Sustainable Energy, Grids and Networks 32 (2022) 100855

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan

# Voltage-sag source detection: Developing supervised methods and proposing a new unsupervised learning

Younes Mohammadi [a,*], Seyed Mahdi Miraftabzadeh [b], Math H.J. Bollen [a], Michela Longo [b]

[a] Department of Engineering Sciences and Mathematics, Luleå University of Technology, Skellefteå Campus, Forskargatan 1, 93187 Skellefteå, Sweden
[b] Department of Energy, Politecnico di Milano, Via Lambruschini 4, 20156 Milano, Italy

ARTICLE INFO

*Article history:*
Received 22 March 2022
Received in revised form 1 July 2022
Accepted 10 July 2022
Available online 16 July 2022

*Keywords:*
Voltage sag (dip)
Source detection (location)
Supervised learning
Signal to noise
K-means
Sparse principal component analysis (SPCA)

ABSTRACT

Recognition and analysis of voltage sags (dips) allow network operators to predict and prevent problems in real-life applications. Clearing the voltage sag source by direction detection methods is the most effective way to solve and improve the voltage sags and their related problems. However, the existing analytical methods use single or two input features as phasor-based (PB) or instantaneous-based (IB) values. Hence, their limited maximum accuracy is given at 93% and 84% when using PB features for noiseless and high-level noise signals, respectively. To increase the detection accuracy, the main contributions of this research by proposing machine learning (ML) methods include: (a) Developing nine supervised methods including support vector machine (SVM)-based, tree-based, others, and an ensemble learning of said methods, and providing a comparative analysis (b) Employing a set of PB, IB, and both PB and IB input features as noiseless and noisy; (c) Finding the best developed supervised methods by highest possible accuracy under subsets said in (b); (d) Proposing a new unsupervised method fed by both PB and IB features using a sparse principal component analysis (SPCA) applied to a k-means clustering with an internal SPCA approach. The proposed unsupervised schema does not use the upstream/downstream labels in developed supervised methods. Extensive simulations of voltage sags due to fault and transformer energizing on a Brazilian regional network show that regardless of the sag sources, input feature subset, and noise levels, the random forest (RF) models yield the best performance so that noiseless-RF (99.84%) using both PB and IB features is the most effective one. The proposed unsupervised method outcomes an overall accuracy of 99.17%-noiseless and about 90% for high-level noises. This performance is higher than analytical methods, very close to SVM-based supervised methods, and uses no predefined labels. Moreover, the results of Slovenian field measurements confirm the effectiveness of the best-developed supervised methods and the proposed unsupervised learning.



## 1. Introduction

Voltage sags are one of the most critical power quality disturbances as short-duration voltage variations [1]. Despite their short duration, such events can cause severe problems in transmission and distribution systems, micro-grids, and industrial or customer facilities. Various reasons, including short-circuit faults, starting/loading induction motors, transformer energizing inrush current, and large loads operations, may cause the sags. Due to all these events (except starting the motors), the sags propagate throughout the power network, affecting connected loads far from the source location [2–4]. Therefore, monitoring, analyzing, and characterizing the sag events and identifying their causes can help mitigate the substantial loss of product of a typical industrial installation and improve power quality [2]. Thus, detecting the sag sources by an accurate formulation (when there is no mitigation equipment) is an important strategy to define the responsibility of both sides of power supply and consumption for the sags caused by fault/transformer energizing in offline applications. There is a need for direction detection methods to operate as a directional function in relays and secured backup protection to prevent the unwanted operations of protection, where more accurate methods are needed. Moreover, a comprehensive comparison of such detection methods as analytical and ML-based is lacking. ML has been introduced as an alternative to analytical methods. It introduces new opportunities, but there are challenges such as the definition of input features and the need for large amounts of data.

* Corresponding author.
*E-mail address:* Younes.mohammadi@ltu.se (Y. Mohammadi).

https://doi.org/10.1016/j.segan.2022.100855

**Nomenclature**

**Abbreviation**

| | |
|---|---|
| CR | Classification rate for each class |
| CV | Cross-validation |
| DT | Decision tree |
| EL | Ensemble learning |
| FAR | False alarm rate for each class |
| FP, FN | False positive and negative values for each class |
| IB | Instantaneous-based |
| KNN | K-nearest neighbor |
| LR | Logistic regression |
| MCR | Misclassification rate |
| ML | Machine learning |
| NN | Neural network |
| PB | Phasor-based |
| PMUs | Phasor measurement units |
| PQM | Power quality monitor |
| $c_j$ | Label of $j$th class ($j = 1, 2$) |
| $c_{jmax}$ | Majority weighted voting class label between two classes in KNN |
| $d$ | Polynomial degree for polynomial kernel of SVM |
| $E$ | Residual matrix used in the process of first SPCA |
| $G_i$ | Gini index to estimate the node impurity in DT |
| $h_i$ | $i$th principal feature vector in the $B^{m\times p}$ |
| $h_i'$ | $i$th principal feature vector in the $C^{m\times 2}$ |
| $I(a, b)$ | Indicator function to show the relationship between a and $b$ |
| $N_{correct}$ | Number of correct results for separated or all ($m$) parts of a dataset |
| $N_k$ | Number of samples in the $k$th test fold of CV |
| $N_{total}$ | Number of total test samples for separated or all ($m$) parts of a dataset |
| $p$ | Probability of predicted class for LR |
| $p_{ij}$ | Gaussian distribution measuring similarity of high-dimensional vectors in t-SNE |
| $P_n$ | Power of noise |
| $P_s$ | Power of sag signal |
| $q_{ij}$ | Student t-distribution measuring similarity of low-dimensional vectors in t-SNE |
| RBF | Radial basic function |
| RF | Random forest |
| SPCA | Sparse principal component analysis |
| SVM | Support vector machine |
| TE | Transformer energizing |
| TP, TN | True positive and negative values for each class |

**Parameters and variables**

| | |
|---|---|
| $A^{m\times n}$ | Original feature matrix, $n = 28, 17, 11$ |
| $a_i$ | Lagrangian multiplierused in SVM |
| $B^{m\times p}$ | Principal feature matrix, output of first SPCA |
| $b$ | Bias term in SVM and LR |
| $C^{m\times 2}$ | Principal feature matrix existing in k-means, after clustering |
| $R^{p\times 28}$ | Sparsity matrix used in the process of first SPCA |
| $v_{i,j}$ | Ratio of class $j$ instances among the training distances in the $i$th node in DT |
| $x_i$ | $i$th normalized feature vector (rows) in the $A^{m\times n}$ |
| $y_i$ | $i$th true (real) label corresponds to $x_i$, {0,1} |
| $\hat{y}$ | predicted label by the classifiers |
| $W_j(x_i, \acute{x})$ | Weight $\geq 0$ of the $i$th sample relative to new sample ẋ in the same tree in RF |
| $\omega_{ij}$ | Assigns a sample to a cluster in k-means according to distance |
| $\gamma$ | Parameter for RBF kernel of SVM, which scale the distance |
| $\beta_1, \beta_2$ | Coefficients for sigmoid kernel of SVM |
| $\mu_j$ | Cluster centers or centroids in the $j$th cluster ($j = 1,2$) |
| $\delta$ | Hypothesis function of LR |
| $\theta_i$ | Parameter vector of the LR |

### 1.1. Related works

Several methods have already been reported for detecting the relative location of voltage sag sources [5–45]. The approaches to distinguish between downstream and upstream sags can be classified into three types. The first and basic type involves the analytical methods, which are based on different criteria, such as power and energy changes during voltage sag (so-called disturbance power and energy criteria) [5–7], voltage–current analysis due to criteria of real current component changes and slope sign of system trajectory [8,9], change of impedance and resistance (incremental positive sequence impedances, negative and positive sequences impedances and impedances based on Park' components) [10–15], analysis of only voltages; comparing the per unit voltage magnitude at primary and secondary of a transformer or the difference between positive sequence voltage magnitudes for each two PQMs covering an area/zone between themselves for the pre and post sag conditions [16,17] and only currents (mostly based on positive sequence current magnitude and/or angle changes) [18–23]. Within this group, statistical methods with a combination of existing rules have been applied to enhance the performance and degree of confidence [24–26]. Higher effectiveness was achieved for several types of methods by using positive-sequence components on the line phasors [27], applying Clark components [28–30], and positive-sequence components on the instantaneous voltage and current signals [31,32]. Moreover, a short [33,34] and detailed [3] analysis has been done on some mentioned methods. The second type of methods includes signal processing techniques, which develop the method presented in [5] based on power and energy changes. These methods are based on the Hilbert transform [35,36], S transform [37], and a combination of S and TT (time-time) transforms [38], which improve the basic method in [5]. The third type focuses on the methods that utilize ML tools. An SVM classifier fed by several features sensitive to source location was used in [39], only for sags due to faults. Later, an EL on DT basic

learners fed by optimum features improved the accuracy [40]. In [41], a feed-forward NN by a criterion as a combination of four features extracted from [5,8–10] and sag characteristics (magnitude, phase-jump, harmonic content, imbalance, and the slope was designed. Further [42], a conventional NN fed by five features extracted from [5,8–10] on three phases, was planned to locate the direction of faults. Later, independent recurrent NN was proposed for sag source location due to only faults [43]. Furthermore, a data-driven approach using random matrix theory found the location of complex sources of sags (multiple faults) with a complicated mathematical method [44]. Far ahead, a genetic algorithm was applied to select the best measurement points and relatively locate the faults [45].

