush conditions     | 201                                        | TC 201 FC 0 Accuracy (%) 100.00 |
| 4       | Miscellaneous         | 1061                                       | TC 1060 FC 1 Accuracy (%) 99.91 |
| Total   |                       | 52,322                                     | TC 52,292 FC 30 Accuracy (%) 99.94 |

RVM = 99.45% accuracy
HE-ELM = 99.55% accuracy

CNN, Convolutional neural network; FC, False classified; HE-ELM, Hierarchical ensemble of extreme learning machine; RVM, Relevance vector machine; TC, True classified; XGBoost, Extreme gradient boosting

accuracy during an internal fault (99.95%) which shows the reliability of the algorithm. On the other hand, the fault discrimination accuracy of 99.93% during external disturbance and 100% during inrush conditions indicate high stability of the proposed technique. However, one case is FC during the miscellaneous condition and it provides 99.91% accuracy. Overall, from 52,322 test cases (Hardware + Software), 52,292 cases are correctly classified and only 30 cases are misclassified and consequently, overall 99.94% accuracy is obtained for the proposed combined CNN-XGBoost scheme. Moreover, during hardware validation, the RVM and HE-ELM scheme also give indistinguishable classification accuracy as that achieved during software analysis. It is concluded according to Table 9 that the anticipated combined CNN-XGBoost methodology provides promising classification accuracy than the RVM- and HE-ELM-based classifier techniques.

The waveform during hardware analysis of various test cases is displayed in Figure 6. These results have been captured in DSO using *.csv files and plotted in MATLAB software.

8 | CONCLUSION

The presented research paper shows the competency of the latest CNN-XGBoost combination to classify internal fault conditions from various abnormalities in the transformer. The method is well described along with the algorithm steps in the paper. CNN is applied as a high-level feature extractor; on the other side, XGBoost is utilized as a classification tool. A seven-bus power system network is considered and modelled in the PSCAD software. Various inrush, internal fault, external fault, over-flux and cross country fault cases are simulated by varying different systems and fault parameters. To collect a huge amount of data with variation, a multi-run feature of PSCAD software is used. Collected data are migrated to MATLAB software to validate the proposed algorithm. The algorithm is successfully validated in the MATLAB software and the classification accuracy of the projected protection method is examined. Moreover, on the basis of classification accuracy as well as DT, the proposed combined CNN-XGBoost technique is compared with RVM and HE-ELM scheme. To authenticate the implementation of the projected scheme, a hardware prototype is developed in the laboratory and various real-time fault data are collected. The collected data are then used for training as well as testing purposes of the presented protection methodology. Again the classification accuracy for hardware data is examined for proposed as well as existing techniques and found that the CNN-XGBoost technique provides promising results against RVM & HE-ELM technique. The proposed method provides around 99.95% classification accuracy within 22 ms to properly recognize the internal fault condition of the transformer. Hence, according to the above analysis, the proposed methodology is competent to classify any type of fault in minimum classification time. The hardware
validation proves that the methodology is implementable and can be used in the real-field to protect the power transformer effectively.

ACKNOWLEDGEMENT

This work is financially supported by the Department of Science and Technology (DST, SERB-India) under project no. DST/EMR/2016/00641.

ORCID

Maulik Raichura https://orcid.org/0000-0002-5568-5575
Nilesh Chothani https://orcid.org/0000-0002-9936-1687

