

A Power Transformer Differential Protection Based on Support Vector Machine and Wavelet Transform

Lucas D. Simões, Hagi J. D. Costa, Matheus N. O. Aires,
Rodrigo P. Medeiros, Flavio B. Costa, Arturo S. Bretas

***Abstract*—This paper presents a power transformer differential protection scheme based on support vector machines (SVM) combined with high-frequency features extracted with the real-time boundary stationary wavelet transform (RT-BSWT). SVM models are derived with synthetic data, considering a wide variety of events, such as inter-turn faults, external faults during CT saturation, and evolving external-to-internal faults. A comparative performance assessment is carried out considering accuracy and other reliability indices, as well as operating time, and good results were achieved. The simplicity of the presented SVM-based relay, without hard-to-derive parameters, built on the classical differential protection framework, highlights potential aspects towards real-life implementation.**

***Keywords*—Support vector machines, power transformers, differential protection, CT saturation, wavelet transform**

I. Introduction

DIFFERENTIAL protection has been commonly employed as the primary protection for power transformers, which are vital components to power system operation and control. Therefore, accurate event detection and fast fault clearance are of utmost importance. However, even though being able to correctly distinguish external from internal faults to the protection zone, delimited by the current transformers (CT), traditional differential protection might be unable to discriminate internal faults from inrush currents, which arise as an effect of power transformer energizations [1].

Harmonic restraint and blocking algorithms have been commercially available to enhance traditional phasor-based differential protection. Nonetheless, these techniques might fail to operate when inrush currents present low harmonic content on one or two phases [2]. Furthermore, these algorithms also present an inherent delay due to the phasor estimation.

Current transformer saturation also plays a significant role in affecting transformer differential protection reliability. Due to its unpredictable nature, CT saturation might incur in misread differential currents that cause the relay to malfunction [3]. Also, evolving external-to-internal faults, i.e., faults that occur during another fault in the same circuit, possibly involving different phases [4], can cause unexpected behavior during CT saturation and are problems to be dealt with by the protection scheme. Furthermore, CIGRE technical brochure [5] presents that most of the significant failures in power transformers have as main contributors failures in the windings.

Many modern digital signal processing techniques and data-driven-based protection schemes have been proposed in the literature in order to overcome the aforementioned problems and limitations [6]–[12]. For instance, [13] proposed the real-time boundary stationary wavelet transform (RT-BSWT) to detect faults and other disturbances. However, this method can only perform the detection, with no possibility to identify a fault. Based on the wavelet-based signal processing in [13], [6]–[8] proposed the traditional phase differential function (87T) and the negative sequence differential function (87Q) to protect power transformers using high-frequency information obtained with the RT-BSWT instead of using low-frequency information obtained with Fourier-based phasor estimation. Therefore, [6]–[8] developed a way to identify internal faults from other disturbances. Despite the RT-BSWT signal processing being strongly sensible to detect faults in [6]–[8], the protection based on thresholds applied in these wavelet signals is susceptible to noise and does not provide an event classification. Also, this method presented a delay of more than a half-cycle for accurate external-to-internal evolving fault detection. In [9] and [10], deep learning-based power transformer differential protection methods were proposed, with good results reported. However, deep learning algorithms are not so easily embedded in hardware, thus potentially halting real-life implementations. [11] and [12] presented interesting approaches for power transformer differential protection based on random forest and support vector machine (SVM) algorithms, respectively, and likewise, promising results were reported. However, these last two works employed a cycle and a half-cycle vector lengths, respectively, to be used as input features to their data-driven models, making their methods slower or similar to phasor-based methods when comparing relay operating time.

This work presents an SVM-based power transformer differential protection scheme. The proposed protection is powered with the operating and restraining wavelet coefficients energies, which are obtained in accordance with [8]. If one of these differential energies exceed a respective threshold, a disturbance detector enables a first SVM classifier in order to

This work was supported by CAPES (Coordination for the Improvement of Higher Education Personnel) and CNPq (Brazilian National Council for Scientific and Technological Development).

Lucas D. Simões, Hagi J. D. Costa, Matheus N. O. Aires and Flavio B. Costa are with the Federal University of Rio Grande do Norte (UFRN), 59078-970, Natal - RN - Brazil (e-mails: lucassimoesee@gmail.com, hagijakobson@gmail.com, matheus2726@gmail.com, flaviocosta@ect.ufrn.br).

Rodrigo P. Medeiros is with the Federal Rural University of the Semi Arid (UFERSA), 59780-000, Caraúbas - RN - Brazil (e-mail: rodrigo.prado@ufersa.edu.br).

Arturo S. Bretas is with University of Florida, Gainesville, FL 32611, USA (email: arturo@ece.ufl.edu).

classify the type of event as an external fault, internal fault or transformer energization. If the event is an internal fault, a tripping command is emitted; if the event is an external fault, a second SVM classifier is enabled to distinguish between possible CT saturations or external-to-internal evolving faults.

The SVM models were trained using the LIBSVM [14] with synthetic data of fault records of an electric power system simulated in MATLAB/Simulink environment. A wide variety of events were simulated, such as: internal turn-to-turn and turn-to-earth faults, external faults with CT saturation, as well as evolving external-to-internal faults.

A comparative performance assessment with conventional methods, including the phase differential function 87T with harmonic restraint and harmonic blocking units, the negative sequence differential function 87Q, and the restricted earth fault (REF) function, is performed, regarding reliability indices and operating time, and the presented protection scheme has shown to be faster and more reliable than conventional methods, and also faster than other methods presented in the literature. In addition, the proposed method was implemented in hardware to show its feasibility in practical applications.