### 1.2. Drawbacks of existing methods

The limited accuracy is the first shortcoming of analytical existing methods. Most of the methods mentioned above (both analytical and ML ones) have been analyzed for sags due to faults, while only a few analytical methods [27–31] considered sags caused by TE. The voltage sags caused by TEs differ from faults since they are always asymmetrical, have harmonic content (especially in current), and have a long recovery time, which can have a different impact on the loads and generators such as wind turbines [46,47]. Therefore, finding the location of TEs is important for the network operators when they do not have information about events such as the energizing after auto-reclosing the protection relays. An example of energizing after a successful auto-reclosing relay was shown in [48]. Hence, the methods must be developed to face the transient harmonic current during a TE event.

Moreover, the analysis of methods was done for mostly noiseless input signals, and only a few research considered the noisy signals [41,44]; meanwhile, the real measurements include different levels of noise, which affect the performance of methods. In addition, developing ML-based methods fed according to the availability of input features (PB, IB, or both) is lacking. For example, for applying the methods on measurements obtained from PMUs, proposing an intelligent method fed by PB features is still needed. Additionally, a comprehensive comparison of such detection methods as analytical and ML-based is lacking to show the effectiveness of ML methods. As another current shortfall, labeling many voltage sags recorded by the PQMs worldwide as an input dataset into supervised ML-based methods is always an issue; hence there is a need for designing an accurate unsupervised ML method to handle the unlabeled inputs.

### 1.3. Contribution and applicability

To cover the shortfalls of existing methods, the *contributions of this paper* for detecting the sag sources, both faults and TEs, include:

(a) A comprehensive comparison of detection methods as analytical and ML-based.

(b) Developing supervised ML methods to enhance the accuracy of analytical methods. The methods include SVM-based methods with kernels of polynomial, RBF, sigmoid, and linear; tree-based ones, i.e., DT and RF; others as KNN and LR; and an EL on all said methods. The selection of the traditional ML methods is based on their speed and relative simplicity compared to neural networks. This part is highly more comprehensive than the literature [39,40].

(c) Considering the different sets of input features, including three feature subsets as PB, IB, and Total, with four noise levels as noiseless, 20, 30, and 40 dB ($3 \times 4 = 12$ models for each of the nine classifiers).

(d) Finding the best (highest accuracy) developed supervised methods among the nine classifiers. Regardless of the sag sources, input feature subset, and noise levels, the RF models will show the best performance from the accuracy perspective.

(e) Proposing a new unsupervised ML method fed by Total features with four noise levels to handle datasets without upstream and downstream labels. This is worth because labeling real data is usually impossible. This proposed method utilizes an SPCA applied to a k-means clustering that already uses an internal SPCA's initializing scheme to start effectively clustering the feature vectors. This part is done as the first work in this area and introduces unsupervised methods that are more effective than analytical methods. They have a very close accuracy to SVM-based supervised methods and do not use the predefined labels.

The developed supervised and proposed unsupervised methods can detect the direction of both faults and TEs as the sag sources, which is different from distinguishing the sag sources. Applying the PB/IB features obtained from the transient period of voltage sags [49] can also increase the speed of the proposed ML methods. Therefore, the *applicability of the methods* can be as:

(a) Directional function in the relays, fault indicators, and secured-supervised backup protection, with high reliability, in case of faults as the source of sags (online applications).

(b) For the sags due to TE, the system operator can recognize the location of the automatic-energized transformer with high performance (offline applications).

(c) At the PCC of transmission systems and the distribution systems with a high penetration of renewable energy resources, which the responsible area of sags occurring in the transmission lines due to faults in the distribution level is in the interest of transmission network operators to assign penalties for the responsible area (offline applications).

The rest of this paper is organized as follows: Section 2 presents the input features, develops supervised methods, and proposes a new unsupervised one. The methods' results using extensive simulations and applying them to field-testing are given in Sections 3 and 4. A detailed discussion on the generality of the paper is given in Section 5, and Section 6 concludes the paper.

## 2. Developed supervised and proposed unsupervised methods

### 2.1. Feature extraction

The first stage for the ML methods in a two-classes (downstream/upstream) problem, developed supervised and proposed unsupervised in this paper, is feature extraction and selection. By registering a voltage sag at the PQMs, the voltage and current waveforms are applied to 23 known-analytical methods, explained in Table A.1. Each analytical method consists of one or two features, and a collection of all of them makes 28 extracted features $f_1-f_{28}$ to be used as input to ML methods, as explained in Table 1. In this way, the challenge of feature selection is solved. The 28 features are divided into three feature subsets in this study: Total features including all 28 features, PB features including 17 features as line and positive sequence phasors, and IB features with 11 features counting instantaneous positive-sequence and Clark's components.

### 2.2. Developed supervised methods

The authors of this paper have already investigated the effectiveness of SVM-based ML methods with the kernels of polynomial, RBF and linear. In the current study, the aim is first to see the effectiveness of another kernel of SVM, i.e., Sigmoid,

**Table 1**
Expression of different features extracted from voltage and current waveforms.

| Features (Category) | Description of features (unit) | Features (Category) | Description of features (unit) | Features (Category) | Description of features (unit) | Features (Category) | Description of features (unit) |
|---|---|---|---|---|---|---|---|
| $f_1$ (PB) | $\Delta\ E$ (pu) | $f_8$ (IB) | $\Delta\ e^+ - q$ (pu) | $f_{15}$ (IB) | slope $\left(i^+, \left\lvert v^+ cos\theta^+\right\rvert\right)$ | $f_{22}$ (IB) | $r_e{}^+$(pu) |
| $f_2$ (PB) | $\Delta\ E^+$ (pu) | $f_9$ (IB) | $\int_0^t \Delta tan\theta^+ dt$ | $f_{16}$ (PB) | $\Delta\ \lvert Z\rvert$ (pu) | $f_{23}$ (IB) | $\Delta\ (p_{a\beta}/\lVert v_{a\beta}\rVert)$ (pu) |
| $f_3$ (IB) | $\Delta\ e^+ - p$ (pu) | $f_{10}$ (PB) | $\Delta(Icos(\theta))_{min}$ (pu) | $f_{17}$ (PB) | $\angle Z_{sag}$ (rad) | $f_{24}$ (PB) | $R_e{}^-$ (pu) |
| $f_4$ (PB) | $\Delta\ Q$ (pu) | $f_{11}$ (PB) | $\Delta\ (I^+cos(\theta)^+)$ (pu) | $f_{18}$ (PB) | $\Delta\ \angle Z$ (rad) | $f_{25}$ (PB) | $\Delta\ \left\lvert I^+\right\rvert$ (pu) |
| $f_5$ (PB) | $tan\theta$ | $f_{12}$ (IB) | $\int_0^t \Delta\left(i^+cos(\theta)^+\right) d\tau$ (pu) | $f_{19}$ (IB) | $\int_0^t \Delta\lvert z\rvert\, d\tau$ (pu) | $f_{26}$ (PB) | $\Delta\ \angle I^+$ (rad) |
| $f_6$ (PB) | $\Delta\ Q^+$ (pu) | $f_{13}$ (PB) | slope$(I, Vcos\theta)_{min}$ | $f_{20}$ (IB) | $\int_0^t \Delta\angle z d\tau$ (rad ms) | $f_{27}$ (IB) | $\int_0^t \Delta\left\lvert i^+\right\rvert d\tau$ (pu) |
| $f_7$ (PB) | $tan\theta^+$ | $f_{14}$ (PB) | slope $\left(I^+, \left\lvert V^+cos\theta^+\right\rvert\right)$ | $f_{21}$ (PB) | $R_e{}^+$ (pu) | $f_{28}$ (IB) | $\int_0^t \Delta\angle i^+ d\tau$ (rad ms) |

min: corresponds phase with deepest voltage sag (Feature extractor methods are described in Appendix (Table A.1).

