IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 2, 2021, accepted May 11, 2021, date of publication May 17, 2021, date of current version May 24, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3081015*

# A Data Mining Based Protection and Classification of Transients for Two-Core Symmetric Phase Angle Regulators

**PALLAV KUMAR BERA, (Student Member, IEEE), AND CAN ISIK, (Life Senior Member, IEEE)**
Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244, USA
Corresponding author: Pallav Kumar Bera (pkbera@syr.edu)

**ABSTRACT** Several conventional and non-conventional transient conditions cause differential relays associated with Phase Angle Regulators to malfunction. For Two-core Symmetric Phase Angle Regulators, this article investigates the suitability of time and time-frequency feature-based estimators to differentiate internal faults from other transient conditions such as overexcitation, external faults with current transformer (CT) saturation, and magnetizing inrush. Subsequently, the faulty core unit (series or exciting) is located, and the transients are identified. Six well-known classifiers are trained on features extracted from one-cycle of post transient 3-phase differential currents filtered by an event detector. Maximum Relevance Minimum Redundancy, Random Forest, and exhaustive search with Decision Trees are used to select the relevant wavelet energy, time-domain, and wavelet coefficient features respectively. The fault detection scheme trained on XGBoost classifier with hyperparameters obtained from Bayesian Optimization gives an accuracy of 99.8%. The reliability of the proposed scheme is verified with varying tap positions, noise levels, and transformer ratings; and under different conditions like CT saturation, fault during magnetizing inrush, series core saturation, low current faults, and integration of wind energy. As a potential application, the methodology can be deployed to supervise microprocessor-based differential relays to improve the security and dependability of the protection system.

**INDEX TERMS** Bayesian optimization, fault detection, feature selection, machine learning, phase shift transformer, transient classification, wavelet transform, xgboost.

## I. INTRODUCTION

Phase Shift Transformers or Phase Angle Regulators (PARs) control the steady-state power flow in parallel transmission lines and sometimes connect two independent grids. They ensure that contingency conditions do not exceed the ratings of transmission equipment. Their performance affects the continuous and stable operation of the power system. With a lower successful operating rate than the transmission lines, transformer protection systems are challenged under various conditions. Internal faults are electrically detected in a transformer mainly with differential, overcurrent, and ground fault relays. Differential relays detect and clear faults faster and locate them accurately. In general, electromechanical, solid-state, analog, and microprocessor-based relays are used as differential relays. Predominantly, differential relays are used to protect the standard and non-standard transformers, and their operation highly depends on appropriate analysis of different electromagnetic transient events [1].

Differential protection, however, suffers from traditional challenges of unwanted tripping in situations of magnetizing inrush, external faults with CT saturation, and overexcitation. These problems are addressed by current-based methods in two ways: using harmonic restraint and waveshape identification methods [2]. The changing complexity and operating modes in the power system have threatened the reliability of these methods. The percentage differential relay with restraint, actuated by restraining current and/or harmonic components of operating current is generally used in differential schemes. The second harmonic component identifies magnetizing inrush, and the fifth identifies overexcitation. The second harmonic restraint method [3] used to detect magnetizing inrush may fail because of lower second harmonics in transformers with a modern core [4]. Moreover,

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

the protection system's sensitivity is compromised due to higher second harmonics during internal faults with CT saturation and presence of distributed and series compensation capacitance [5]. The fifth harmonic restraint may also fail in case of internal faults during overexcitation. While the use of the fourth harmonic with the second in case of inrush and adaptive fifth harmonic pick up in case of overexcitation improves security, the challenges remain. External faults with CT saturation may also cause false trips if the settings of the dual-slope current differential relays are not set effectively [6]. Differential relays also fail to detect ground faults near neutral of grounded wye-connected transformer winding [1].

PARs are classified on the basis of the number of magnetic cores and the magnitude of source- and load-side voltages. Two-core Symmetric PARs or Indirect Symmetric PARs (ISPARs) have the same source- and load-side voltages with two cores: series and exciting (Fig.1A). They are the conventionally used PARs with higher security against high voltage levels. To regulate power flow, the exciting unit creates the required phase difference through the load tap changer, and the forward/backward transition can be achieved in the series secondary with a switch or change-over selectors in the exciting secondary [7]. Taking into account the high repair and replacement cost and to limit further damages PARs require a sensitive, secure, and dependable protection system. Maintaining dependability for in-zone faults and security against no-fault conditions is a challenge. Besides the traditional challenges associated with transformer differential protection, high sensitivity to detect turn-to-turn (*t-t*) and winding phase-to-ground faults, and security against series winding saturation are specific to PARs [8]. Furthermore, the phase compensation techniques used in standard differential protection with fixed phase shifts cannot be applied for the compensation of the phase-shift across the PARs with a non-standard phase shift [9], [10]. Consequently, special considerations are required while designing their protection system.

A two-element based differential protection is proposed in [11]. It performs well for internal faults and series saturation, but it suffers from other traditional and PAR specific challenges. Phase/magnitude compensation is proposed to address the non-standard phase shift in [12]. However, it requires tracking the tap positions and has a lower sensitivity for low current faults. Reference [13] proposes differential protection, which does not need the knowledge of tap positions. But it applies to hexagonal PARs only. Reference [8] proposes directional comparison-based protection, which provides overall protection addressing various challenges, however, it needs both current and voltage information to function. The present work attempts to provide a comprehensive alternative solution to the conventional and non-conventional protection challenges associated with a PAR using Machine Learning (ML).

Data Mining and ML-based methods which do not require predefined threshold values and mathematical models have been proposed to distinguish faults and disturbances in transformers in the last two decades [14]. Neural Networks (NN) [15], [16], Support Vector Machines (SVM) [17], [18], Decision Tree (DT) [16], [19], k-nearest neighbor (kNN) [20], and Random Forest (RF) [21] are some of the popular algorithms that has been used for differential protection of transformers. Although several such studies exist in transformer protection, this problem is insufficiently explored for PARs. Few in the literature have considered using ML to detect faults and other transients in PARs. In [22] internal faults were differentiated from inrush currents using the Wavelet Transform and classified with a NN. The problem has not been studied consistently because a complete PAR model with a provision to simulate internal faults is scarce. To adequately address this question, the ISPAR was modeled, internal faults were detected, and transients were classified for PARs and transformers in [23]; however, many problems like overexcitation, series core saturation, fault during magnetizing inrush, integration of renewables were not considered.

The present article studies the suitability of time, and time-frequency domain features to discriminate faults from transient disturbances like the magnetizing inrush, external faults with CT saturation, and the overexcitation for a PAR. The ISPAR is modeled in PSCAD using 2- and 3-winding transformers to simulate the transients. A series of time and wavelet features are extracted and then selected using feature selection algorithms. An exhaustive search strategy with DT algorithm as a wrapper is used to select wavelet coefficients. Maximum Relevance Minimum Redundancy and Random Forest algorithms are used to get the relevant wavelet energy and time-domain features respectively. Six classifiers trained and tested on 60552 transient cases simulated by changing the system parameters demonstrate the proposed scheme's validity. The scheme's stability is also examined under a variety of conditions, including fault during magnetizing inrush, series-winding saturation, CT saturation, the addition of an inverter-interfaced wind turbine, as well as with various transformer ratings, tap positions and noise levels.

The remainder of this paper is arranged in the following order. The transient events in the ISPAR are modeled and simulated in Section II. Section III presents the proposed differential protection scheme with the event detection, extraction and selection of features, and introduces the six classifiers. The performance of the classifiers for detection of faults, localization of faulty units, and classification of transients are presented in Section IV. Section V includes the assessments for the various non-conventional challenges that the PAR may encounter. The last section concludes the article.