II. SVM Theory

The support vector machine is a binary supervised learning algorithm based on the statistical learning theory [15]. Its main goal is, upon training, to construct a decision surface, called hyperplane, in a fashion to maximize the margin between positive and negative samples [15]. This hyperplane is delimited by the support vectors, which are the training examples closest to the optimal hyperplane. With the help of the support vectors, by solving an optimization problem, it is possible to set up a maximized separation margin.

The SVM performs separation of nonlinearly separable data by adopting the kernel functions, which map the data into higher-dimensional spaces, where it is possible to perform linear pattern separation. In the case of using non-linearly separable data, the trained SVM classifier can be written as:

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i y_s^{(i)} K(\mathbf{x}_s^{(i)}, \mathbf{x}) + b, \quad (1)$$

where l is the number of support vectors; $\mathbf{x}_s^{(i)}$, $y_s^{(i)}$ and α_i are the i-th support vector, its associated label, and weight, respectively; $\mathbf{x}$ is the data observation to be classified; $K(\mathbf{x}_s^{(i)}, \mathbf{x})$ is the kernel function; and b is a bias [16]. The decision function, for a given unclassified feature vector $\mathbf{x}$, uses the sign of the classifier in (1) to perform data classification.

One key point of SVM algorithms is their offline parameterization, which requires expert knowledge, since different combinations of parameters might yield different results for a given dataset, making it essential to find the best parameters. This is typically solved by employing a grid search, which consists of testing a different range of hyperparameters and evaluating their performance. The concerning hyperparameters may vary depending on the chosen kernel, yet, one that is always present is the regularization parameter C. For instance, the choice of the radial-basis function (RBF) as kernel function leads to an additional parameter to be derived in the grid search, called γ. Moreover, since the SVM is essentially a binary classifier, some techniques are needed so the SVM algorithms can deal with problems with more than two classes. Among these methods, some of the most popular are the *one-versus-all* and the *one-versus-one* approaches [17]. In this paper, for the problems regarding more than two classes, the one-versus-all approach was adopted.

III. SVM-based Differential Protection Relay

Fig. 1 depicts the flowchart of the proposed differential protection. This protection algorithm is designed to run at every sampling time k. Details about each of the processing blocks presented in Fig. 1 are addressed in this section.

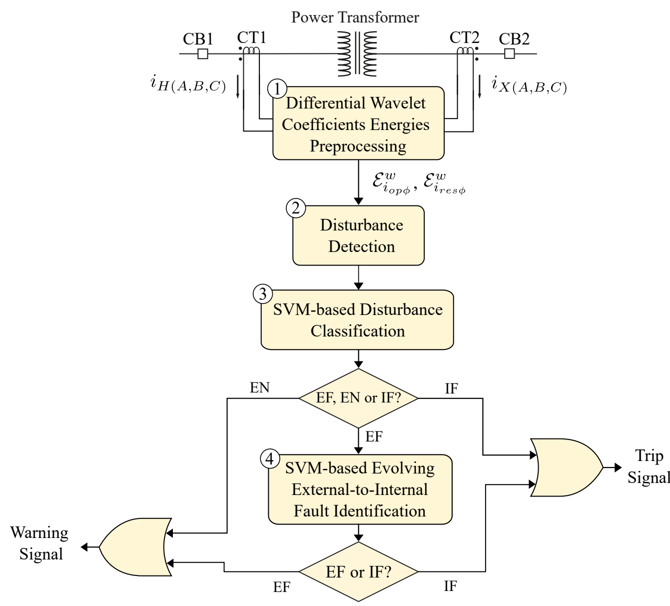

Fig. 1. Proposed SVM-based power transformer differential protection.

A. Differential Wavelet Coefficients Energies Preprocessing (Block 1)

Fig. 1 illustrates the data processing of the presented SVM-based protection algorithm. The data processing comprises several stages: currents measurement, performing of RT-BSWT, phase and magnitude adjustments, and computation of differential wavelet coefficients and differential energies. Data processing is needed to obtain the differential boundary wavelet coefficients energies, which were initially introduced by [6], and will be used as inputs for the subsequent algorithm analytics. The steps are described as follows:

1) Current Measurement: Initially, the relay acquires current samples from the CT secondary currents employing anti-aliasing filters and AD converters. This procedure yields the time-discrete secondary currents $i_{H\phi} = \{i_{HA}, i_{HB}, i_{HC}\}$ and $i_{X\phi} = \{i_{XA}, i_{XB}, i_{XC}\}$, where H and X stand for the transformer primary and secondary windings, respectively, whereas ϕ corresponds to phases A, B, C, and also negative sequence Q, which is computed from previously-stored current samples as follows [6]:

$$i_{HQ}(k) = \frac{1}{3}\left[i_{HA}(k) + i_{HB}\left(k - \frac{f_s}{3f}\right) + i_{HC}\left(k - \frac{2f_s}{3f}\right)\right], \tag{2}$$

$$i_{XQ}(k) = \frac{1}{3}\left[i_{XA}(k) + i_{XB}\left(k - \frac{f_s}{3f}\right) + i_{XC}\left(k - \frac{2f_s}{3f}\right)\right], \tag{3}$$

where k is the current sampling; $\frac{f_s}{3f}$ and $\frac{2f_s}{3f} \in \mathbb{N}$ correspond to delays of 120ž and 240ž, respectively; f_s is the sampling frequency; f is the fundamental frequency. Although it uses delayed samples, this computation also includes actual current samples from phase A. Therefore, in the occurrence of a fault, the negative sequence has relevant information of this fault already in the first post-fault samples. The instantaneous values of the currents are used to perform the RT-BSWT.