along with the others in a more generalized dataset, including the different noise levels and different feature subsets. The second aim is to check the efficiency of the tree-based methods (DT and RF), others (KNN and LR), and an EL on all mentioned methods. The results will further show more effectiveness of tree and other-based methods for the voltage sag source detection goal. In general, the selection of the traditional ML methods is based on their speed and relative simplicity compared to neural networks. This part is highly more comprehensive than the literature [39,40]

The training process of the developed supervised methods is shown in Fig. 1a for a binary classification problem with labels as 0 (an upstream location of sag sources) and 1 (a downstream location). The normalized dataset/feature in three different categories as Total ($A^{m\times 28}$), PB ($A^{m\times 17}$), and IB ($A^{m\times 11}$) are inputted to learning methods (developed supervised methods in this study) for noiseless and noisy (20, 30, and 40 dB) conditions; hence $3 \times 4 = 12$ models are obtained for each developed method. Matrixes. We have used a min–max normalization applied on the matrixes' columns to scale the values within [−1,1]. A 5-fold CV technique is also used to reduce the risk of training overfitting. The Total, PB, and IB datasets with four different values of $m$ are divided into 5 folds, and in each step, one-fold is considered a test fold and 4-fold as a training fold. The process is repeated 5 times, so for each configuration of the learning methods, all 5 folds work as a test set once. The Bayesian algorithm optimizes the parameters of the kernels and methods themselves. The used fitness function is evaluated by MCR on the test folds at the iteration of $g$ as (1) and (2). The main decision function for each developed classifier (except the EL) is explained in Table 2.

$$\mathrm{MCR}(I_i^g) = \frac{1}{5}\sum_{k=1}^{5}\frac{1}{N_k}\sum_{j=1}^{N_k} \mathrm{I}(\hat{y}_i, y_j),\ i = 1, \ldots, m \tag{1}$$

$$I(\hat{y}_i, y_j) = \begin{cases} 1, & \hat{y}_i \neq y_j \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

As another developed supervised method, this paper proposes an EL on the members explained in Table 2. The output decision of all the classifiers is combined to produce a single output using the rule of majority/hard voting. In majority voting, the predicted class label for a sample $x_i$ is the class label representing the majority of the class labels predicted by each of the eight individual classifiers. Using the EL increases the probability of more accurate classification as every classifier behaves differently with the training and testing samples. Therefore, the aggregated decision is taken after combining the decision of the classifiers. While combining the output decision of the classifiers, misclassification done by any one of the classifiers gets suppressed due to the majority of correct classification done by other remaining members. In this manner, we have used voting = *'hard'* in *Phyton*.

Finally, 12 optimum models (optimum means all parameters and variables are optimized during the training and testing process for each model) are obtained for each of the nine methods. Fig. 1b shows the best (highest accuracy) Total, PB, and IB models for noiseless and noisy conditions (Tables A.2a and A.2b for the parameters of best models). As seen in Fig. 1b, the developed RF methods have been selected in a majority of different categories. Since the methods have been trained offline in this study, thus, the best model is defined as the model with the highest accuracy only, i.e., finding the most effective classifier among nine ones for each of 3 input feature subsets and four noise levels. The best models are chosen based on Table 3, Section 5.1, and Appendix A.2.

### 2.3. Proposed unsupervised method

This section proposes a framework including the well-known unsupervised methods (SPCA followed by k-means). The framework is worth because it is the first attempt to design a schema that can handle unlabeled datasets for the voltage sag source detection problem. Since the labeling of voltage sag samples as upstream and downstream is usually impossible in reality. The results will further show the good efficiency of the proposed method, higher than the analytical methods and very close to the supervised SVM-based methods. The process of the proposed unsupervised scheme for noiseless and noisy conditions (four separate models) consists of four modules, as can be seen in Fig. 2 (optimizations by the grid search): (i) feature extraction and normalization; (ii) applying SPCA on the Total feature vectors, normalized high-dimensional, $x_i$ (a column min–max normalization), which results in the vectors $h_i$ with $p$ principal features; (iii) employing k-means clustering initialized by an internal SPCA for grouping principal features; (iv) visualizing the original and clustered features in 2D space using t-SNE.

*(i) Feature extraction and normalization*

By registration of a voltage sag at the PQMs, the voltage and current waveforms are applied to the 23 known-analytical methods (Table A.1). Then, 28 features are extracted as the Total feature subset. A min–max normalization is then applied to the columns of matrix $A^{m\times 28}$ to scale the values within [−1,1]. The $m$ would have four values for noiseless, 20 db, 30 db and 40 db conditions.

*(ii) Applying SPCA*

The dataset extracted from part (i) was first inputted directly to k-means. However, the clustering results were inefficient (see, for example, Table 20). Hence, the authors decided to use some feature size reduction algorithms. Principal Component Analysis (PCA) is one of the most powerful algorithms in data preprocessing for dimensionality reduction in many applications. PCA is a linear algorithm that transforms the original data into a linear combination of the new uncorrelated features. The new features aid in bringing non-obvious patterns in the data to the forefront and can improve the performance of the ML methods [57,58]. However, in our work, an SPCA algorithm was chosen