## II. MODELING AND SIMULATION

PSCAD/EMTDC is used to model the ISPAR and simulate the electromagnetic transients. The rating of the ISPAR are: $S_n$ = 500MVA, $V_n$ = 230kV, maximum phase shift = $\pm 25°$. CT1 and CT2 are located on the two sides of the PAR. The fault model of ISPAR is not available in most simulation software. The single-phase 2-winding transformer

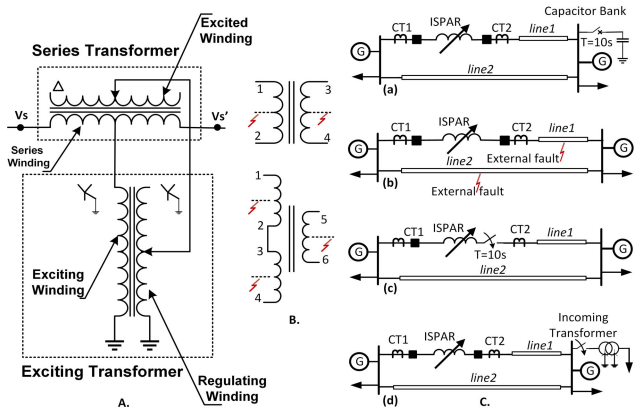


**FIGURE 1.** (A) ISPAR model, (B) 2-& 3-winding transformer fault model, (C) Simulation models for overexcitation, external fault with CT saturation, magnetizing inrush, and sympathetic inrush (top to bottom).

fault model needed for faults in the exciting unit and the single-phase 3-winding transformer fault model needed for faults in the series unit (Fig.1B) were designed in [23]. Equation (1) describes the voltage-current relationship for six-coupled coils of the 3-winding transformer. The self (*Li*) and mutual (*Mij*) inductances of the $6 \times 6$ matrix in equation (1) and *Li* and *Mij* of $4 \times 4$ matrix of the 2-winding transformer are computed from the voltage ratios, reactive part of the no-load current ($I_m$), and short-circuit tests. The consistency of the proposed protection scheme is highly dependent on the accuracy of the developed models since the learning data is obtained from system transients that depend on these models. Hence, the 2- and 3-winding transformer fault models are validated first in PSCAD. The saturation characteristics, percentage of turns faulted can be changed in the 2-and 3-winding transformers. The Appendix Section includes the Fortran script of the 1-phase 3-winding transformer.

$$\begin{bmatrix} V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \end{bmatrix} = \begin{bmatrix} L1 & M12 & M13 & M14 & M15 & M16 \\ M21 & L2 & M23 & M24 & M25 & M26 \\ M31 & M32 & L3 & M34 & M35 & M36 \\ M41 & M42 & M43 & L4 & M45 & M46 \\ M51 & M52 & M53 & M54 & L5 & M56 \\ M61 & M62 & M63 & M64 & M65 & L6 \end{bmatrix} \cdot \frac{d}{dt} \begin{bmatrix} I1 \\ I2 \\ I3 \\ I4 \\ I5 \\ I6 \end{bmatrix} \quad (1)$$

In the present analysis, the internal faults, overexcitation, external faults with CT saturation, and magnetizing and sympathetic inrush conditions for ISPAR are considered. These scenarios are studied successively in the sections that follow. In the simulations, the total run-time is 10.2s, switching time (ST) is 10.0s, and the duration of faults is 0.05s (3 cycles). The multi-run component in the master library is employed as needed during the simulations.

### A. INTERNAL FAULTS

The internal faults are simulated in primary (P) and secondary (S) sides of exciting and series units in the ISPAR. They include the faults occurring inside the enclosure and inside the CT locations. They are usually caused by insulation breakdown and require faster action by protective relays to limit the extent of the damage. The basic internal faults include short circuits and phase (ph) faults, *t-t*, and winding-winding faults. 46872 faults are simulated by varying the percentage of turns shorted (PTS), fault resistance (FR), faulty unit (FU), fault type (FT), fault inception time (FIT), phase shift (PS): forward & backward, and the PAR tap positions.

*Phase & Ground Faults:* These include winding ph-g faults (a-g, b-g, c-g), winding ph-ph-g faults (ab-g, ac-g, bc-g), winding ph-ph faults (ab, ac, bc), 3-ph and 3-ph-g faults. The values of different parameters (param) of the ISPAR used to simulate 33264 instances are shown in Table 1a.

*Turn-to-Turn (t-t) Faults:* Insulation failures are responsible for a major percentage of faults in a transformer. The insulation degrades over time with thermal, electrical, and mechanical stresses causing *t-t* faults which can develop into serious faults if go undetected [24]. They are challenging to detect, particularly when PTS is low. The values of different parameters resulting in 9072 cases are displayed in Table 1b.

*Winding-winding (w-w) faults:* Transformer aging and short circuits degrade the insulation between LV and HV winding and cause winding failure [24]. The values of different parameters used to obtain 4536 cases are listed in Table 1b.

**TABLE 1.** Parameters: (a) internal ph and g faults, and (b) *t-t* and *w-w* faults.

(a)

| *Param* | *Values* |
|---|---|
| *FR* | 0.01, 0.1 & 1 Ω (3) |
| *PTS* | 20%, 50%, 70% (3) |
| *FT* | lg, llg, ll, lll & lllg (11) |
| *FIT* | 10s to 10.0153s (12) |
| *FU* | Exciting (P & S) (2) & Series (P & S) (2) |
| *PS* | forward & backward (2) |
| *tap* | .2,.4,.6,.8,1[1&0.5 in exciting] |

(b)

| *Param* | *Values* |
|---|---|
| *FR* | 0.01, 0.5 & 1 Ω (3) |
| *PTS* | 20%, 50%, 70% (3) |
| *FIT* | 10s to 10.0153s (12) |
| *FU* | Exciting ph A,B,C (P & S) (6) & Series ph A,B,C (P & S) (6) |
| *PS* | forward & backward (2) |
| *tap* | .2,.4,.6,.8,1 [1&0.5 in exciting] |

### B. OVEREXCITATION

Faults due to over fluxing develop slowly and cause deterioration of insulation and may lead to major faults. They cause heating and vibration and can damage the transformer [25]. Since it is difficult for differential protection to control the amount of overexcitation a transformer can tolerate, tripping of the differential element during overexcitation is undesirable. Generally, 5th harmonic restraint is used to restrain the operation of differential relays [26]. Several conditions may lead to overexcitation in electrical systems. Here, two such situations have been modeled: overvoltage during load rejection and capacitor switching (Fig.1C). The typical differential current waveforms for these are shown in Fig.2a and Fig.2b. Parameter values are listed in Table 2a.

### C. EXTERNAL FAULT WITH CT SATURATION

External short circuit stresses the PAR and reduces the transformer life. A non-zero differential current flows due to

**TABLE 2.** Parameters: (a) Overexcitation, and (b) Magnetizing inrush and Sympathetic inrush.

(a)

| *Param* | *Values* |
|---|---|
| *switch* | load (3) & capacitor (3) |
| *ST* | 10s to 10.0153s (12) |
| *tap* | 0.2, 0.4, 0.6, 0.8, 1 (5) |
| *PS* | forward & backward (2) |
| Total=6×12×5×2=720 | |

(b)

| *Param* | *Values* |
|---|---|
| *RFD* | ±80, ±60, ±40, 0% in 3-phs (21) |
| *ST* | 10s to 10.0153s (12) |
| *tap* | 0.2, 0.4, 0.6, 0.8, 1 (5) |
| *PS* | forward & backward (2) |
| Total = 21×12×5×2 = 2520 | |

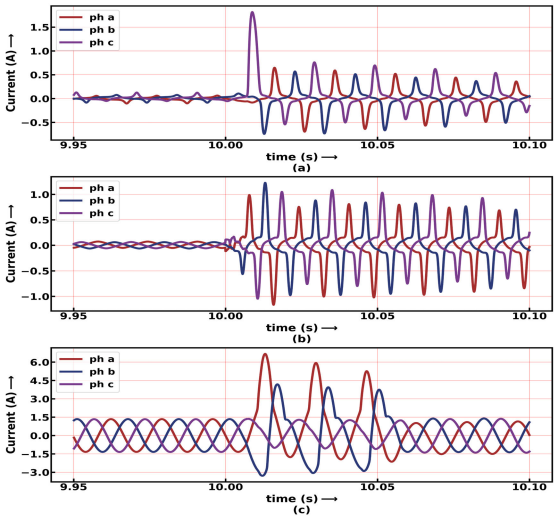


**FIGURE 2.** 3-phase differential currents for (a) load rejection, (b) capacitor switching, and (c) CT saturation during external faults.

**TABLE 3.** Parameters for external faults on *line1 & line2*.

| *Param* | *Values* |
|---|---|
| *FR* | 0.01, 0.1 & 1 Ω (3) |
| *FT* | lg, llg, ll, lll & lllg in 3 phs (11) |
| *FIT* | 10s to 10.0153s in steps of 1.38ms (12) |
| *tap* | 0.2 to full tap in steps of 0.2 (5) |
| *PS* | forward & backward (2) |
| *FL* | line1 & line2 (2) |
| Total = 3×11×12×5×2×2 = 7920 | |

saturation of CTs during external faults and may lead to false trips [27]. Raising the bias threshold of differential relays may ensure stability, but then the sensitivity for in-zone internal faults is reduced. The fault location (FL) is varied while simulating these cases (Fig.1C) besides FR, FT, FIT, tap position, and PS (Table 3). Fig.2c shows the differential current for an external *lg* fault with PS = forward, FIT = 10.0083s, FL = line1, and FR = 1Ω at full tap.