2) RT-BSWT: The differential protection algorithm aims to distinguish events in a power transformer by first extracting high-frequency components of fault-induced transients. This extraction is performed using the RT-BSWT, which is a time-frequency decomposition method, where a discrete signal is decomposed into scaling and/or wavelet coefficients, obtained by using low-pass (scaling filter, s) and high-pass (wavelet filter, h) filters, respectively. Since the objective in this paper is to extract high-frequency information, only the wavelet decomposition is accounted for. The RT-BSWT wavelet components are computed through the inner product of the wavelet filter h of L coefficients, and with L samples of the current signal i as follows [13]:

$$w(l,k) = \frac{1}{\sqrt{2}} \sum_{n=0}^{L-1} h(n) i(k - L + n + 1 + l), \tag{4}$$

where $0 \leq l < L; \Delta k \geq L$ is the length of a sliding window; L is the number of wavelet coefficients computed at each sampling time; $k \geq \Delta k - 1$ is the current sampling; $i(k+m) = i(k - \Delta k + m)$ with $m \in \mathbb{N}$ (periodized signal in Δk samples).

3) Phase/Magnitude Adjustments: The magnitude, phase, and zero-sequence adjustments, needed due to the CTs configuration, are performed on the wavelet coefficients according to [7].

4) Differential Wavelet Coefficients: The differential wavelet coefficients are given by [8]:

$$w_{i_{op\phi}}(0,k) = \frac{1}{2}(w'_{i_{H\phi}}(0,k) + w'_{i_{X\phi}}(0,k)), \tag{5}$$

$$w_{i_{op\phi}}(l \neq 0,k) = w'_{i_{H\phi}}(l,k) + w'_{i_{X\phi}}(l,k), \tag{6}$$

$$w_{i_{res\phi}}(l,k) = w'_{i_{H\phi}}(l,k) - w'_{i_{X\phi}}(l,k), \tag{7}$$

where $0 \leq l < L$ and $w'_i = \{w'_{i_{H\phi}}, w'_{i_{X\phi}}\}$ are the wavelet coefficients computed with (4).

5) Differential Energies: The differential boundary wavelet coefficients energies $\mathcal{E}^w_{diff} = \{\mathcal{E}^w_{i_{op\phi}}, \mathcal{E}^w_{i_{res\phi}}\}$ are calculated from the respective differential wavelet coefficients $w_{diff} = \{w_{i_{op\phi}}, w_{i_{res\phi}}\}$, as follows [8]:

$$\mathcal{E}^w_{diff}(k) = \mathcal{E}^{wa}_{diff}(k) + \mathcal{E}^{wb}_{diff}(k), \tag{8}$$

in which the terms $\mathcal{E}^{wa}_{diff}$ and $\mathcal{E}^{wb}_{diff}$ are computed as [13]:

$$\mathcal{E}^{wa}_{diff}(k) = \sum_{l=1}^{L-1} w^2_{diff}(l,k), \tag{9}$$

$$\mathcal{E}^{wb}_{diff}(k) = \sum_{n=k-\Delta k+L}^{k} w^2_{diff}(0,n). \tag{10}$$

B. Disturbance Detection (Block 2)

Any transient event, such as transformer energizations, external or internal faults, can be detected when [6], [13]:

$$\begin{cases} \mathcal{E}^w_{diff}(k-1) \leq E_{diff}, \\ \mathcal{E}^w_{diff}(k) > E_{diff}. \end{cases} \tag{11}$$

When both inequalities in (11) are valid, the current sampling k is recorded as $k_d = k$, where k_d is the sample in which a disturbance is detected; $E_{diff} = \{E_{op\phi}, E_{res\phi}\}$ are the adaptive energy thresholds, estimated by [6]:

$$E_{diff} = \frac{3}{k_2 - k_1 + 1} \sum_{n=k_1}^{k_2} \mathcal{E}^w_{diff}(n), \tag{12}$$

where $[k_1/f_s \; k_2/f_s]$ is a time range of two cycles; k_2 and k_1 are last and first samples of the adaptive threshold sliding window.

C. SVM-based Disturbance Classification (Block 3)

Immediately after a disturbance is detected by block 2, the SVM-based disturbance classification block is enabled and loaded with the differential wavelet coefficients energy samples. For convenience, this block will hereafter be referred to as SVM 1. After the disturbance detection, the block accumulates four post-disturbance samples from the differential energies $\mathcal{E}^w_{diff}$, to form a signature vector composed of 32 features (4 post-disturbance samples for operating and restraining energies, from phases A, B, C, and negative sequence Q). The post-fault window with four differential energies samples is illustrated in Fig. 2(b). This vector is then normalized with minimum and maximum values for each feature obtained during the training stage, discussed in detail in Section IV-A.

The now normalized feature vector is ready to be used as input for the SVM model, which classifies the disturbance as an external fault (EF), internal fault (IF), or energization (EN). If an energization is classified, a warning signal is generated; if an internal fault is detected, a trip signal is issued to open the circuit breakers; if an external fault is identified, the evolving external-to-internal identification block is enabled.

Fig. 2 illustrates an internal turn-to-turn fault, being detected at k_d, and thus enabling SVM 1. After accumulating the first four post-disturbance differential energies samples, at sample k_d+3, it identifies the event as an internal fault, issuing a trip signal, as shown in Fig. 2(a). However, the fault is only cleared after the circuit-breaker opening time delay. In this paper, in the simulation environment, the adopted circuit-breaker time delay is approximately two cycles. Therefore, two cycles after the trip signal is issued, and when the currents go to zero, the circuit breakers open to clear the fault. The opening of the breakers results in transients, as shown in Fig. 2(b), which could make the protection algorithm detect a disturbance and perform a false classification of event. In order to avoid this problem, the SVM 1 classification is temporarily blocked after the trip signal is issued.