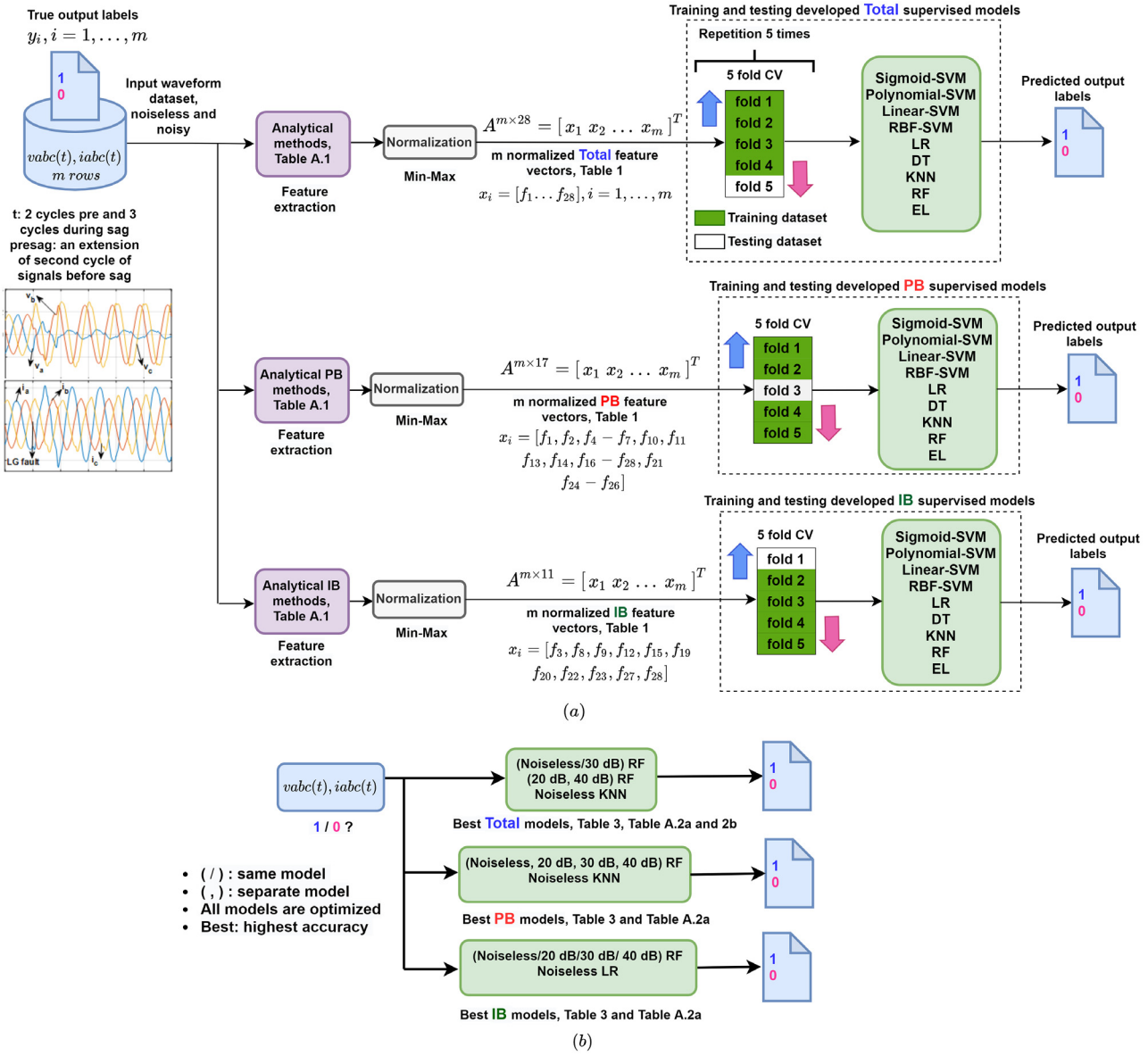


**Fig. 1.** Process of developed supervised methods (a) Training the models for noiseless and noisy datasets (b) Implementation of best-developed models.

and employed to extract principal features after using and testing PCA and even kernel PCA [59] on the dataset. SPCA, in our work, takes the high-dimensional data sequences ($x_i - 28D$) from $A^{m\times 28}$, uses eigenvectors of the covariance matrix, does an eigen analysis, and projects the feature vectors on the first $p$ dominant eigen vectors (principal components). Hence, the main idea underlying SPCA is similar to PCA. The difference is that each principal feature in SPCA is a linear combination of just a few (sparse) weighted uncorrelated original features. Finally, SPCA results in the corresponding low-dimensional principal feature vectors ($h_i - pD$) into $B^{m\times P}$. SPCA offers more right solutions under high dimensions than PCA [57]. Eq. (11) shows a low-rank approximation of the feature matrix $A$. SPCA tries to minimize the least square criterion (12) with the sparsity matrix $R^{28\times P}$ and coefficients $r_{ls}$. The second part of (12) is the lasso penalty, which is tuned by metaparameter of $\lambda \geq 0$ [60].

$$A = BR^T + E \tag{11}$$

$$\min_{B,R} \| A - BR^T \|^2 + \lambda \sum_{l=1}^{28} \sum_{s=1}^{p} |r_{ls}| \tag{12}$$

*(iii) Internal SPCA and k-means clustering*

Principal feature vectors ($h_i$) from the output of SPCA are inputted to the k-means clustering block. K-means is one of the most famous/efficient methods employed for clustering goals. In order to have more accurate clusters, firstly, the vectors ($h_i - pD$) are converted to ($h_i' - 2D$) into $C^{m\times 2}$ by an internal SPCA, and then initial centroids ($\mu_j$) are calculated based on the two principal features and using a *k-means++* initialization scheme [61]. The internal SPCA may help to better initialize centroids [62]. The k-means clustering aims to group the vectors $h_i'$ into $K$ clusters. In our case, $K = 2$, since a two-class source detection problem is studied in this paper. Each feature vector is assigned to the cluster with the shortest 'distance' to one of the cluster centers. Centroids are then updated once all feature vectors are assigned. The inertia for minimization by k-means is Euclidean distance (13), and it is trained by alternatively applying the following two steps until convergence (14) and (15):

$$\min \sum_{i=1}^{m} \sum_{j=1}^{2} \omega_{ij} \| h_i' - \mu_j \|^2 \tag{13}$$

**Table 2**
Expression of the employed classifiers in Fig. 1.

| Type | Classifier | Explanation | Main decision function | Ref. | Eq. |
|---|---|---|---|---|---|
| SVM-based | RBF-SVM | SVM is commonly used in binary classification and transfers data to higher n-dimensional space to find an optimal hyperplane to separate classes. The SVs are coordinates of a new n-dimensional system. Thus, SVM uses different kernels. The SVMs' performance is independent of the number of features, but slow training speed does not endorse it for online applications [50]. | $\hat{y} = \text{sign}(\sum_{i=1}^{m} a_i y_i e^{(-\gamma \|x_i - x_j\|^2)} + b),\ \gamma > 0$ | [40] | (3) |
| | Polynomial-SVM | | $\hat{y} = \text{sign}(\sum_{i=1}^{m} a_i y_i (x_i^T x_j + 1)^d + b),\ d > 0$ | [40] | (4) |
| | Linear-SVM | | $\hat{y} = \text{sign}(\sum_{i=1}^{m} a_i y_i x_i^T x_j + b)$ | [40] | (5) |
| | Sigmoid-SVM | | $\hat{y} = \text{sign}(\sum_{i=1}^{m} a_i y_i\ tanh(\beta_1 x_i^T x_j + \beta_2) + b),\ \beta_1 > 0,\ \beta_2 < 0$ | [51] | (6) |
| Tree-based | DT | DT is based on different hierarchical steps that lead to certain decisions. It applies a treelike structure to represent decision paths with induction and pruning steps. In the induction step, the tree structure is built, while, in the pruning step, the complexities of the tree are reduced. The inputs are mapped to outputs by traversing each path through different tree branches. | $G_i = 1 - \sum_{j=1}^{2} v_{i,j}{}^2,\ (i = 1, \ldots, m)$ | [52] | (7) |
| | RF | RF uses several DTs instead of having only one in DT. It can be applied in massive datasets to classify data or measure the importance of each feature in the final decision. The RF is preferred over the DT because it is more accurate and overcomes the overfitting issue of DT. RF is not recommended for real-time purposes because it is generally slower than other models [53]. | $\hat{y} = \frac{1}{n\_tree} \sum_{j=1}^{n\_tree} \sum_{i=1}^{m} W_j(x_i, \acute{x}) y_i$ | [54] | (8) |
| Other | KNN | KNN is one of the most basic used classifiers. It generally finds data with similar characteristics and groups them in the same class without assumptions on data distribution. The groups are constructed by considering the attributes of the neighboring samples. It is used in online data mining and pattern recognition. | $\hat{y} = c_{jmax} = \underset{c_j \epsilon \{0,1\}}{\text{argmax}} \sum_{i=1}^{n\_neighbor} \frac{1}{i} \times I(c_j, y_i), j = 1, 2$ | [55] | (9) |
| | LR | Like SVM, LR is commonly used for binary classification. In LR, the threshold indicates examples labeled into which class using hypothesis and logistic function. The hypothesis determines the likelihood of generating data and fitting them into the logarithm function that forms a sigmoid curve. Then, the function is used to predict the class of new inputs [50]. | $\hat{p} = \delta\left(\sum_{i=1}^{m} \theta_i x_i + b\right),\ \hat{y} = \begin{cases} 0\ if\ \hat{p} < 0.5 \\ 1\ if\ \hat{p} \geq 0.5 \end{cases}$ | [56] | (10) |

$$\omega_{ij} = \begin{cases} 1,\ if\ j = \underset{j}{argmin} \| h_i' - \mu_j \|^2 \\ 0,\ otherwise \end{cases} \quad (14)$$

$$\mu_j = \frac{\sum_{i=1}^{m} \omega_{ij} h_i'}{\sum_{i=1}^{m} \omega_{ij}} \quad (15)$$

where (14) assigns each feature vector $h_i'$ to its closest $\mu_j$, and (15) updates the $\mu_j$ by averaging all feature vectors within the $j$th cluster.