### D. MAGNETIZING INRUSH

When a transformer is energized, a high inrush current of the order of 10-15 times of normal current flows because of the saturation of the transformer core. Second harmonic restraint relays may fail to detect inrush currents in modern transformers having high flux density. In addition to the tap positions and PS, $\Phi_R$ and t′ present in the flux equation expressed as:

$$\Phi = \Phi_R + \Phi_m cos\omega t' - \Phi_m cos\omega(t + t') \qquad (2)$$

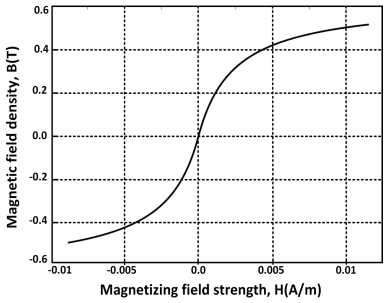


**FIGURE 3.** B-H curve of exciting transformer unit.

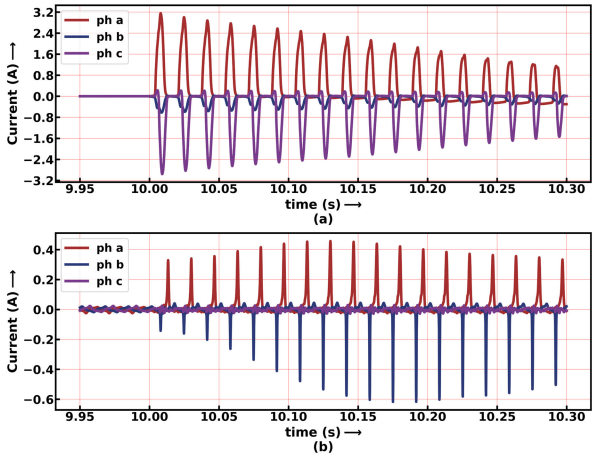


**FIGURE 4.** 3-phase differential currents for (a) Magnetizing inrush, and (b) Sympathetic inrush.

where $\Phi_R$ = residual flux density (RFD), $\Phi_m$ = maximum flux, and t′ = switching time (ST) are the important parameters [28]. To get the desired residual fluxes, DC currents are injected in the phases. The current values are obtained from the x-coordinates of the B-H curve (Fig.3). Table 2b shows the values of RFD, ST, PAR taps, and PS used to obtain the 2520 cases. Fig.4(a) shows typical differential currents for a magnetizing inrush with tap = full, ST = 10s, PS = forward, and RFD = 0 in all phases. The exciting transformer unit in the ISPAR is considered to be responsible for the inrush currents [9].

### E. SYMPATHETIC INRUSH

Sympathetic inrush can occur when a power transformer is switched on in a power network with already energized PARs (Fig.1C). The flux change per cycle which drives the PAR to saturation is given by equation (3).

$$\Delta\Phi = \int_{t}^{2\pi+t} [(R_{sys} + R_{par})i_1 + R_{sys} \cdot i_2] \qquad (3)$$

where $R_{sys}$ = system resistance, $R_{par}$ = resistance of PAR, and $i_1$ and $i_2$ are magnetizing currents of PAR and the incoming transformer [29]. This interaction between the incoming transformer and the PAR may lead to failure of differential relays. Some factors responsible for such mal-operations are: cores with soft magnetic material, application

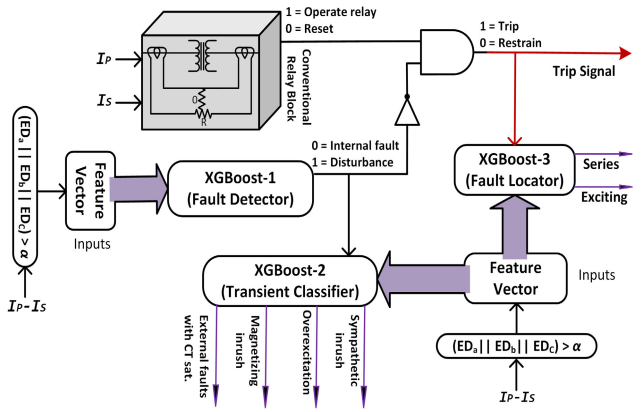


**FIGURE 5.** Flowchart for fault detection and localization.

of superconducting technology in windings, and CT partial transient saturation [4]. Here the incoming transformer is energized at t = 10s and the values of $\Phi_R$ and t′ which significantly influence the inrush currents are varied (See Table 2b). Fig.4(b) shows the differential currents for tap = 0.2, ST = 10.0069s, PS = backward, and no RFD.

The differential currents obtained from the 60552 transients of internal faults, overexcitation, external faults, and inrush currents simulated in this section will be preprocessed to obtain the relevant time and time-frequency features and used as inputs to the six classifiers for detection and classification of the transients in the succeeding sections. The complete dataset is available on IEEE Dataport [30].

## III. THE PROPOSED PAR DIFFERENTIAL PROTECTION

Fig.5 depicts the proposed time and time-frequency domain based protection and classification scheme having three applications: detection of internal faults (DIF), localization of faulty unit (LFU), and classification of transient disturbances (CTD). The event detector (ED) detects the change in the differential currents ($I_P$-$I_S$) if the ED index in any phase is more than the threshold $\alpha$ and registers one cycle of post transient 3-phase differential currents. These currents are preprocessed to obtain detailed wavelet-coefficients (WC), wavelet-energy (WE), and time-domain (TD) features. The proposed scheme can be seen as a design having three classifiers. The fault detector is the first classifier (Xgb-1). It recognizes the internal faults with "0" and transient disturbances with "1". Thus, Xgb-1, together with the NOT gate regulates the operation of the trip/restrain function block by obstructing the transient disturbances and allowing internal faults. The transient classifier is the second classifier (Xgb-2), which examines an event further if the output of Xgb-1 is "1". It can identify the disturbance responsible for faulty operation of the conventional relay block (CRB) (Xgb-1 is "1" & CRB is "1"). The fault locator is the third classifier (Xgb-3). It locates the defective transformer core unit as series or exciting (Xgb-1 is "0" & CRB is "1").

### A. EVENT DETECTION

The differential currents become non-zero when a power system transient occurs. The ED which detects this change and computes the fractional increase between cumulative sum of modulus of samples of two successive cycles is defined by equation (4).

$$ED(t) = \frac{\sum_{i=t}^{n_c+t} |Id_\phi(i)| - \sum_{i=t}^{n_c+t} |Id_\phi(i-n_c)|}{\sum_{i=t}^{n_c+t} |Id_\phi(i)|} \quad (4)$$

where $n_c$ is number of samples in one cycle, $Id$ is differential current, $\phi$ denotes the 3-phases, and i is the sample number starting at the second cycle. The 3-phase differential current samples are recorded by the ED filter from the time when

$$ED(t) \geq \alpha = 0.05 \quad (5)$$

in any of the three phases. In the absence of transients, ED(t) values are negligibly small [31]. These recorded samples are used for the feature extraction.

### B. FEATURE SELECTION METHODS

The success of any classification algorithm highly depends on the input features. Feature selection is critical in reducing the classification error. Given a dataset with features X = {$x_j$; j = 1,..,N} and target $y$, feature selection obtains a subset of $S$ features from the N-dimensional space to distinguish $y$, boosting the interpretability and reliability of predictions, and reducing the time complexity.

#### 1) MAXIMUM RELEVANCE MINIMUM REDUNDANCY (mRMR)

Feature selection methods based on mutual information, F-test select the top features without considering the relationship among the selected features. They calculate the mutual information as a score between the joint distribution of all features $(x_i)$ and target $y$ and select the features with the largest score. However, the selected features might be correlated and not cover the whole space. mRMR penalizes a feature's relevancy using the mutual information score by its redundancy when other features are also present. It searches for features $S$ satisfying equation (6) to select the features with highest mutual information $I(x_i;y)$ to target variable $y$ and satisfying equation (7) to reduce the redundancy of the features selected using maximum relevance (equation (6)) [32].

$$maxD(S, y), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; y) \quad (6)$$

$$minR(S), R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \quad (7)$$

Here $I(x_i;y)$ and $I(x_i; x_j)$ are mutual information that determine the amount of difference between the joint distribution and product of marginal distributions of the pair of random variables involved.