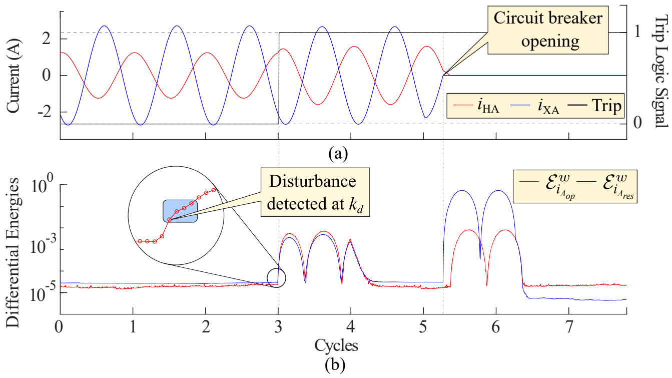

Fig. 2. Turn-to-turn internal fault involving 2% of the wye winding of phase A: (a) high and low sides currents, i_{HA} and i_{XA}, respectively, as well as the trip logic signal; (b) operating and restraining wavelet coefficients energies of phase A

D. SVM-based evolving external-to-internal Fault Identification (Block 4)

For the sake of clarity, likewise previously stated for SVM 1, this processing block will be referred to as SVM 2. Accordingly, following the flowchart depicted in Fig. 1, in the occurrence of an external fault classification by SVM 1, this block is enabled.

After the fault detection, at sample k_d, the algorithm waits for 1/2 of a cycle to begin gathering samples to build a half-cycle length feature vector. The need for 1/2 of a cycle is so because it is usually when the first effects of CT saturation begin to appear. This feature vector is formed by a sliding window of half a cycle length, which slides at steps of 1/8 of a cycle, up to 2 cycles after the first event detection at k_d, and performs a classification about the external fault at each of these steps. Differently from SVM 1, which has a signature vector composed of operating and restraining energies from phases A, B, C, and negative sequence Q, SVM 2 does not take negative sequence into account, since its high sensitivity could make the SVM underperform when distinguishing events with CT saturation from events with evolving external-to-internal fault. Therefore, SVM 2 uses feature vectors composed of 768 features (128 samples for half-cycle window length, of phases A, B, and C). The sliding window procedure is illustrated in Fig. 3. Accordingly, based on the output of SVM 2, if an ordinary external fault, with or without CT saturation, is detected, a warning signal is issued. Alternatively, if an external-to-internal evolving fault is detected, a trip signal is issued to open the circuit breakers, as depicted in Fig. 1.

Fig. 3 illustrates an external double-line-to-ground (DLG) fault between phases B and C (BCG fault) on the high voltage side, with the occurrence of saturation of the CT of the fault side. Firstly, the external fault is detected by SVM 1, as shown in Fig. 3(b). Then, the presented algorithm waits for 1/2 of a cycle to gather samples up to a half-cycle length and perform a classification. This window slides until two cycles after

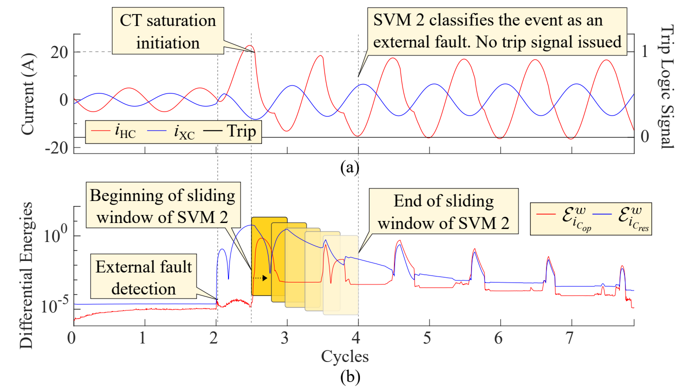

Fig. 3. External DLG fault on the high voltage side of the transformer, with CT saturation on the fault side: (a) high and low sides currents and trip logic signal; (b) operating and restraining wavelet coefficients energies of phase A and an illustration of the sliding window used for SVM 2

the first disturbance detection. As portrayed in Fig. 3(a), no trip was issued by SVM 2, even in the occurrence of severe CT saturation, and its implications regarding the increase of operating and restraining energy levels, illustrated in Fig 3(b).

Besides external faults with CT saturation, SVM 2 can also detect and classify, if they occur inside its sliding windows, evolving external-to-internal faults, i.e., internal faults which take place during external faults. An evolving external-to-internal fault is illustrated in Fig. 4, whereas an external single-line-to-ground (SLG) fault on phase A (AG fault), correctly classified by SVM 1, evolves to a turn-to-earth fault on phase A of the wye winding, as depicted in Fig. 4(a). Upon the inception of the fault, the algorithm has some delay in classifying the event. This delay is due to the sliding windows, which yields a maximal delay, from the fault inception time, of 1/8 of a cycle ($\approx$ 2.1 ms) to perform an event diagnosis. This delay, likewise the identification of the evolving external-to-internal fault and its resulting trip signal, is depicted in Fig. 4. As previously stated for SVM 1, issuing a trip signal prevents the SVM blocks from performing classifications.