*(iv) 2D visualization of original and clustered features by t-SNE*

To visualize the original vectors ($x_i - 28D$) from $A^{m \times 28}$ (with labels 0/1, which are available from simulations) and the clustered vectors (clustered $h_i' - 2D$) from clustered $C^{m \times 2}$, a nonlinear dimensionality reduction method, t-SNE [63], is used. T-SNE is another embedding method for converting high-dimensional Gaussian distributed feature points into low-dimensional (in our case, two) points in a t-student distribution. In this study, one t-SNE is used. It is once applied to the ($x_i - 28D$) with original labels, and the second time applied to the ($x_i - 28D$) with obtained labels from the proposed model. Hence, theoretically, first, the similarity between two feature vectors $i$ and $j$ belong to spaces of $x_i$ are modeled by $p_{ij}$ and $q_{ij}$ in the input and output of t-SNE, respectively. The mapping from input to output of t-SNE is then obtained by minimizing the KL divergence between those two distributions:

$$\text{KL}(\text{P} \parallel \text{Q}) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (16)$$

In our case, using t-SNE is only for 2D visualization of original feature vectors and clustered principal feature vectors.

To obtain the accuracy of our proposed unsupervised method, a label as 0 or 1 is assigned for some samples in each cluster by checking the existing true labels, which are already existing from simulations. Then, the rest of the samples can have a label. Finally, the optimized models of the proposed unsupervised method are employed for real applications in noiseless and noisy conditions (Fig. 2b). Later on, Table 20 will show the impact of using dimension reduction by SPCA as the first step in the overall proposed schema.

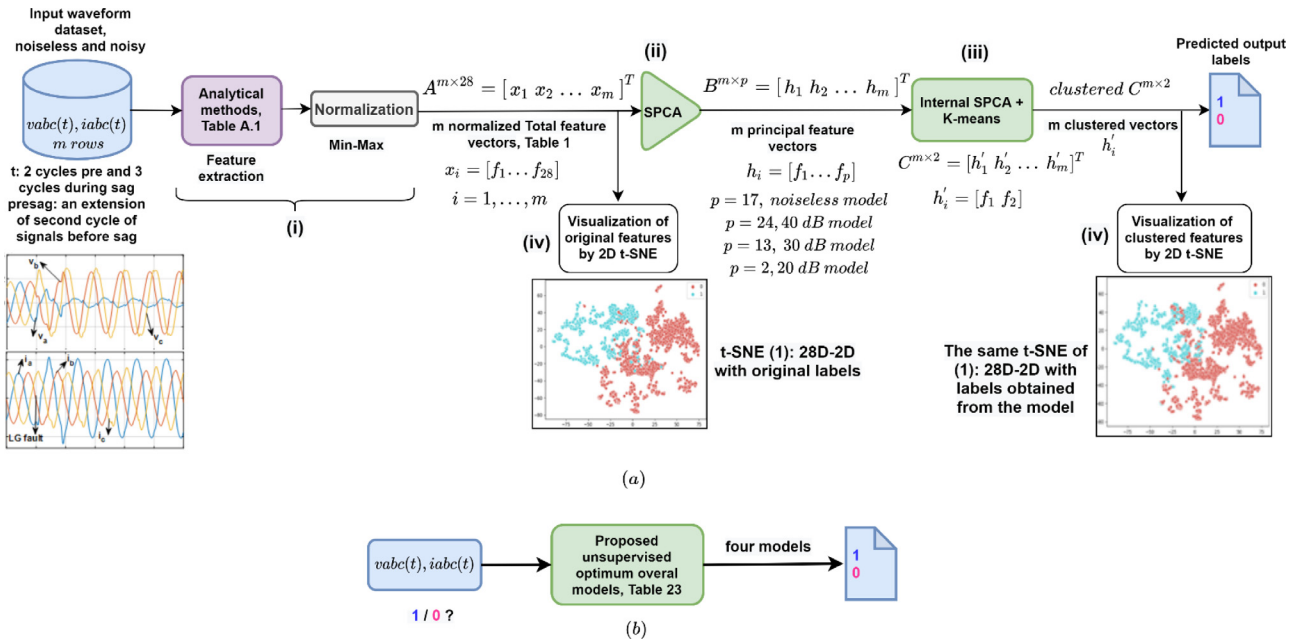


**Fig. 2.** Process of proposed unsupervised method (a) Training the models for noiseless and noisy dataset (b) Implementation of proposed models.

## 3. Performance evaluation

### 3.1. Explanation of training and testing samples

The simulated case study is a regional power system in the Mato Grosso state of Brazil [40], as shown in Fig. 3. This real system is modeled by PSCAD/EMTDC. The PQMs are installed at 6 points (different topologies[1]), with M1 at a 138 kV line, M2…M5 on the boundary between the 230 and 138 kV systems, and M6 between the 138 kV and 13.8 kV system. The downstream direction is the forward side of the power flow direction before voltage sag, and the upstream direction is backward. The training and testing datasets are generated as follows by considering various scenarios. The loads are constant impedance, constant power, and induction motors, and their sizes changed randomly in different simulations. The impedances of lines are also changed randomly during simulations to show the possibility of installing new lines. The duration of faults is set between 0.1 and 0.3 s randomly and is simulated under different system conditions: 15 fault locations (F1–F15), 11 fault types (LLLG, LLL, LG-a, LG-b, LG-c, LLG (LL)-ab, LLG (LL)-bc, LLG (LL)-ca), 5 fault impedances (0.001, 1, 10, 40 and 80 Ω) and 6 PQMs installed over the different typologies of systems (M1-M6). Consequently, $15 \times 11 \times 5 = 825$ fault cases are evaluated in the 6 PQMs, totaling 4950 fault samples. Five TE points are simulated (TE1-TE5) with 7 transformer capacities (20–140 MVA with 20 MVA steps); hence $5 \times 7 \times 6 = 210$ TE samples are also obtained. The voltage sags due to simultaneous occurring two faults, two TEs, or a fault and TE at two downstream and upstream sides are very rare events ([10] and many other works of literature); hence they are not considered in this study. However, a TE followed by two fault cases is shown in Section 4. 5160 noiseless samples result in $m = 4385$ noiseless voltage sag data as voltage and current signals. The noiseless 5160 groups of data are then polluted to white Gaussian noises with SNRs of 40 dB, 30 dB, and 20 dB, which result in $m = 4469$, 4579 and 4783 noisy voltage sag data, respectively. The $m$ voltage and current signals are processed by analytical methods (Table A.1) via MATLAB codes in a sampling frequency of 7.68 kHz to extract the three types of feature matrixes as $A^{m\times 28}$, $A^{m\times 17}$, and $A^{m\times 11}$. Then, $m$ corresponding label as downstream class (1) or upstream class (0) is set. The feature matrixes are then normalized and inputted to intelligent methods in *Python* programming. The different parts of fault samples, leading to voltage sags, for noiseless and noisy datasets are as: LLLG (LLL) (symmetrical samples), LG-a, LG-b, LG-c, LLG-ab, LLG-bc, LLG-ca, LLLG (earth fault samples), LG-a, LG-b, LG-c, LLG (LL)-ab, LLG (LL)-bc, LLG (LL)-ca (asymmetrical samples). The other samples are TE and samples with the downstream and upstream labels. The $N_{total}$ for separated parts of each dataset is given in the next sections, and the accuracy of either developed supervised or proposed unsupervised models is obtained as follows:

$$\text{ACC} = \frac{N_{correct}}{N_{total}} \times 100 \tag{17}$$

### 3.2. Results of developed supervised methods

This section presents the results of developed supervised methods with the predicted outputs as 1 and 0. Different models using 5-fold cross-validation are trained for various inputted features as Total, PB, and IB and for different levels of noise as noiseless, 20, 30, and 40 dB. Table 3 shows the overall classification results of the nine developed methods (12 models for each) and the best analytical method. The accuracies are sorted from highest (dark green) to lowest (light red) in each column, and the highest accuracy and the closest accuracy to the highest one are indicated by ✓and ⍁, respectively. The following conclusions obtain from Table 3:

1- The developed supervised methods, especially non-SVM-based methods (DT, RF, LR, KNN), and the EL show much higher accuracy than the best analytical methods.
2- On average, the developed RF methods are the best for any input features and noise level. Total-RF, Total/PB-KNN,

[1] Note that the PQMs are installed on the border of systems with different network topologies. M1 is at the border of a single source radial network with constant impedance and power loads. M2 is at the border of a two-source radial network due to the presence of a 15MVA DG at the DS side. M3, M4 and M5 are at the border of a two-source interconnected network, and M6 is at the border of a single-source radial network with a large induction machine load (13.8 kV, 3200 HP).