#### 2) RANDOM FOREST FEATURE SELECTION

Random forest as a classifier performs implicit feature selection during training for classification, which results in higher accuracy. This implicit feature selection is utilised to rank a feature $x_i$ which adds the impurity decrease $\Delta i(\tau, T)$ for all nodes $\tau$ where $x_i$ is used and is averaged over all trees, T [33]. The feature importance is defined by equation (8).

$$Imp(x_i) = \frac{1}{T}\sum_{T}\sum_{\tau}\Delta i(\tau, T) \tag{8}$$

Here $i(\tau)$ is the 'gini impurity' at node $\tau$, expressed as:

$$i(\tau) = 1 - \sum_{i}^{c}(p_i|\tau)^2 \tag{9}$$

where $p_i$ is the fraction of samples that belongs to the $i$th class of the $c$ classes.

The input features for the six classifiers are obtained using the feature selection methods that consider time-domain and time-frequency domain features.

### C. FEATURES SELECTED

The composition of a signal can be analyzed by different time-domain statistics and frequency components. Time-domain analysis provides the transitory response of a system and allows a better understanding of the flow of electrical quantities. Wavelet transform is suitable for decomposing an aperiodic signal into frequency bands, and their time-frequency analysis has been used in several applications that require time and frequency information simultaneously: gait analysis, fault detection, ultra-wideband wireless communications, etc.

#### 1) WAVELET COEFFICIENT (WC)

Discrete Wavelet Transform (DWT) quantifies the similarity between the original signal and the wavelet function by the detail ($d_l$) and approximation ($a_l$) coefficients [34]. The low and high-frequency components are obtained at each decomposition level $l$ using equation (10) and equation (11).

$$a_l(k) = \sum_{l_k} w_\varphi(l_k - 2k)a_{l-1}(l_k) \tag{10}$$

$$d_l(k) = \sum_{l_k} w_\psi(l_k - 2k)a_{l-1}(l_k) \tag{11}$$

where $w_\varphi$, $w_\psi$ are the low and high pass filters. The mother wavelet and decomposition level influence the detail coefficients and thus the classification accuracy. However, researchers [15], [16], [18]–[20] have arbitrarily chosen the wavelet function and decomposition level without justifying their use. To address this issue, [35] used Particle Swarm Optimization to obtain the optimal wavelet functions combination to extract the most prominent features for classification of faults and [36] used harmony search algorithm to determine the suitable wavelet functions and decomposition levels.

Here multilevel 1D DWT is used with wavelet families 'Daubechies', 'Symlets', 'Coiflets', 'Biorthogonal', 'Reverse biorthogonal', and 'Discrete Meyer' to extract the WCs. The wavelet functions in each wavelet family ('Daubechies'- db1 to db38, 'Symlets'- sym2 to sym20, 'Coiflets'- coif1 to coif14, 'Biorthogonal'- bior1.1 to bior6.8, 'Reverse biorthogonal'- rbio1.1 to rbio6.8, 'Discrete Meyer'- dmey) are decomposed at different levels. The maximum useful level of decomposition chosen to avoid edge effects caused by signal extension is given by the equation (12).

$$Maximum\ level = \lfloor log_2(\frac{signal\ length}{filter\ length - 1})\rfloor \tag{12}$$

Features (wavelet functions + decomposition level) for DIF are chosen using a classifier-involved method. The detail coefficients of the 3-phase differential currents obtained from each of these wavelet functions at the permissible decomposition levels are used to train and test DT (the baseline classifier here), finding the one which minimizes the error rate. Five WCs with the best balanced accuracies averaged over 10 runs are selected. Thus, **bior2.2 at level 3, db4 at level 4, rbio3.3 at level 3, rbio4.4 at level 4, and sym4 at level 4** are obtained for DIF. The same features are used for LFU and CTD as well.

#### 2) DIFFERENTIAL WAVELET ENERGY (WE)

Wavelet energy is also a powerful tool to extract features. The differential WE is employed for differential protection of transformers in [37], [38]. The detail WC energy of the different wavelet functions that belong to the above mentioned wavelet families are combined to form a new set of inputs. The energy associated with the WCs for each wavelet function at all permissible levels considering one cycle post-transient 3-phase differential is calculated using

$$E_{dl}^{w} = \sum_{k}|d_l(k)|^2 \tag{13}$$

The top 10 WE features are then obtained using mRMR feature selection method, which finds the optimal feature subset considering the importance of the features and their correlations. An exhaustive search over $2^{10}$-1 combinations of the 10 features obtained with mRMR is performed using kNN and DT as the baseline classifiers to obtain the optimum number of features. It is noticed that the accuracy vs the number of features curve of both kNN and DT improved up to 6 features and then started decreasing as the number of features increases (Fig.6a). These 6 WE features, namely **rbio3.1, sym17, bior3.9, rbio3.9, coif13, and dmey** are thus selected and combined to form the inputs to the classifiers.

#### 3) TIME-DOMAIN (TD) FEATURES

The 3-phase differential currents are also used to extract a comprehensive number of TD features. The entire feature list consisting of 63 features can be obtained from [39]. Random forest feature selection method is used to rank these features in the order of information gain. Subsequently, the number

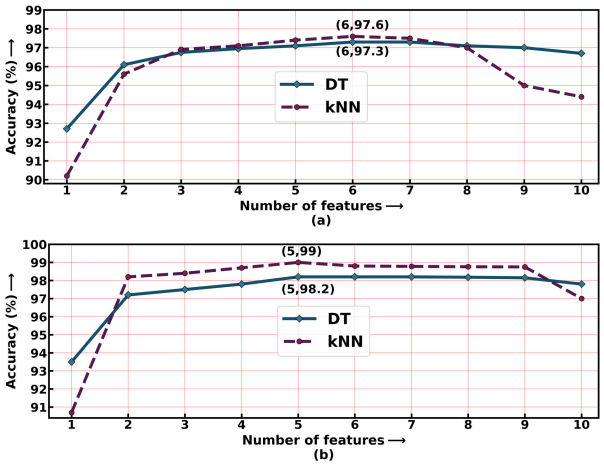


**FIGURE 6. Selecting optimal number of features: (a) Wavelet energy, and (b) Time-domain features.**

and combination of most relevant features are obtained by an exhaustive search over $2^{10}$-1 combinations of the top 10 features ranked by Random Forest feature selection using kNN and DT classifiers as the baseline again. It is observed that the accuracy vs the number of features curve of both kNN and DT improved up until 5 features and then started decreasing with any further increase in the number of features (Fig.6b). These 5 TD features, namely **average change quantile, sample entropy, excess kurtosis, variance, and complexity invariant distance** are detailed in the following part.

- *Average Change Quantile* $= \frac{1}{n'}\sum_{t=1}^{n'-1}|Id_{\phi_{t+1}} - Id_{\phi_t}|$, computes mean of absolute consecutive changes in the signal inside two values: *qh* and *ql* having $n'$ samples.
- *Sample Entropy* measures time complexity by computing the negative logarithm of the probability that subseries of length $m$ have distance $<$ r, then subseries of length m+1 also have distance $<$ r.
- *Excess Kurtosis* $= \frac{\mu 4}{\sigma^4} - 3$, is the fourth standardized moment with mean $\mu$ and standard deviation $\sigma$.
- *Variance* $= \frac{1}{n}\sum_{t=1}^{n}(Id_{\phi_t} - \mu)^2$ where $n$ is total samples.
- *Complexity Invariant Distance*$=\sqrt{\sum_{t=1}^{n-1}(Id_{\phi_{t+1}} - Id_{\phi_t})^2}$, estimates the time series complexity. A time series having more peaks, valleys etc. has a higher value.

Once the wavelet functions and the corresponding decomposition levels are obtained using the DT as baseline, the WCs are used to train and test RF, Xgb, NB, SVM, NN, and kNN classifiers. Similarly, the WE and TD features selected using the DT and kNN as baseline classifiers are used to train and test the six classifiers.

### D. CHOICE OF CLASSIFIERS

A variety of classifiers are used to evaluate the validity of the proposed feature-based protection scheme. Tree-based ML estimators: random forest (RF), and XGBoost (Xgb) having superior performance are very popular in data mining. The other classifiers used are Naive Bayes (NB)- a probabilistic classifier competitive in many applications; Support-vector machines (SVM)- basically a non-probabilistic classifier; Neural Networks (NN)- inspired by the human brain and adapted in a variety of applications ranging from social networking to cancer diagnosis; and k-nearest neighbors (kNN) where the system generalizes the training data after receiving a query.