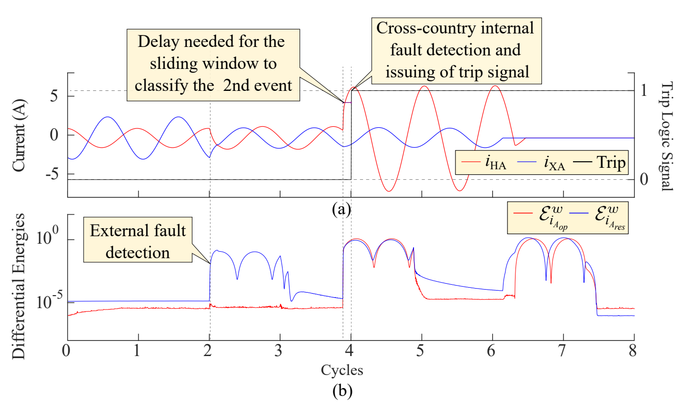

Fig. 4. External AG fault the high voltage side, with an evolving external-to-internal turn-to-earth fault on the wye winding: (a) high and low sides currents and trip logic signal; (b) operating and restraining wavelet coefficients energies

IV. Case Study

The proposed SVM-based differential protection was assessed using the MATLAB/Simulink platform. Fig. 5 illustrates a single line diagram of the test system, which consists of a 100 MVA rated power transformer, with its primary and secondary windings connected to two Thevenin equivalents, represented by S1 and S2, respectively, and their corresponding impedances, Z_{S1} and Z_{S2}. Details about the system parameters can be found in the appendix of this paper. Data were acquired with SNR = 65 dB and at a sampling frequency of 15360 Hz, i.e., 256 samples per cycle in a 60 Hz system. According to [13], this sampling frequency is ideal for methods based on fault-induced transients. The employed mother wavelet was the Daubechies with four coefficients (db(4)).

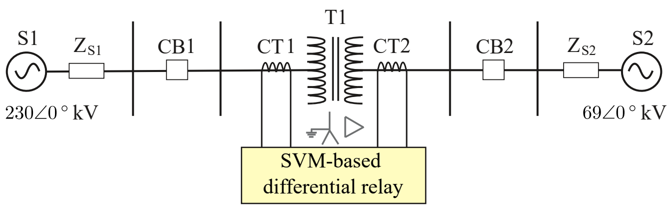

Fig. 5. Electrical system single line diagram

The following databases, comprising a wide variety of events, including critical internal faults, external faults with CT saturation, and external-to-internal evolving faults, were generated to provide data for the training and assessment of the SVM models:

- Database 1 (external faults): AG, BG, CG, AB, BC, CA, ABG, BCG, CAG, and ABC faults, on high and low voltage sides of the transformer, with varying fault inception angle $\theta_f = \{0, 15, ..., 165, 180\}$ electrical degrees, and fault resistance $R_f = \{0.1, 1, 5, 10, 20, 50, 100\}$ Ω (1820 records).
- Database 2 (internal faults): AG, BG, CG, AB, BC, CA, ABG, BCG, CAG, and ABC faults, on sides of both high and low voltage windings of the transformer, while varying fault inception angle $\theta_f = \{0, 15, ..., 165, 180\}$, and fault resistance $R_f = \{0.1, 1, 5, 10, 20, 50, 100\}$ Ω (1820 records).
- Database 3 (critical internal faults):
  1) turn-to-turn and turn-to-earth faults on the phases A, B, and C wye windings. With varying percentage of turns affected by the fault $e = \{1, 2, 3, ..., 98\}\%$ (588 records).
  2) turn-to-turn and turn-to-earth faults on the delta side between phases A-to-B, B-to-C, and C-to-A windings. With varying percentage of turns affected by the fault $e = \{1, 2, 3, ..., 98\}\%$ (588 records).
- Database 4 (transformer energizations): switching performed on the high voltage side (230 kV), with the secondary terminal opened, and varying the high-voltage circuit breaker closing time at angles of $\theta_s = \{0, 1, 2, ..., 179, 180\}$, considering the presence and the absence of residual flux in each assessed angle (362 records).
- Database 5 (external faults with CT saturation): same as database 1, but with forced CT saturation due to burden load resistance of $R_b = 15\Omega$, as well as a purposely change in CT magnetizing curves to smaller tapes (200-5 A). Only the CT of the faulted side was designed for saturation, whereas the other CT held the standard non-saturation parameters (1820 records).
- Database 6 (external fault + evolving external-to-internal faults): AG, BG, CG, AB, BC, CA, ABG, BCG, CAG, and ABC external faults on the high and low voltage sides of the transformer, evolving to turn-to-turn and turn-to-earth faults on the windings of both sides, applied to phases A, B and C. Variation of a percentage of turns affected by the fault with $e = \{5, 35, 65, 95\}\%$, similarly to what is done in database 3 (800 records). The internal faults take place between a cycle and a cycle and a half after the external fault inception time, at randomized fault inception angles.

A. SVM Models Training

Data-driven classification methods need previous training. In this subsection, the dataset description and partitioning are presented, and the training and test process is further described. The datasets used to train and test the SVM models are described as follows:

1) SVM 1 Dataset: the dataset in which SVM 1 is trained and tested is composed of the records from databases 1, 2, 3, and 4, which refer to external faults, internal faults, critical internal faults, and energization events, respectively.
2) SVM 2 Dataset: the dataset used to train and test SVM 2 is composed of the records from databases 1, 5, and 6, which refer to external faults, external faults with CT saturation, and evolving external-to-internal faults, respectively.

The datasets are formed by gathering the signature vectors, associated with their respective event labels, from the various records. These datasets are then split randomly into an 80% and 20% proportion to form the training and test sets, respectively. Although this split is performed randomly, the proportions of each type of event, regarding the total of cases in the respective dataset, holds for divided training and test sets. For instance, if the energization records represent 7% of the total records of the dataset used for SVM 1, the training and test set obtained from splitting this dataset should keep the same 7% proportions of energization records in the training and test sets.