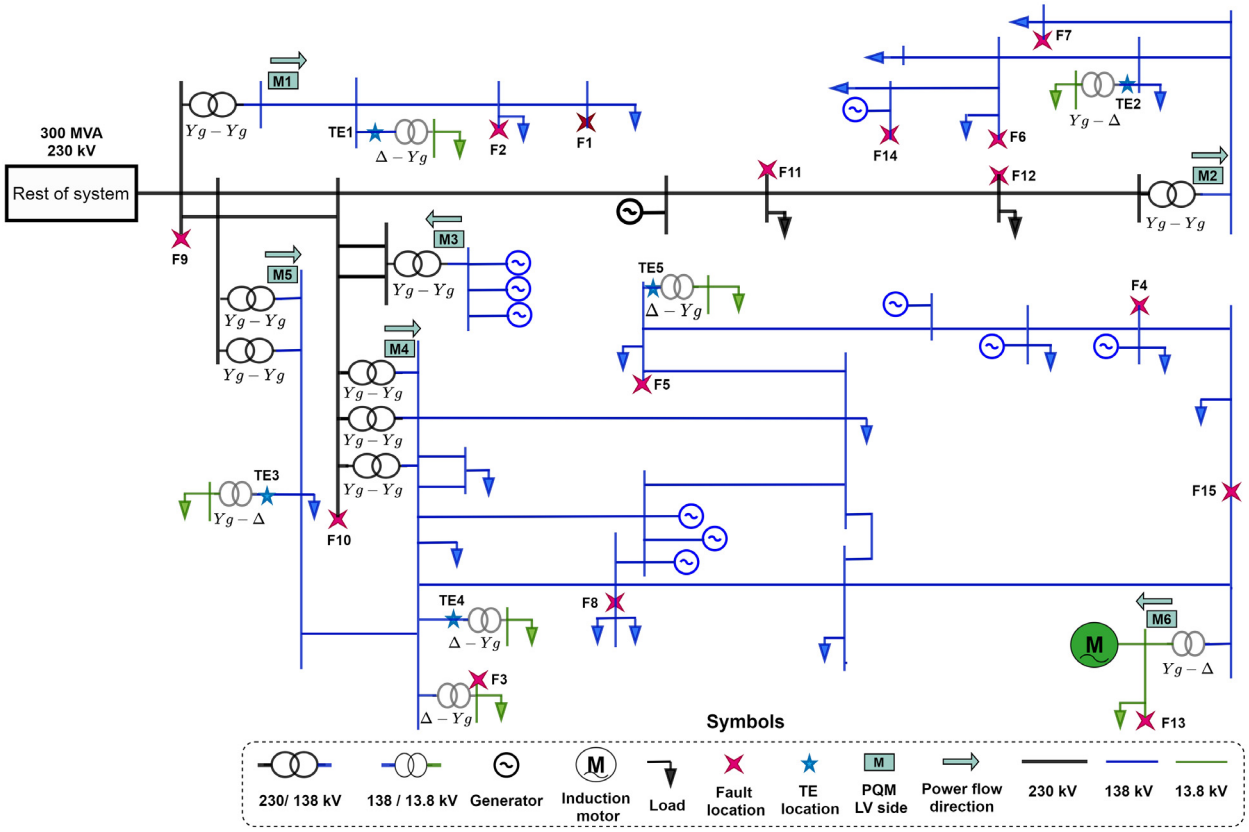


**Fig. 3.** Testing system for numerical simulations of voltage sag samples due to faults and TEs.

and IB-LR models have shown the highest performance for noiseless datasets.

3- Comparison between SVM-based, non-SVM-based, and EL methods shows that tree-based methods (DT, RF), others (LR, KNN), and EL have better overall accuracy (Tables A.2a and A.2b, optimum parameters of best models).
4- On average, the SVM-based models with RBF kernel are more accurate, and models using polynomial kernel are less accurate than the other SVM-based models.
5- Adding noises to pure datasets decreases the classification accuracy for all the best analytical methods and most of the developed supervised models, even by training models individually for each noise level. The impact is much less on the non-SVM-based and EL models, showing their inherent anti-noise characteristics.
6- The best analytical methods using PB features are ahead of those utilizing IB features. This result may change for the short duration sags, sags with transient behavior [30–32], and sags caused by TE events. Hence, the developed supervised methods fed by PB features are more accurate than those fed by IB features. This result is more accurate for non-SVM-based and EL methods. The learning methods fed by Total features show almost similar performance to those fed by PB features. It shows that using only PB features or IB features for sags that are short, transient, and caused by TE events can build a high-performance classifier.
7- The developed EL in this paper included SVM-based and non-SVM-based (tree-based and others) models. The EL had the second score among others for Total and PB features. SVM-based members had a negative effect on the EL accuracy. Therefore, an enhanced EL model could be built by employing only non-SVM-based models. However, the developed models of KNN/RF fed by total or PB features, and LR/RF fed by IB features reached a very high accuracy under noiseless data.

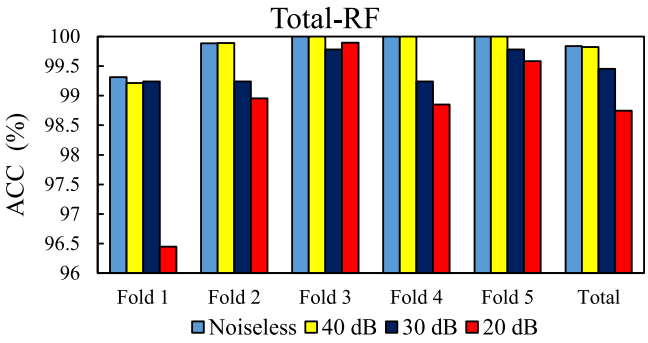


**Fig. 4.** Overall accuracy of developed Total-RF method at different folds.

The performance of developed Total-RF methods (on average, the highest performances) and Total-EL methods for different cross-validation folds are shown in Figs. 4 and 5. The EL methods are very close to the RF method and perform very well for different noise levels.

In order to check in depth, the performance of the developed supervised methods against the best analytical ones, Table 4 gives the results in terms of $N_{correct}$ and ACC for the noiseless $A^{m\times 28}$ related to separated samples of datasets. Again, the non-SVM-based and EL methods show much higher accuracy than the best analytical and SVM-based methods. Excellent accuracy is obtained for analytical methods of RS (16) (100%) and RP (6) (99.39%) for upstream (class 0) and TE samples, respectively. However, the developed EL and KNN methods show 100% accuracy for class 0 samples, and all non-SVM-based and EL methods

**Table 3**
Classification overall accuracy (%) of developed supervised and best analytical methods under noiseless and noisy data fed by Total, PB, and IB features.

| | Method | Total features | | | | PB features | | | | IB features | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Noiseless | 40 dB | 30 dB | 20 dB | Noiseless | 40 dB | 30 dB | 20 dB | Noiseless | 40 dB | 30 dB | 20 dB |
| 1 | Best analytical | 93,09[1] | 91,77[2] | 89,28[2] | 84,61[3] | 93,09[1] | 91,77[2] | 89,28[2] | 84,61[3] | 90,54[4] | 87,54[6] | 84,56[5] | 81,12[5] |
| 2 | Sigmoid-SVM | 96,44 | 91,99 | 92,25 | 79,57 | 93,89 | 94,94 | 80,20 | 78,10 | 93,84 | 92,19 | 92,10 | 84,13 |
| 3 | Polynomial-SVM | 98,77 | 99,22 | 64,48 | 86,99 | 96,74 | 99,40 | 92,06 | 68,28 | 90,81 | 63,59 | 83,45 | 62,74 |
| 4 | Linear-SVM | 98,88 | 98,10 | 98,46 | 90,09 | 98,95 | 98,25 | 94,19 | 90,69 | 92,22 | 95,32 | 88,73 | 73,65 |
| 5 | RBF-SVM | 98,95 | 98,46 | 97,59 | 92,85 | 99,13 | 98,93 | 97,09 | 95,65 | 96,62 | 96,62 | 89,69 | 87,93 |
| 6 | LR | 99,54 | 99,37 | 98,54 | 96,34 | 99,57 | 99,53 | 98,06 | 96,05 | 99,34✓ | 97,99 | 95,32 | 89,59 |
| 7 | EL | 99,64 | 99,66 | 99,09 | 96,47 | 99,66 | 99,71 | 98,93 | 97,09 | 97,06 | 97,02 | 92,84 | 88,18 |
| 8 | DT | 99,66 | 98,72 | 98,38 | 97,62 | 99,16 | 98,68 | 98,78 | 97,51 | 98,02 | 96,58 | 95,72 | 90,38 |
| 9 | RF | 99,84✓ | 99,82✓ | 99,46✓ | 98,81✓ | 99,79☑ | 99,89✓ | 99,52✓ | 98,66✓ | 99,22☑ | 98,95✓ | 97,59✓ | 93,96✓ |
| 10 | KNN | 99,84✓ | 99,15 | 98,61 | 96,05 | 99,82✓ | 99,66 | 98,76 | 96,70 | 94,82 | 94,12 | 91,34 | 85,01 |

1: RS (16); 2: RCC (8); 3: CBM (20); 4: RS (18); 5: DPE (3); 6: CBM (21); Colors: accuracies sorted from highest (dark green) to lowest (light red); ✓: Best accuracy for each column, ⧄: Closest accuracy to highest one by tolerance of (–0.1%).