#### 1) RANDOM FOREST

RFs are ensemble learning methods with reduced variance and better performance, which constructs numerous decision trees during learning and predicts the target, which is the mode of all predictions. Each tree constitutes a random subset of the training set and each node is split considering a subset of input features. While individual trees may overfit, averaging the predictions of all trees reduces the variance [33]. RF has also been used to select the important time-series features in Section III-C3.

#### 2) EXTREME GRADIENT BOOSTING (XGBoost)

Xgb is a supervised learning algorithm that sequentially combines weak learners into a stronger one, with each new model attempting to correct the previous model minimizing the objective function given by

$$J^{(t)} \approx \sum_{i=1}^{n}[a_i w_{c(x_i)} + \frac{1}{2}b_i w^2_{c(x_i)}] + \gamma L + \frac{1}{2}\lambda\sum_{i=1}^{n}[w_j^2] \quad (14)$$

where a(.) and b(.) are the first and second-order derivatives of mean square error loss, c(.) assigns data to the corresponding leaf, $w$ is score vector on leaves, $\gamma$ is complexity, $\lambda$ scales the penalty, and L is the number of leaves. The regularization term expressed as:

$$\Omega = \gamma L + \frac{1}{2}\lambda\sum_{i=1}^{n}[w_j^2] \quad (15)$$

present in the objective function is added as an improvement to reduce overfitting [40]. Xgb is one of the best gradient boosting machine frameworks and has become popular as the algorithm of choice for many winning teams of ML competitions.

#### 3) NAIVE BAYES

NB is the simplest Bayesian Network model that applies Bayes' Theorem to classify the target on the basis of conditional independence of every pair of features given the value of the class variable $y$ [41]. It is based on estimating P(A|B), the probability density of features A given class B. It has lesser training time and requires smaller training data. NB has shown good performance for applications such as text categorization, spam filtering, and medical diagnosis.

#### 4) SUPPORT VECTOR MACHINES

SVMs are memory-efficient classifiers that use the kernel method to create hyperplanes that separate the input data in high dimensional feature spaces [42]. The training samples

and the boundaries are called the support vectors and hyperplanes respectively. Generally, a larger distance between the hyperplane and the nearest training sample leads to a lower generalization error of the classifier. Radial Basis Function and polynomial kernels were used in the study.

#### 5) NEURAL NETWORK

The NN used is a fully connected feedforward network consisting of two hidden layers of perceptrons between the input and the output layer. It learns a non-linear function approximator $f(.) : R^S \rightarrow R^c$ with $S$ features and $c$ outputs through back-propagation [43]. It is an effective and efficient pattern recognition technique for ML applications.

#### 6) K-NEAREST NEIGHBOR

kNN is an instance-based non-parametric supervised learning used in applications of data mining, pattern recognition, and image processing, which computes the class of an instance by majority voting of the k (an integer) nearest neighbors of each query point. The training phase involves storing the features and target labels [44]. kNN has also been used as the baseline to select the optimum number of features in Section III-C2 and Section III-C3.

### E. BAYESIAN HYPERPARAMETER OPTIMIZATION

The performance of an ML algorithm depends on the choice of hyperparameters. Bayes' Theorem and Gaussian Process (GP) are used to optimize the hyperparameters of the classifiers used. Specifically, to get the optimal parameters for computationally intensive training of Xgb, which has numerous hyperparameters, the Bayesian Optimization has been used. It constructs a probabilistic surrogate of the objective function from the previous observations and then generates the next candidate of parameter list $z_{i+1}$ by optimizing the surrogate function. GP is used to model prior on $f$. The acquisition function $u$ proposes the next sampling points in the search space. The Bayesian Optimization with GP is described in Algorithm 1 [45].

**Algorithm 1** Bayesian Optimization

Collect initial observations $\mathcal{D}_n = \{z_i, f(z_i); i = 1, \ldots, n\}$.
**for** $n = 1, 2, \ldots$ **do**
Obtain the next sampling point $z_{n+1}$ by optimizing the acquisition function over the $GP$: $z_{n+1} = arg\ max_z\ u(z|\mathcal{D}_n)$.
Calculate $y_{n+1} = f(z_{n+1})$.
Augment observations $\mathcal{D}_{n+1} = \{\mathcal{D}_n, (z_{n+1}, y_{n+1})\}$ and update the $GP$.
**end of for**

## IV. RESULTS

The 3-phase differential currents acquired from CT1 and CT2 are sampled at a frequency of 10kHz. The features extracted and selected from the 167 post transient samples per phase and registered by the ED are used for training the six classifiers. The input dimension of the training and testing cases varies depending on the level of decomposition and wavelet function chosen when WCs are used as features. In the case of TD features, the input dimension is 15 ($5 \times 3$), and with WE as feature, it is 18 ($6 \times 3$). To reduce the classification error and improve the generalization, 10-fold stratified cross-validation and Bayesian search are applied, which use the available data effectively and train the classifiers on optimized hyperparameters. Normally, the performance of an estimator is evaluated with the accuracy metric. However, in the case of data imbalance between classes, the results are biased. Hence, balanced accuracy ($\bar{\eta}$) calculated using

$$\bar{\eta} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{16}$$

for a two-class problem is used to compute the performance. T, F, P, and N represent true, false, positive, and negative respectively in the equation, which gives the average accuracy obtained for all classes [46].

**TABLE 4.** Performance with WCs for DIF.

| *Wavelet* | *Classifier*($\bar{\eta}$) | | | | | |
|---|---|---|---|---|---|---|
| | RF | **Xgb** | NN | kNN | NB | SVM |
| bior2.2 | 99.5 | 99.7 | 99.7 | 99.4 | 71.0 | 90.2 |
| db4 | 99.5 | 99.7 | 99.5 | 99.4 | 77.2 | 93.0 |
| rbio3.3 | 99.6 | 99.7 | 99.5 | 99.1 | 76.5 | 93.0 |
| **rbio4.4** | 99.7 | **99.8** | 99.7 | 99.7 | 76.2 | 97.0 |
| sym4 | 99.7 | 99.7 | 99.6 | 99.5 | 77.7 | 93.7 |

### A. DETECTION OF INTERNAL FAULTS (DIF)

Since the occurrence of any power system transient event is unpredictable in time, the use of an ED becomes imperative. The correct distinction of internal faults from the other transients is the foremost classification task. The security and dependability of the proposed method depend on the type 1 error (FP) and type 2 error (FN) of this binary classification problem. The lower the classification error, the better is the performance of the entire scheme. To achieve this, the six classifiers are trained on 48442 cases and tested on 12110 cases of one-cycle of the post fault differential currents simulated in section II. The classifiers are trained with three sets of features, and the testing accuracies are reported. First, the selected WCs obtained using exhaustive search by training DTs are used as the inputs, and the classification performance is shown in Table 4. Xgb gives the best $\bar{\eta}$ of 99.8% on 'rbio4.4' at level 4. Second, the classifiers are trained on the 6 WE features obtained by an exhaustive search of $2^{10}$-1 different combinations of the top 10 WEs ranked using the mRMR algorithm. Table 7 shows the classification performance on the 6 features of the different classifiers. Xgb overshadows the rest of the classifiers with $\bar{\eta}$ of 99.5%. Third, the 5 features obtained again from an exhaustive search over $2^{10}$-1 different combinations of the 10 TD features ranked using RF are put-to-use. Table 8 shows the performance of the six classifiers. Again, Xgb gives the best performance with $\bar{\eta} = 99.8\%$.