Since SVM 1 has a fixed window used as input, and SVM 2 employs a sliding window, their training and assessment differ a little. For the SVM 2 dataset, since one simulated case represents eight observations (due to the sliding window process, which slides eight times), this might yield a large dataset, which could slow down the training process without bringing relevant information to the model. Therefore, an evaluation of the size of the training set versus the test set's overall accuracy was performed. After 60% of the total size of SVM2's training set, no accuracy improvements were identified. Thus, SVM 2 was trained with only 60% of its total training set.

Once the training and test sets are ready, feature scaling is performed on both sets using min-max normalization, taking into account only the minimum and maximum values of the features from the training set. After the feature scaling, optimal SVM parameters are found by performing a grid search associated with 10-fold cross-validation. SVM 1 was modeled using a radial basis function (RBF) kernel, which requires an associated γ parameter, in addition to the regularization hyperparameter C, which are both defined using the grid search. Since SVM 2 is based on a considerable energy window length, meaning many features, it was modeled with a linear kernel, since there was no need to map the signature vectors into higher dimensions to provide accurate classification. Moreover, using a linear kernel also ensures faster training and classification time. The preprocessing steps were performed using MATLAB, and the SVM models were trained using the LIBSVM [14].

As the best parameters are found, the SVM models are validated against new, unseen data from the test sets. Their performance during this stage is presented in the remainder of this section.

B. SVM 1 Performance Assessment

After training and parameter tuning, the performance of SVM 1 was assessed individually considering a test set of 20% of its total dataset, and the confusion matrix in Fig. 6 was obtained. The evaluated test set consists of 363 external fault cases, 599 internal fault cases, and 72 transformer energization cases. To tackle this three classes problem, the one-versus-all algorithm was adopted, with three SVM models, each one associated with a respective class of event. The C and γ parameters were the same for the three models, with values of 500 and 100, respectively. Furthermore, 10-fold cross-validation was performed during the training stage to avoid overfitting the SVM models. SVM 1 successfully classified all of the tested cases, presenting an accuracy of 100%. Therefore, no false trips were issued, all the internal faults were properly identified, including turn-to-turn and turn-to-earth faults, even when only a low percentage of turns were involved in the fault.

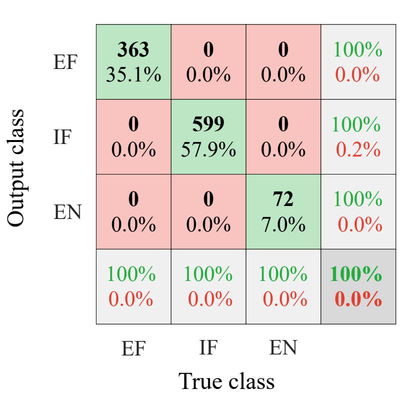

Fig. 6. SVM 1 confusion matrix, obtained from the respective test set

C. SVM 2 Performance Assessment

The performance of SVM 2 was assessed with its respective test set, which is composed of records of 182 ordinary external faults, 182 external faults with CT saturation, and 160 evolving external-to-internal faults, whereas each record yields 8 feature vectors, represented as observations in the dataset. The performance of SVM 2 against its test set is presented in Fig. 7, with an overall accuracy of 98.7% in distinguishing external faults, with and without CT saturation, from evolving external-to-internal faults which evolved from external faults. However, in a few cases of severe CT saturation, external faults were mistakenly classified as internal faults, resulting in false trips. Conversely, only 1.3% of the observations regarding external faults were misclassified as internal faults, whereas no external-to-internal evolving faults were misclassified as external faults.

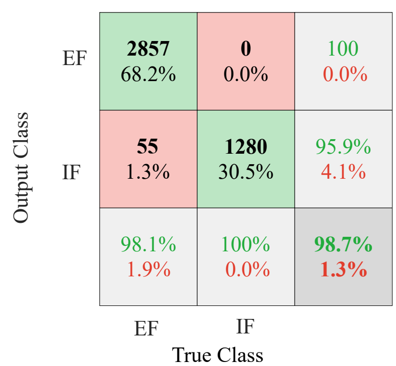

Fig. 7. SVM 2 confusion matrix, obtained from the respective test set

D. Protection Scheme Overall Performance

A further evaluation of the presented confusion matrices is performed employing the protection metrics proposed in [18]. These metrics are: accuracy, which stands for the overall accuracy of the confusion matrix; dependability, which represents the ability of the protection algorithm to detect and isolate faults; security portrays the method's capability to detect faults and not perform false trips selectively; safety is associated with the ability to classify faulty conditions correctly as faults; and lastly, sensibility is related to the protection scheme proneness to perform false trips. A comparison regarding these protection metrics and the operating time of SVM 1, SVM 2, and a conventional algorithm is presented in Table I. The used traditional phasor-based power transformer differential protection algorithm includes [19]: the phase differential function 87T with harmonic restraint and harmonic blocking units, the negative sequence differential function 87Q, and the REF function. The same test sets were used to evaluate both the proposed and the conventional protection schemes

Test results presented in Table I demonstrate the overall reliability of the proposed protection scheme, especially SVM 1, which presents a 100% success rate in all four evaluated metrics. Likewise, SVM 2 also has a good performance on the assessed metrics, including 100% dependability and safety percentages. However, it fails a little on security and sensibility indices, i.e., it presents some false trips, mainly due to severe CT saturation during some external faults. Conversely, the conventional protection presented no false trips; however, it performed poorly when dealing with

TABLE I
PROTECTION SCHEME METRICS

Metric	Test Set 1		Test Set 2	
	SVM 1	Conventional protection	SVM 2	Conventional protection
Dependability	100%	95.5%	100%	63,1%
Security	100%	100%	98.11%	100%
Safety	100%	90.1%	100%	86,1%
Sensibility	100%	100%	95.88%	100%
Operating time	260 μs (fixed)	19.3 ms (average) $\sigma = 4.5$ ms	2.1 ms (maximal)	18.6 ms (average) $\sigma = 1.5$ ms

critical internal faults involving low percentages of turns, presenting a bad performance in dependability and safety indices, for both tested datasets. Therefore, it is shown that the presented SVM-based differential protection outperforms the conventional one on almost all of the evaluated metrics.