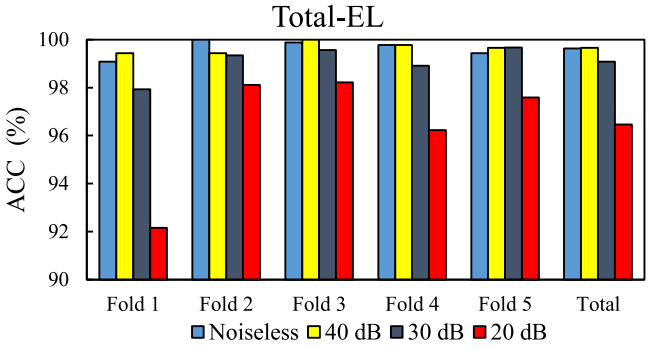


**Fig. 5.** Overall accuracy of developed Total-EL method at different folds.

are ahead of the analytical method of RP (6) for TE samples. The weakest accuracy of analytical methods was assigned to PCSC (22) (91.96%) for downstream (class 1) samples, while the developed RF method gave 99.68%. The same overall accuracy of 99.84% was obtained by developed RF and KNN methods and 99.63% for the proposed EL, which is higher than the used DT-EL method in [40] by overall accuracy of 99.2%, even by employing optimal features. The developed RF method performs better for symmetrical, earth fault, class 1, and TE samples, whereas the KNN method is more accurate for the asymmetrical and class 0 samples.

In the following, the performance ($N_{correct}$ and accuracy) of best PB/IB analytical and developed supervised methods (shown by ✓in Table 3 as PB-KNN and IB-LR (noiseless), PB/IB-RF (all noise levels)) related to separated different parts of samples are distinctly given in Tables 5–8 for noiseless, 40 dB, 30 dB, and 20 dB conditions, respectively. The findings from these tables are as follows:

- ***For PB methods***: The analytical methods had the best accuracy for noiseless and 40 dB earth faults and symmetrical 30 and 20 dB samples, whereas the KNN and RF methods showed much more accuracy. The weakest accuracy of analytical methods was related to TE samples, while a very high accuracy of KNN and RF was obtained for such samples.
- ***For IB methods***: The analytical methods had the best accuracy for TE samples, i.e., sags due to TEs, which are a good candidate for IB methods. It is because of the second-order harmonic of current signals on the TE events. It was expected to see such results for the LR and RF methods. However, they showed accuracy even less than analytical methods (except for 20 dB data), so that the weakest accuracy of KNN and RF was obtained for the noiseless and 40 dB TE samples, respectively. While a low accuracy of analytical methods was for overall samples (noiseless, 30 and 20 dB) and 40 dB earth faults, the LR and RF methods showed considerably more accuracy in the cases.
- ***Cases in which IB methods are more accurate than PB***: TEs regarding analytical methods and only symmetrical noiseless samples: IB-LR (highest accuracy (99,73%)), PB-KNN (99.6%).

Confusion matrixes are shown in Tables 9–12 for noiseless, 40 dB, 30 dB, and 20 dB conditions, respectively. This way, the best PB/IB analytical and developed learning methods can be compared for both classes of 0 and 1. The diagonal entries of the confusion matrix show the number of correctly predicted instances in each class, whereas the non-diagonal entries show the incorrect classifications.

We use the CR and FAR [64] for each class of 1 and 0 separately to measure the performance of the methods mentioned above for each class separately, besides their overall accuracy. For each class, the CR and FAR are defined as follows:

$$CR = \frac{TP}{TP + FN} \times 100 \tag{18}$$

$$FAR = \frac{FP}{FP + TN} \times 100 \tag{19}$$

CR can be used for measuring the performance of each method between two classes, while for measuring the performance of two different methods, both CR and FAR must be checked. Table 13 gives the performances of the best PB/IB analytical and developed supervised methods (methods as shown in Tables 9–12) for each class in terms of CR and FAR, and it is concluded that:

- Best methods among analytical ones for classes 0 and 1: PB methods: RS (16) under noiseless data, RCC (8) under 40 and 30 dB noises, and CBM (20) for 20 dB data.
- Best methods among all for classes 0 and 1: PB developed supervised: KNN under noiseless data, and RF for noisy data.
- Performance of methods for the classes separately: $CR_{class\ 1} < CR_{class\ 0}$ for analytical, and $CR_{class\ 1} \leq CR_{class\ 0}$ for supervised methods.
- Although some analytical methods have high CR close to the developed supervised methods, the much less FAR seen for developed supervised makes them more accurate; for instance, a 0.5% FAR for PB-KNN and 1.44% for IB-LR under noiseless signals.

### 3.3. Results of proposed unsupervised method

The results of the proposed unsupervised method on the noiseless and noisy $A^{m\times 28}$ are explained in this section. Table 14

**Table 4**
Performance of developed supervised methods and best analytical ones fed by Total features under noiseless separated samples of sags.

| Method | Performance | Sym. | Asym. | Earth faults | class 1 | class 0 | TE | Overall |
|---|---|---|---|---|---|---|---|---|
| ----- | $N_{total}$ | 754 | 3468 | 2671 | 1593 | 2792 | 163 | 4385 |
| Best analytical | $N_{correct}$ | 702[1] | 3231[1] | 2491[2] | 1465[3] | 2792[1]✓ | 162[4] | 4082[1] |
| | ACC (%) | 93,1[1] | 93,17[1] | 93,26[2] | 91,96[3] | 100[1]✓ | 99,39[4] | 93,09[1] |
| Sigmoid-SVM | $N_{correct}$ | 703 | 3373 | 2584 | 1499 | 2730 | 153 | 4229 |
| | ACC (%) | 92,24 | 97,26 | 96,74 | 94,09 | 97,78 | 93,86 | 96,44 |
| Polynomial-SVM | $N_{correct}$ | 748 | 3431 | 2635 | 1543 | 2788 | 152 | 4331 |
| | ACC (%) | 99,2 | 98,93 | 98,65 | 96,86 | 99,85 | 93,25 | 98,77 |
| Linear-SVM | $N_{correct}$ | 749 | 3433 | 2638 | 1547 | 2789 | 154 | 4336 |
| | ACC (%) | 99,34 | 98,99 | 98,76 | 97,11 | 99,89 | 94,47 | 98,88 |
| RBF-SVM | $N_{correct}$ | 745 | 3436 | 2654 | 1555 | 2784 | 158 | 4339 |
| | ACC (%) | 98,8 | 99,08 | 99,36 | 97,61 | 99,71 | 96,93 | 98,95 |
| LR | $N_{correct}$ | 747 | 3455 | 2665 | 1585 | 2780 | 163✓ | 4365 |
| | ACC (%) | 99,07 | 99,62 | 99,77 | 99,49 | 99,57 | 100✓ | 99,54 |
| EL | $N_{correct}$ | 752 | 3459 | 2664 | 1577 | 2792✓ | 158 | 4369 |
| | ACC (%) | 99,73 | 99,74 | 99,73 | 98,99 | 100✓ | 96,93 | 99,63 |
| DT | $N_{correct}$ | 750 | 3457 | 2661 | 1583 | 2787 | 163✓ | 4370 |
| | ACC (%) | 99,47 | 99,68 | 99,62 | 99,37 | 99,82 | 100✓ | 99,65 |
| RF | $N_{correct}$ | 754✓ | 3461☑ | 2669✓ | 1588✓ | 2790 | 163✓ | 4378✓ |
| | ACC (%) | 100✓ | 99,79☑ | 99,92✓ | 99,68✓ | 99,92 | 100✓ | 99,84✓ |
| KNN | $N_{correct}$ | 752 | 3463✓ | 2667☑ | 1586☑ | 2792✓ | 163✓ | 4378✓ |
| | ACC (%) | 99,73 | 99,86✓ | 99,85☑ | 99,56☑ | 100✓ | 100✓ | 99,84✓ |

1: RS (16); 2: RCC (8); 3: PCSC (22); 4: RP (6); Colors: ACC/$N_{correct}$ sorted from highest (dark green) to lowest (light red); ✓: Best ACC for each column, ⧄: Closest ACC to highest one by tolerance of (−0.12%).