**TABLE 5.** Performance with WCs for LFU.

| *Wavelet* | *Classifier($\bar{\eta}$)* | | | | | |
|---|---|---|---|---|---|---|
| | RF | **Xgb** | NN | kNN | NB | SVM |
| bior2.2 | 93.6 | 97.6 | 93.1 | 94.1 | 63.6 | 85.5 |
| db4 | 95.2 | 97.2 | 93.2 | 94.3 | 57.2 | 88.8 |
| rbio3.3 | 95.0 | 97.7 | 92.1 | 92.9 | 64.1 | 86.9 |
| **rbio4.4** | 95.5 | **97.8** | 92.9 | 94.4 | 57.7 | 89.7 |
| sym4 | 95.9 | 97.4 | 93.7 | 94.5 | 57.0 | 89.6 |

**TABLE 6.** Performance with WCs for CTD.

| *Wavelet* | *Classifier($\bar{\eta}$)* | | | | | |
|---|---|---|---|---|---|---|
| | RF | Xgb | **NN** | kNN | NB | SVM |
| bior2.2 | 98.6 | 98.8 | 99.2 | 99.2 | 74.1 | 96.5 |
| db4 | 98.0 | 98.7 | 98.8 | 97.8 | 66.7 | 96.2 |
| rbio3.3 | 98.6 | 98.8 | 99.3 | 99.3 | 73.1 | 98.1 |
| **rbio4.4** | 98.6 | 98.7 | **99.4** | 99.3 | 68.4 | 97.7 |
| sym4 | 98.2 | 98.9 | 98.7 | 98.9 | 67.4 | 97.3 |

**TABLE 7.** Performance with WE.

| *Model* | *Classifier($\bar{\eta}$)* | | | | | |
|---|---|---|---|---|---|---|
| | RF | **Xgb** | NN | KNN | NB | SVM |
| *DIF* | 93.2 | **99.5** | 86.0 | 99.2 | 78.4 | 60.0 |
| *LFU* | 93.2 | **98.3** | 82.4 | 94.0 | 57.3 | 72.5 |
| *CTD* | 95.6 | **98.7** | 96.1 | 98.7 | 62.4 | 88.8 |

**TABLE 8.** Performance with TD.

| *Model* | *Classifiers($\bar{\eta}$)* | | | | | |
|---|---|---|---|---|---|---|
| | RF | **Xgb** | NN | kNN | NB | SVM |
| *DIF* | 96.2 | **99.8** | 94.6 | 98.6 | 77.5 | 87.0. |
| *LFU* | 94.0 | **98.8** | 89.2 | 95.2 | 61.3 | 85.9 |
| *CTD* | 99.2 | **99.9** | 98.8 | 99.7 | 75.2 | 95.3 |

### B. LOCALIZATION OF FAULTY UNIT (LFU)

After the fault detector recognizes an internal fault, the faulty unit (exciting or series) is identified using the one-cycle data from the post fault differential currents. The six classifiers are trained on 37498 fault cases and tested on 9374 cases for LFU. Table 5 shows the classification performance on selected WCs as features, and Table 7 shows the same for WE features. Table 8 shows the classification performance on TD. Xgb performs better than the other classifiers with an $\bar{\eta}$ of 98.8% obtained using TD features, $\bar{\eta}$ of 97.8% with 'rbio4.4' at level 4, and $\bar{\eta}$ of 98.3% with WE as feature.

### C. CLASSIFICATION OF TRANSIENT DISTURBANCES (CTD)

The different transient disturbances: overexcitation, external faults with CT saturation, magnetizing and sympathetic inrush are also classified after the fault detector identifies them as no-fault transients. Table 6 shows the performance on selected WCs, table 7 on WE features, and table 8 shows the same for TD features of the six classifiers. 10944 cases are used for training and 2736 cases are used for testing the classifiers. Xgb outperforms the other classifiers with an $\bar{\eta}$ of 99.9% obtained with the TD features, $\bar{\eta}$ of 98.7% with WE as feature, and NN gives the best $\bar{\eta}$ of 99.4% with 'rbio4.4' at level 4.

**TABLE 9.** Computational time of the three models.

| *Model* | *Training time(s)* | | | *Testing time(ms)* | | | *Feature extract time(ms)* | | |
|---|---|---|---|---|---|---|---|---|---|
| | TD | WC | WE | TD | WC | WE | TD | WC | WE |
| *DIF* | 123 | 456 | 183 | 2.4 | 1.6 | 2.5 | 13.5 | 0.13 | 0.52 |
| *LFU* | 90 | 407 | 128 | 2.5 | 2.1 | 2.4 | 13.5 | 0.13 | 0.52 |
| *CTD* | 30 | 383 | 66 | 2.6 | 2.2 | 2.5 | 13.5 | 0.13 | 0.52 |

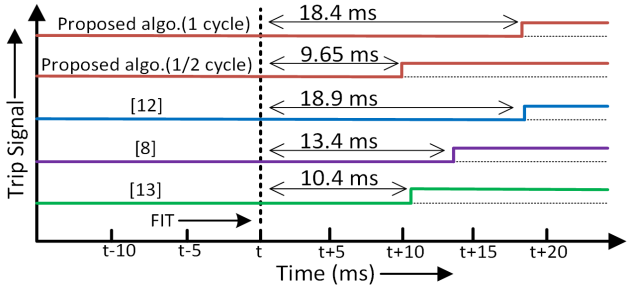


**FIGURE 7.** Comparison of operating time of different protection techniques.

### D. EXECUTION TIME

The proposed method can be faster than the operation time of 1-2 cycles of a conventional relay with harmonic blocking. The execution time (average time of 100 runs) for the feature extraction, training, and testing of the Xgb models for the three tasks with WC, WE, and TD as features are computed on Intel Core i7-8665U CPU @1.90 GHz, 16 GB RAM (See Table 9). The in-service operating time of the fault/no-fault decision would include time to extract the feature for a single instance and then testing it on the already trained Xgb model. Xgb trained on 'rbio4.4' is the fastest taking (16.67+1.6+0.13) = 18.4ms with a $\bar{\eta}$ of 99.8%. It takes 32.6ms with TD and 19.7ms with WE. To test the scheme for further reduction in computation time, the Xgb is trained and tested on 84 samples (1/2 cycle) of the 60552 cases. The results show that the proposed technique performs well with (8.34+1.2+0.12) = 9.65ms operating time and $\bar{\eta}$ of 99.2%. The time taken for LFU and CTD can be obtained from the columns 'Testing time' and 'Feature extract time' of the table. Noting that computations can be further optimized, these processing times are suitable for future real-time implementation.

*Comparison of Operating Time:* Fig.7 shows the operating time of the proposed technique on one cycle and 1/2 cycle, current differential-based techniques [12], and [13]; and directional-based technique [8]. The computational time of 9.65ms of the proposed scheme on 1/2 cycle suggests that ML-based differential protection schemes can compete with the previously proposed techniques [8], [12], [13].

## V. PERFORMANCE EVALUATION FOR NON-CONVENTIONAL AND ADDITIONAL CHALLENGES

The security and dependability of the proposed method are also tested for various system conditions in addition to the aforementioned traditional challenges in Section II. These conditions, namely the integration of type-4 wind turbine,

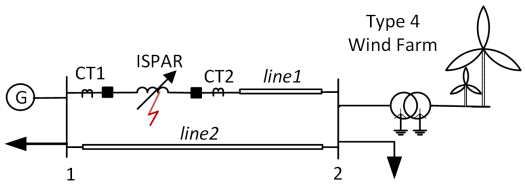


**FIGURE 8.** WTG connected at bus 2.

fault during magnetizing inrush, series winding saturation, change in tap positions, change in rating, saturation of CT, presence of noise, and low current faults which can jeopardize the reliability of the relay are discussed in this section.

### A. EFFECT OF INTEGRATION OF WTG

The type 4 Wind Turbine Generators (WTG) have complex fault characteristics and are very different from conventional generators. It is also expected that systems with high wind penetration may experience larger frequency deviations after system disturbances and in the absence of accurate modeling of its dynamics and fault behavior, the transformer differential protection may mal-operate [47]. A permanent magnet synchronous machine connected to the grid by a full-scale converter is considered in this study where the converter limits the fault current from 1.1 to 1.5 times the rated load current. The stability of the proposed scheme with the WTG is validated by the accuracy of 100% obtained on 5049 cases of internal faults and 6360 cases of transient disturbances. The fault cases are simulated by varying the tap positions, PS, FR, FIT, and FT (Fig.8). Due to grid side interface similarities, this analysis is also applicable to systems with photovoltaic generations [48].

### B. EFFECT OF SERIES WINDING SATURATION

Since the voltage rating of the series winding connecting the source and load is lower than the rating of the overall system, it may saturate when subjected to considerable voltage increase. The security of traditional differential protection is tested in such conditions [11]. The stability of the proposed scheme during series winding saturation is tested by increasing the source voltage from 120% to 150% of the nominal voltage in steps of 10%. 3000 cases of internal faults and 720 cases of series winding saturation are simulated by varying the tap positions, PS, and magnitude of overvoltage. Since the number of cases is imbalanced, Synthetic Minority Over-sampling Technique (SMOTE) [49] is used to oversample the series winding saturation cases. Xgb gives an accuracy of 100% on 3000 cases of each class.

### C. EFFECT OF CHANGE IN TAP POSITION

Generally, the transformer tap changer effect is taken into account with a corrected input of the primary voltage. The proposed technique considers different tap positions and tackles possible mal-operations in case of transients due to non-standard phase shifts without tracking the tap-changer positions. 3000 cases of internal faults and 648 cases of tap-change cases are simulated by varying the tap positions, PS, and ST. It gives an accuracy of 99.9% on 3000 cases of each obtained by oversampling the tap-change cases using SMOTE.