REFERENCES

1. Patel, D., et al.: Design and development of fault classification algorithm based on relevance vector machine for power transformer. IET Electr. Power Appl. 12(4), (2018)
2. 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 Delivery 28(3), 1508–1515 (2013)
3. Tripathy, M., Maheshwari, R.P., Verma, H.K.: Power transformer differential protection based on optimal probabilistic neural network. IEEE Trans. Power Delivery 25(1), 102–112 (2010)
4. 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. 14(1), 111–121 (2020)
5. Chorowski, J., Wang, J., Zurasda, J.M.: Review and performance comparison of SVM- and ELM-based classifiers. Neurocomputing. 128(Supplement C), S07–S16 (2014)
6. Bucurca, M., Dogaru, R., Dogaru, I.: A comparison of extreme leaning machine and support vector machine classifiers. In: 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 471–474, IEEE (2015)
7. Balaga, H., Gupta, N., Vishwakarma, D.N.: GA trained parallel hidden layered ANN based differential protection of three phase power transformer. Int. J. Electr. Power Energy Syst. 67, 286–297 (2015)
8. Shah, A.M., Bhavesh, B.: A new adaptive differential protection scheme for tap changing power transformer. Int. J. Emerg. Electr. Power Syst. 16, 339 (2015)
9. Bagheri, S., Moravej, Z., Gharehpetic, G.B.: Effect of transformer winding mechanical defects, internal and external electrical faults and inrush currents on performance of differential protection. IET Gener. Transm. Distrib. 11(10), 2508–2520 (2017)
10. Ballal, M.S., et al.: Online condition monitoring system for substation and service transformers. IET Electr. Power Appl. 11(7), 1187–1195 (2017)
11. Ghanbari, T., Samei, H., Ghafoori-fard, J.: New approach to improve sensitivity of differential and restricted earth fault protections for industrial transformers. IET Gener. Transm. Distrib. 10(6), 1486–1494 (2016)
12. Dubie, G., Cukane, A.: New algorithm for detecting power transformer faults based on M-robust estimation of sound signals. IET Gener. Transm. Distrib. 8(6), 1117–1126 (2014)
13. Hooshary, A., Sanaye-Pasand, M.: Waveshape recognition technique to detect current transformer saturation. IET Gen. Transm. Distrib. 9(12), 1430–1438 (2015)
14. Aboos, A.A., Detection of current transformer saturation based on variational mode decomposition analysis. IET Gener. Transm. Distrib. 10(11), 2658–2669 (2016)
15. Chothani N.G., Bhalja B.R. New Algorithm for current transformer saturation detection and compensation based on derivatives of secondary currents and Newton's backward difference formulae. IET Gener. Transm. Distrib. 8, (5), 841–850 (2014). https://doi.org/10.1049/iet-gtd.2013.0324
16. Mostafaei, M., Hadhjoo, F.: Flux-based turn-to-turn fault protection for power transformers. IET Gener. Transm. Distrib. 10(5), 1154–1163 (2016)
17. Dashi, H., et al.: Discriminating transformer large inrush currents from fault currents. Int. J. Electr. Power Energy Syst. 75, 74–82 (2016)
18. Lin, X., et al.: Internal fault fast identification criterion based on superimposed component comparison for power transformer. Int. J. Electr. Power Energy Syst. 73, 491–497 (2015)
19. Oliveira, M.O., Bretas, A.S., Ferreira, G.D.: Adaptive differential protection of three-phase power transformers based on transient signal analysis. Int. J. Electr. Power Energy Syst. 57, 366–374 (2014)
20. Lin, X., Weng, H., Wang, B.: Identification of cross-country fault of power transformer for fast unblocking of differential protection. IEEE Trans. Power Delivery. 24(3), 1079–1086 (2009)
21. ABB Oy, Distribution Automation Handbook (prototype) Power System Protection, Section 8.6. MV Feeder Earth-fault protection, 1MRS757287. Online: DAHandbook_Section_08p06_Feeder_EF_Protection_757287_ENa.pdf
22. Ren, X., et al.: A novel image classification model with CNN-XGBoost model. In: Kraetzer, C., Shi, Y.-Q., Dittmann, J., Kim, H.J., (Eds): Digital Forensics and Watermarking, pp. 378–390. Springer International Publishing, New York (2017)
23. Afrasiabi, S., et al.: Integration of accelerated deep neural network into power transformer differential protection. IEEE Trans. Ind. Inf. 16(2), pp. 865–876 (2019)
24. Gun, Y., et al.: Advances in Neural Information Processing Systems 2, pp. 396–404. Morgan Kaufmann Publishers Inc., San Francisco, CA (1990)
25. Borovyykh, A., Bohte, S., Oosterlee, C.W.: Conditional time series forecasting with convolutional neural networks. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 10614 LNCS, 729–730. (2017)
26. Kiranyaz, S., et al: Convolutional Neural Networks for Patient-Specific ECG Classification. Paper presented at the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015
27. Abdeljaber, O., et al.: Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. J. Sound Vib. 388, 154–170 (2017)
28. Kiranyaz, S., Ince, T., Gabbouj, M.: Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans. Biomed. Eng. 63(3), 664–675 (2016)
29. Ince, T., et al.: Real-time motor fault detection by 1-D convolutional neural networks. IEEE Trans. Ind. Electron. 63(11), 7067–7075 (2016)
30. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. Paper presented at the proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016
31. Wiot, D.: A new adaptive transient monitoring scheme for detection of power system events. IEEE Trans. Power Delivery. 19(1), 42–48 (2004)
32. Mohanty, S.R., Pradhan, A.K., Routray, A.: A cumulative sum-based fault detector for power system relaying application. IEEE Trans. Power Delivery. 23(1), 79–86 (2008)
33. Chen, X., et al: A self-adaptive 1D convolutional neural network for flight-state identification. Sensors (Switzerland). 19(2), (2019)
34. Song, R., et al: EXtreme gradient boosting for identifying individual users across different digital devices, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 43–54. Springer-Verlag, New York, NY (2016)
35. IEEE Guide for Protecting Power Transformers (Revision of IEEE Std C37.91-2000) (IEEE Power Engineering Society Sponsored by the Power System Relaying Committee, (2008)

How to cite this article: Raichura M, Chothani N, Patel D. Efficient CNN-XGBoost technique for classification of power transformer internal faults against various abnormal conditions. IET Gener Transm Distrib. 2021;15:972–985.
https://doi.org/10.1049/gtd2.12073