Besides the efficiency in performing the correct distinction of events, another fundamental metric of power systems protection is the operation speed for rapid fault clearance. Thus, the conventional protection's operating time regarding both test sets is presented in Table I. An average operating time of 19.3 ms, with a standard deviation σ of 4.5 ms, was obtained with the same test set of SVM 1; whereas to deal with evolving external-to-internal faults, it presented an average operating time of 18.6 ms, with a standard deviation σ of 1.5 ms. Furthermore, since the presented protection algorithm takes into account, for the first event classification, i.e., SVM 1, four post-fault samples, for a 60 Hz system and a sampling rate of 15360 Hz, this means a classification time of approximately 260 μs. Additionally, considering the sliding window step of 32 samples from SVM 2, for the aforementioned sampling rate, this yields a maximal operating time of 2.1 ms for accurate external-to-internal evolving fault identification. This last result is quite promising since it presents an improvement for accurate external-to-internal evolving fault detection as compared with [7], which had an average operating time delay of 10 ms for identifying this type of event. It also demonstrates the presented method to be faster than other data-driven-based methods presented in the literature, since [9]–[12] needed a time delay of 4.12 ms, 9.7 ms, 20 ms, and 10 ms, respectively, to perform classification of similar events.

The other aforementioned protection metrics are not brought to comparison with the performances presented in [9]–[12] since it is difficult to replicate both the protection methods, as well as the databases presented in these papers. Therefore, the most coherent metric to evaluate among all these papers is the operating time needed for the event classification.

E. Hardware Implementation

According to Fig. 1, the proposed protection method can be divided into two parts: the wavelet-based preprocessing algorithm and SVM-based algorithms. Only the wavelet-based preprocessing method needs to run in each sampling time, i.e., the computational burden in a hardware implementation must be less than 65 μs to run in the real-time by considering $f_s = 15360$ Hz. Conversely, the SVM blocks, which are time-consuming algorithms, run only when a disturbance is detected. In a hardware implementation, however, its computational burden can take more than one sampling time.

The wavelet-based preprocessing of the proposed protection requires only 730 floating-point operations (FLOPs) per sampling time. This method was based on [6], which used the floating-point DSP TMS230F28335 to demonstrate its real-time implementation feasibility. The computational burden, per sampling, of all boundary wavelet coefficients energy variables, was about 11.12 μs, which is less than 65 μs. FLOPs are considered to be addition and multiplication operations. Memory management is not considered.

SVM 1 classification requires about 73728 FLOPs, 768 square root, and 768 exponential operations to perform the event classification. Therefore, the computational burden is much higher than that required by the wavelet-based preprocessing method. However, currently, there are powerful DSPs able to perform millions of FLOPs per second (MFLOPS). For instance, the DSP TMS320C6748 performs up to 2746 MFLOPS [20], i.e., 2746 FLOPs per μs or about 178,490 FLOPs in 65 μs. In addition, the computational burden of SVM 1 can be partitioned in several sampling times without compromising the protection operation time. For instance, the DSP TMS320C6748 would provide about 356,980 FLOPs to compute the SVM 1 in two sampling times (130 μs), which would be enough for this purpose.

SVM 2 uses a linear kernel, i.e., it involves no square root nor exponential operations. However, it uses large feature vectors, requiring about $751,593$ FLOPs to perform the classification only when requested. Therefore, the real-time implementation of SVM 2 would be possible in about five sampling times (325 μs) by using the DSP TMS320C6748 because 892,450 FLOPs would be enough for this purpose.

To further validate the feasibility of hardware implementation of the proposed protection scheme, the method was implemented in the *NI sbRIO 9637* board, which has a built-in real-time processor. Even though this board is not designed to run high time-consuming digital signal processing and machine learning algorithms, it is a dedicated hardware that runs a real-time operating system. Therefore, this limited DSP could perform the event classification of SVM 1 in approximately 3 ms, which is still a good operating time for protection purposes since the conventional method can operate in about one cycle (≈16 ms).

V. CONCLUSIONS

This paper presented a novel data-driven power transformer differential protection based on SVMs. The proposed protection algorithm employs the RT-BSWT to extract high-frequency components of the current signals and uses this transient information as inputs to the SVM algorithm to perform event classification. When an internal fault is identified, a trip is issued to protect the power transformer. After a thorough assessment of the proposed method, the main contributions of this paper can be summarized as follows:

- The SVM-based protection method presented 100% of success rate in distinguishing external faults, internal faults, and energization events. The relay operating time in a hardware implementation is from a few hundred μs to a few ms, depending on the used DSP.

- External transformer faults followed by a CT saturation or an internal fault (evolving external-to-internal faults) are challenging cases. However, the method presented a 98.7% accuracy in distinguishing external faults with and without CT saturation from evolving external-to-internal faults. The proposed relay is able to perform this event distinction with an operating time of up to 2.6 ms.
- The proposed SVM-based protection scheme outperformed the conventional one in dealing with critical turn-to-turn internal faults and presented the fastest operating time.
- Part of the presented data-driven protection scheme was implemented in the *NI sbRIO 9637* board. Although this board is not designed to run high time-consuming digital signal processing and machine learning algorithms, it was able to run the necessary preprocessing and SVM 1, issuing a trip signal in about 3 ms, which is faster than conventional protection and other existing data-driven power transformers protection schemes.