### D. EFFECT OF FAULT DURING INRUSH

The harmonic restraining or blocking differential relays are used to ensure security during magnetizing inrush; however, conventional relays' operation is delayed if faults occur during magnetic inrush. To ensure dependability, 12292 cases of inrush and faults during inrush are simulated by varying the parameters discussed in section II. $\bar{\eta}$ of 100% suggests that the proposed scheme performs well in the event of faults during magnetizing inrush.

### E. EFFECT OF CHANGE IN RATING

The PAR model, as well as the system parameters, influence the simulation and hence the training. Tolerances also exist in the transformer's, line's, and other parameters. The proposed scheme works even if an ISPAR of a different rating is considered. 6912 internal faults and other transients are simulated for an ISPAR with $S_n = 450$MVA, $V_n = 345$kV by varying FR, PAR tap position, FT, FIT, ST, etc. The same Xgb-1 model, which was trained on $S_n = 500$MVA, $V_n = 230$kV, is used to test these 6912 cases. The stability of the scheme is substantiated by $\bar{\eta}$ of 99.3%.

### F. EFFECT OF CT SATURATION

The impedance of CT secondary may influence the level of harmonics in the differential currents. To study the effect of saturation of the CTs, the burden and CT secondary impedance are changed. $\bar{\eta}$ of 100% on 6912 cases of internal faults and other transient disturbances obtained by varying the different parameters discussed in section II validate the reliability of the proposed scheme for CT saturation conditions.

### G. EFFECT OF NOISE

In the real-world presence of noise during the capture and processing of differential currents may affect the protection system's stability. The white Gaussian noise of different Signal-to-Noise-ratio (SNR) is added to the data to study its effect on the proposed method. Table 10 shows the accuracy of Xgb for noise levels from 40dB to 20dB. The classifier performs poorly with lower SNR, but even then always above 80.2% ($\frac{93.4+67.8}{2}$). The $\bar{\eta}$ varies from 99.8% with no noise to 80.2% for SNR of 20dB. The accuracy dips down further to 67% for a SNR of 10dB which is understandable as the ratio of the desired information to the undesired signal is only about 3.16.

### H. LOW CURRENT t-t & WINDING PH-G FAULTS

The proposed algorithm performs well for both high resistive winding ph-g faults and *t-t* faults also. To test its sensitivity, *t-t* faults with 2% of the series winding shorted, and winding ph-g faults with high resistance of 50Ω in the series winding are simulated. The ED was able to detect the 48 winding ph-g

**TABLE 10. Effect of noise.**

| Fault/ Disturbances | SNR (dB) | Number of cases | Predicted class | | Accuracy (%) |
|---|---|---|---|---|---|
| | | | Faults | Disturbances | |
| Internal faults | $\infty$ | 9406 | 9399 | 7 | 99.9 |
| | 40 | 9406 | 9324 | 82 | 99.1 |
| | 30 | 9431 | 9246 | 185 | 98.0 |
| | 20 | 9318 | 8700 | 618 | 93.4 |
| Other disturbances | $\infty$ | 2705 | 13 | 2692 | 99.5 |
| | 40 | 2705 | 96 | 2609 | 96.5 |
| | 30 | 2680 | 337 | 2343 | 87.4 |
| | 20 | 2793 | 898 | 1895 | 67.8 |

and 144 *t-t* faults obtained by varying the tap positions, FR, and FIT.

## VI. CONCLUSION

This article addresses the problem of detection and localization of faults and classification of transients for an ISPAR. The internal faults are distinguished from overexcitation, external faults with CT saturation, and inrush conditions. After that, depending on the detection of a fault, the faulty unit (ISPAR series or exciting) is located, or the transient disturbances are classified. Wavelet and time-domain features obtained from one cycle of post transient 3-phase differential currents registered by the event detector are used to train six prominent classifiers. Firstly, the classifiers are trained with the most important WCs obtained by exhaustive search using DT. Secondly, the top WE features obtained using mRMR are put-to-use. Lastly, TD features selected by maximizing the information gain are used. It is observed that overall XGBoost trained with the TD features outperforms the other models for DIF, LFU, and CTD; and when both accuracy and computation time are considered the XGBoost model trained on 'rbio4.4' WC is superior for DIF. On top of fault detection with $\bar{\eta} = 99.8\%$, localization with $\bar{\eta} = 98.8\%$, and classification of transients with $\bar{\eta} = 99.9\%$, the proposed scheme has several benefits over the conventional differential relays:

- the proposed scheme is more dependable for fault during magnetic inrush and sensitive to low current turn-to-turn and winding ph-g faults;
- it is secure to magnetic and sympathetic inrush, overexcitation, external fault with CT saturation, series winding saturation, CT saturation, tap position changes, and integration of WTG;
- it takes care of the non-standard phase shift in the PAR without tracking the exciting unit tap positions;
- the proposed technique is robust to change in PAR ratings and noise in measurements;
- it does not need additional voltage or phase information.

The protection scheme advanced in this article can integrate with standard microprocessor-based differential relays and offer supervisory control over their operation, thus improving the stability of the power system and providing a complete solution to the problem of PAR protection.

## APPENDIX

Fortran code for single-phase 3-winding transformer fault model

1. nw = 6, w = 2*pi*f
2. $I_m3 = I_m2 = I_m1$
3. fa = fault1*0.01
4. fb = 1.0-fa
5. fc = fault2*0.01
6. fd = 1.0-fc
7. fe = fault1*0.01
8. ff = 1.0-fe
9. z1=v1/i1, z2=v2/i2, z3=v3/i3
10. l1 = v1/(w*$I_m$1*i1)
11. l2 = v2/(w*$I_m$2*i2)
12. l3 = v3/(w*$I_m$3*i3)
13. X1 = (x13-x23+x12)/2
14. X2 = (x23-x13+x12)/2
15. X3 = (x13-x12+x23)/2
16. Lk1 = X1*z1/w
17. Lk2 = X2*z2/w
18. Lk3 = X3*z3/w
19. L1l= Lk1*$fa$, L2l=Lk1*$fb$
20. L3l= Lk2*$fc$, L4l=Lk2*$fd$
21. L5l= Lk3*$fe$, L6l=Lk3*$ff$
22. L1m =v1/(w*$I_m$1*i1)*$fa^2$
23. L2m =v1/(w*$I_m$1*i1)*$fb^2$
24. L3m = v2/(w*$I_m$2*i2)*$fc^2$
25. L4m = v2/(w*$I_m$2*i2)*$fd^2$
26. L5m =v3/(w*$I_m$3*i3)*$fe^2$
27. L6m =v3/(w*$I_m$3*i3)*$ff^2$
28. L1 = L1l + L1m, L2 = L2l + L2m
29. L3 = L3l + L3m, L4 = L4l + L4m
30. L5 = L5l + L5m, L6 = L6l + L6m
31. M12 = sqrt(L1m*L2m)
32. M13 = sqrt(L1m*L3m)
33. M14 = sqrt(L1m*L4m)
34. M15 = sqrt(L1m*L5m)
35. M16 = sqrt(L1m*L6m)
36. M23 = sqrt(L2m*L3m)
37. M24 = sqrt(L2m*L4m)
38. M25 = sqrt(L2m*L5m)
39. M26 = sqrt(L2m*L6m)
40. M34 = sqrt(L3m*L4m)
41. M35 = sqrt(L3m*L5m)
42. M36 = sqrt(L3m*L6m)
43. M45 = sqrt(L4m*L5m)
44. M46 = sqrt(L4m*L6m)
45. M56 = sqrt(L5m*L6m)

## REFERENCES

[1] *IEEE Draft Guide for Protecting Power Transformers*, Standard IEEE PC37.91/D16, Oct. 2019, pp. 1–163.

[2] X.-n. Lin, P. Liu, and O. P. Malik, "Studies for identification of the inrush based on improved correlation algorithm," *IEEE Trans. Power Del.*, vol. 17, no. 4, pp. 901–907, Oct. 2002.

[3] R. Hamilton, "Analysis of transformer inrush current and comparison of harmonic restraint methods in transformer protection," *IEEE Trans. Ind. Appl.*, vol. 49, no. 4, pp. 1890–1899, Jul. 2013.

[4] T. S. Sidhu, M. S. Sachdev, H. C. Wood, and M. Nagpal, "Design, implementation and testing of a microprocessor-based high-speed relay for detecting transformer winding faults," *IEEE Trans. Power Del.*, vol. 7, no. 1, pp. 108–117, Jan. 1992.

[5] P. Liu, O. P. Malik, D. Chen, G. S. Hope, and Y. Guo, "Improved operation of differential protection of power transformers for internal faults," *IEEE Trans. Power Del.*, vol. 7, no. 4, pp. 1912–1919, Oct. 1992.

[6] M. Stanbury and Z. Djekic, "The impact of current-transformer saturation on transformer differential protection," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1278–1287, Jun. 2015.

[7] M. A. Ibrahim and F. P. Stacom, "Phase angle regulating transformer protection," *IEEE Trans. Power Del.*, vol. 9, no. 1, pp. 394–404, Jan. 1994.

[8] U. N. Khan and T. S. Sidhu, "A phase-shifting transformer protection technique based on directional comparison approach," *IEEE Trans. Power Del.*, vol. 29, no. 5, pp. 2315–2323, Oct. 2014.

[9] *IEEE Guide for the Application of Protective Relaying for Phase-Shifting Transformers*, IEEE Standard C37.245-2018, May 2019, pp. 1–71.

[10] U. N. Khan and T. S. Sidhu, "Protection of standard-delta phase shifting transformer using terminal currents and voltages," *Electr. Power Syst. Res.*, vol. 110, pp. 31–38, May 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779614000030

[11] M. Ibrahim, *Protection of Phase Angle Regulating Transformers*, Standard IEEE Special Publ. PSRC WG K1, 1999.

[12] Z. Gajic, "Use of standard 87T differential protection for special three-phase power transformers—Part I: Theory," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1035–1040, Jul. 2012.

[13] B. Kasztenny and E. Rosolowski, "Modeling and protection of hexagonal phase-shifting transformers—Part II: Protection," *IEEE Trans. Power Del.*, vol. 23, no. 3, pp. 1351–1358, Jul. 2008.

[14] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep neural network into power transformer differential protection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 865–876, Feb. 2020.

[15] P. L. Mao and R. K. Aggarwal, "A novel approach to the classification of the transient phenomena in power transformers using combined wavelet transform and neural network," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 654–660, Oct. 2001.
[16] S. Bagheri, Z. Moravej, and G. B. Gharehpetian, "Classification and discrimination among winding mechanical defects, internal and external electrical faults, and inrush current of transformer," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 484–493, Feb. 2018.
[17] Z. Jiao and Z. Li, "Novel magnetization hysteresis-based power-transformer protection algorithm," *IEEE Trans. Power Del.*, vol. 33, no. 5, pp. 2562–2570, Oct. 2018.
[18] A. M. Shah and B. R. Bhalja, "Discrimination between internal faults and other disturbances in transformer using the support vector machine-based protection scheme," *IEEE Trans. Power Del.*, vol. 28, no. 3, pp. 1508–1515, Jul. 2013.
[19] Y. Sheng and S. M. Rovnyak, "Decision trees and wavelet analysis for power transformer protection," *IEEE Trans. Power Del.*, vol. 17, no. 2, pp. 429–433, Apr. 2002.
[20] P. B. Thote, M. B. Daigavane, P. M. Daigavane, and S. P. Gawande, "An intelligent hybrid approach using KNN-GA to enhance the performance of digital protection transformer scheme," *Can. J. Electr. Comput. Eng.*, vol. 40, no. 3, pp. 151–161, 2017.
[21] A. M. Shah and B. R. Bhalja, "Fault discrimination scheme for power transformer using random forest technique," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 6, pp. 1431–1439, Apr. 2016.
[22] S. K. Bhasker, P. K. Bera, V. Kumar, and M. Tripathy, "Differential protection of indirect symmetrical phase shift transformer and internal faults classification using wavelet and ANN," in *Proc. IEEE Region Conf. (TENCON)*, Nov. 2015, pp. 1–6.
[23] P. K. Bera, C. Isik, and V. Kumar, "Discrimination of internal faults and other transients in an interconnected system with power transformers and phase angle regulators," *IEEE Syst. J.*, early access, Jul. 29, 2020, doi: 10.1109/JSYST.2020.3009203.
[24] S. Kulkarni and S. Khaparde, *Transformer Engineering*. Boca Raton, FL, USA: CRC Press, 2013.
[25] A. Zollanvari, K. Kunanbayev, S. A. Bitaghsir, and M. Bagheri, "Transformer fault prognosis using deep recurrent neural network over vibration signals," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, Sep. 2021, Art. no. 2502011.
[26] C.-H. Einvall and J. R. Linders, "A three-phase differential relay for transformer protection," *IEEE Trans. Power App. Syst.*, vol. PAS-94, no. 6, pp. 1971–1980, Nov. 1975.
[27] T. Zheng, T. Huang, Y. Ma, Z. Zhang, and L. Liu, "Histogram-based method to avoid maloperation of transformer differential protection due to current-transformer saturation under external faults," *IEEE Trans. Power Del.*, vol. 33, no. 2, pp. 610–619, Apr. 2018.
[28] P. Pachore, Y. Gupta, S. Anand, S. Sarkar, P. Mathur, and P. K. Singh, "Flux error function based controlled switching method for minimizing inrush current in 3-phase transformer," *IEEE Trans. Power Del.*, vol. 36, no. 2, pp. 870–879, Apr. 2021.
[29] A. Q. Zhang, T. Y. Ji, M. S. Li, Q. H. Wu, and L. L. Zhang, "An identification method based on mathematical morphology for sympathetic inrush," *IEEE Trans. Power Del.*, vol. 33, no. 1, pp. 12–21, Feb. 2018.
[30] P. K. Bera and C. Isik. (2020). *Data: Transients in Indirect Symmetrical Phase Shift Transformers*. [Online]. Available: https://dx.doi.org/10.21227/d8fv-6257
[31] O. Dharmapandit, R. K. Patnaik, and P. K. Dash, "A fast time-frequency response based differential spectral energy protection of AC microgrids including fault location," *Protection Control Modern Power Syst.*, vol. 2, no. 1, p. 30, Aug. 2017.
[32] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[33] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[34] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[35] T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu, "A new approach for fault classification in microgrids using optimal wavelet functions matching pursuit," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4838–4846, Sep. 2018.
[36] T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu, "A new harmony search approach for optimal wavelets applied to fault classification," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 521–529, Mar. 2018.
[37] R. P. Medeiros and F. B. Costa, "A wavelet-based transformer differential protection with differential current transformer saturation and cross-country fault detection," *IEEE Trans. Power Del.*, vol. 33, no. 2, pp. 789–799, Apr. 2018.
[38] Y. Deng, S. Lin, L. Fu, K. Liao, L. Liu, Z. He, S. Gao, and Y. Liu, "New criterion of converter transformer differential protection based on wavelet energy entropy," *IEEE Trans. Power Del.*, vol. 34, no. 3, pp. 980–990, Jun. 2019.
[39] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time series FeatuRe extraction on basis of scalable hypothesis tests (tsfresh—A Python package)," *Neurocomputing*, vol. 307, pp. 72–77, Sep. 2018.
[40] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.
[41] T. M. Mitchell, *Machine Learning*, 1st ed. New York, NY, USA: McGraw-Hill, 1997.
[42] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer-Verlag, 1995.
[43] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
[44] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 2006.
[45] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, New York, NY, USA, 2012, pp. 2951–2959.
[46] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3121–3124.
[47] H. Nguyen-Duc and Y. Nakanishi, "Effect of DFIG wind farm fault currents on the transformer differential relaying performance," in *Proc. IEEE Innov. Smart Grid Technol. Asia (ISGT Asia)*, May 2018, pp. 916–921.
[48] A. Hooshyar, E. F. El-Saadany, and M. Sanaye-Pasand, "Fault type classification in microgrids including photovoltaic DGs," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2218–2229, Sep. 2016.
[49] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

**PALLAV KUMAR BERA** (Student Member, IEEE) received the B.S. degree in electrical engineering from the Haldia Institute of Technology, West Bengal, India, in 2011, and the M.S. degree in systems and control from the IIT Roorkee, India, in 2014. He is currently pursuing the Ph.D. degree with Syracuse University, Syracuse, NY, USA. His research interests include the protection of electric power grids and transient and small signal stability analysis.

**CAN ISIK** (Life Senior Member, IEEE) received the B.S. and M.S. degrees from Middle East Technical University, Ankara, Turkey, in 1978 and 1980, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1985. He is currently a Professor with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. His research interests include intelligent systems applications, indoor environment control, modeling and forecasting for financial systems, and medical instrumentation. He is a Member of Eta Kappa Nu, Golden Key, and is listed in Who is Who in American Education.