The promising results, combined with the demonstrated feasibility of hardware implementation of the SVMs, as well as the RT-BSWT, make the proposed protection a good possibility for enhancing power transformer differential protection. However, to increase robustness of the proposed method, more types of event should be evaluated in the future, such as transformer energization with presence of an internal fault, sympathetic inrush, and overexcitation conditions.

VI. Appendix

The modeling of the power transformer evaluated in this paper is based on the system presented in [6]. The power transformer has a rated power of 100 MVA, a voltage ratio of 230:69, and a YNd1 configuration. The impedances related to the primary and secondary winding are Z_p=$2.04 + j12.54$ Ω and Z_s=$1.44 + j38.04$ Ω, respectively. The impedances related to the Thevenin equivalents of the transmission system, illustrated in Fig. 5 and named as Z_{S1} and Z_{S2}, are presented in Table II.

TABLE II
THEVENIN EQUIVALENT IMPEDANCES OF THE TRANSMISSION SYSTEM

Thevenin equivalent	Z_0	Z_1
Z_{S1}	$16.07 + j25.04\ \Omega$	$12.05 + j18.78\ \Omega$
Z_{S2}	$5.52 + j8.61\ \Omega$	$4.02 + j6.26\ \Omega$

TABLE III
NONLINEAR CHARACTERISTICS OF THE MAGNETIZING BRANCH OF THE USED TRANSFORMERS

T1		CT1 (800-5 A)		CT2 (1200-5 A)		Sat. CT (200-5 A)	
i(A)	Φ(Wb)	i(A)	Φ(Wb)	i(A)	Φ(Wb)	i(A)	Φ(Wb)
0.144	498.137	0.052	0.112	0.054	0.338	0.283	0.038
0.478	523.044	0.075	0.225	0.132	1.606	0.376	0.079
1.211	547.951	0.135	0.450	0.175	1.876	0.655	0.188
2.540	572.858	0.165	1.125	0.189	2.251	1.095	0.338
6.446	579.085	0.301	1.501	0.341	2.626	1.620	0.379
8.954	585.312	0.555	1.688	0.561	2.926	6.512	0.386
15.595	591.538	0.687	1.876	0.976	3.001	43.852	0.390
20.396	597.765	44.856	2.251	9.440	3.477	449.60	0.394
35.461	603.992	-	-	-	-	-	-

The nonlinear characteristics of the power transformer T1 and the current transformers CT1 and CT2 are shown in Table III as current versus flux (i, Φ) points of the transformer magnetizing curves. The current versus flux (i, Φ) points of the magnetizing curve of the CT used to force CT saturation is also presented in Table III with the name of Sat. CT.

References

[1] A. Guzman, S. Zocholl, G. Benmouyal, and H. Altuve, "A current-based solution for transformer differential protection-part i: problem statement," *IEEE Trans. Power Del.*, vol. 16, no. 4, 2002.

[2] K. Behrendt, N. Fischer, and C. Labuschagne, "Considerations for using harmonic blocking and harmonic restraint techniques on transformer differential relays," *Journal of Reliable Power*, vol. 2, no. 3, pp. 36–52, 2011.

[3] 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, June 2015.

[4] X. Lin, H. Weng, and B. Wang, "Identification of cross-country fault of power transformer for fast unblocking of differential protection," *IEEE Transactions on Power Delivery*, vol. 24, no. 3, pp. 1079–1086, July 2009.

[5] "Transformer reliability survey," CIGRE, Tech. Rep., 2015.

[6] R. P. Medeiros, F. B. Costa, and K. M. Silva, "Power transformer differential protection using the boundary discrete wavelet transform," *IEEE Trans. Power Del.*, vol. 31, no. 5, pp. 2083–2095, Oct 2016.

[7] 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.

[8] R. P. Medeiros and F. B. Costa, "A wavelet-based transformer differential protection: Internal fault detection during inrush conditions," *IEEE Transactions on Power Delivery*, vol. 33, no. 6, pp. 2965–2977, Dec 2018.

[9] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep neural network into power transformer differential protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 865–876, 2020.

[10] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Designing a composite deep learning based differential protection scheme of power transformers," *Applied Soft Computing*, vol. 87, p. 105975, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494619307562

[11] A. M. Shah and B. R. Bhalja, "Fault discrimination scheme for power transformer using random forest technique," *IET Generation, Transmission Distribution*, vol. 10, no. 6, pp. 1431–1439, 2016.

[12] 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, 2013.

[13] F. B. Costa, "Fault-induced transient detection based on real-time analysis of the wavelet coefficient energy," *IEEE Trans. Power Del.*, vol. 29, no. 1, pp. 140–153, Feb 2014.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[15] S. S. Haykin, *Neural networks and learning machines*, 3rd ed. Upper Saddle River, NJ: Pearson Education, 2009.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, p. 273297, Sep. 1995. [Online]. Available: https://doi.org/10.1023/A:1022627411411

[17] A. Rocha and S. K. Goldenstein, "Multiclass from binary: Expanding one-versus-all, one-versus-one and ecoc-based approaches," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 289–302, Feb 2014.

[18] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y. Shin, "High-impedance fault detection in the distribution network using the time-frequency-based algorithm," *IEEE Transactions on Power Delivery*, vol. 30, no. 3, pp. 1260–1268, 2015.

[19] K. A. Tavares and K. M. Silva, "Evaluation of power transformer differential protection using the atp software," *IEEE Latin America Trans.*, vol. 12, no. 2, pp. 161–168, March 2014.

[20] *TMS320C6748 Fixed- and Floating-Point DSP Datasheet*, Texas Instruments, 2009.