**Table 5**
Performance of best PB/IB analytical and developed supervised methods under noiseless separated samples of sags.

| Noiseless sample type | $N_{total}$ | PB features | | | | IB features | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Best analytical | | KNN | | Best analytical | | LR | |
| | | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) |
| Sym. | 754 | RS(16)-702 | 93,1 | 751 | 99,6 | CBM(21)-699 | 92,71 | 752 | 99,73 |
| Asym. | 3468 | RS(16)-3231 | 93,17 | 3463 | 99,86 | RS(18)-3144 | 90,66 | 3446 | 99,37 |
| Earth faults | 2671 | RCC(8)-2491 | 93,26 | 2666 | 99,81 | RCC(9)-2433 | 91,09 | 2657 | 99,48 |
| TE | 163 | RCC(7)-151 | 92,64 | 163 | 100 | RP(6)-162 | 99,39 | 158 | 96,93 |
| Overall | 4385 | RS(16)-4082 | 93,09 | 4378 | 99,82 | RS(18)-3970 | 90,54 | 4356 | 99,34 |

Colors: ACC sorted from highest (dark green) to lowest (light red).

**Table 6**
Performance of best PB/IB analytical and developed supervised methods under 40 dB separated samples of sags.

| 40 dB sample type | $N_{total}$ | PB features | | | | IB features | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Best analytical | | RF | | Best analytical | | RF | |
| | | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) |
| Sym. | 803 | RS(16)-740 | 92,15 | 803 | 100 | CBM(21)-733 | 91,28 | 798 | 99,38 |
| Asym. | 3491 | RCC(8)-3225 | 92,38 | 3486 | 99,86 | CBM(21)-3040 | 87,08 | 3458 | 99,05 |
| Earth faults | 2708 | RCC(8)-2509 | 92,65 | 2706 | 99,93 | DPE(3)-2344 | 86,56 | 2685 | 99,15 |
| TE | 175 | PCSC(22)153 | 87,43 | 175 | 100 | RP(6)-175 | 100 | 166 | 94,86 |
| Overall | 4469 | RCC(8)-4101 | 91,77 | 4464 | 99,89 | CBM(21)-3912 | 87,54 | 4422 | 98,95 |

Colors: ACC sorted from highest (dark green) to lowest (light red).

**Table 7**
Performance of best PB/IB analytical and developed supervised methods under 30 dB separated samples of sags.

| 30 dB sample type | $N_{total}$ | PB features | | | | IB features | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Best analytical | | RF | | Best analytical | | RF | |
| | | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) |
| Sym. | 837 | CBM(20)-760 | 90,8 | 836 | 99,88 | CBM(21)-734 | 87,69 | 828 | 98,92 |
| Asym. | 3560 | RCC(8)-3205 | 90,03 | 3540 | 99,44 | DPE(3)-3034 | 85,22 | 3463 | 97,28 |
| Earth faults | 2771 | RCC(8)-2495 | 90,04 | 2755 | 99,42 | DPE(3)-2386 | 86,11 | 2702 | 97,51 |
| TE | 200 | PCSC(22)-175 | 87,5 | 199 | 99,5 | RP(6)-198 | 99 | 195 | 97,5 |
| Overall | 4579 | RCC(8)-4088 | 89,28 | 4575 | 99,52 | DPE(3)-3872 | 84,56 | 4486 | 97,59 |

Colors: ACC sorted from highest (dark green) to lowest (light red).

shows the optimal parameters of the proposed overall model, required time, and accuracy of known classes. A label was assigned for some samples in each cluster by checking the existing true labels obtained from the simulations, and then all samples were labeled. The method's accuracy can be calculated by having input and obtained labels from our method. Increasing the noise level

**Table 8**
Performance of best PB/IB analytical and developed supervised methods under 20 dB separated samples of sags.

| 20 dB sample type | $N_{total}$ | PB features | | | | IB features | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Best analytical | | RF | | Best analytical | | RF | |
| | | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) | $N_{correct}$ | ACC (%) |
| Sym. | 860 | CBM(20)-757 | 88,02 | 854 | 99,3 | CBM(21)-703 | 81,74 | 824 | 95,81 |
| Asym. | 3714 | CBM(20)-3135 | 84,41 | 3659 | 98,52 | DPE(3)-3036 | 81,74 | 3479 | 93,67 |
| Earth faults | 2885 | CBM(20)-2423 | 83,99 | 2838 | 98,37 | DPE(3)-2392 | 82,91 | 2705 | 93,76 |
| TE | 208 | PCSC(22)-173 | 83,17 | 205 | 98,56 | RP(6)-189 | 90,87 | 190 | 91,35 |
| Overall | 4782 | CBM(20)-4046 | 84,61 | 4718 | 98,66 | DPE(3)-3879 | 81,12 | 4493 | 93,96 |

Colors: ACC sorted from highest (dark green) to lowest (light red).

**Table 9**
Confusion matrix for best PB/IB analytical and developed supervised methods under noiseless data.

| PB-RS(16) | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2792 | 0 |
| | 1 | 303 | 1290 |

(a)

| PB-KNN | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2792 | 0 |
| | 1 | 7 | 1586 |

(b)

| IB-RS(18) | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2751 | 41 |
| | 1 | 374 | 1219 |

(c)

| IB-LR | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2786 | 6 |
| | 1 | 23 | 1570 |

(d)

Colors: sorted from lowest value (light blue) to highest value (dark blue).

**Table 10**
Confusion matrix for best PB/IB analytical and developed supervised methods under 40 dB data.

| PB-RCC(8) | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2783 | 59 |
| | 1 | 309 | 1318 |

(a)

| PB-RF | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2840 | 2 |
| | 1 | 3 | 1624 |

(b)

| IB-CBM(21) | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2750 | 92 |
| | 1 | 465 | 1162 |

(c)

| IB-RF | | Predicted class 0 | Predicted class 1 |
|---|---|---|---|
| True class | 0 | 2825 | 17 |
| | 1 | 30 | 1597 |

(d)

Colors: sorted from lowest value (light blue) to highest value (dark blue).

decreases the accuracy of known classes while the training times of SPCA and clustering are increased. The noiseless model showed maximum accuracy of 97.17% with a minimum time of 0.013 s To check the efficiency of the proposed methods, function 2D t-SNE was used to visualize two classes (clusters), whose parameters were set as Barnes–Hut algorithm, Euclidean distance metrics, perplexity = 30. The best 2D embedding space for visualization was selected by selecting the minimum loss values from running t-SNE 100 times. The t-SNE was applied on the original input feature vectors with original labels and obtained labels (clustered) from the proposed methods as shown in Figs. 6–9 for noiseless, 40 dB, 30 dB, and 20 dB datasets, respectively. Figs. 6b, 7b, 8b, and 9b show that the proposed methods work well on clustering and labeling feature vectors. Some feature vectors that the proposed methods have failed are indicated by dashed circles, where the number of failing cases increases by increasing the noise level.

The proposed unsupervised methods have been compared in Table 15 with the best analytical and SVM-based supervised methods (the ones investigated in Section 3.2). Checking the overall accuracy shows that the proposed unsupervised methods show accuracy higher than the best analytical methods for any level of noise, so for the noiseless dataset, high accuracy of 97.17% was obtained. Moreover, the proposed methods show a very close accuracy to SVM-based supervised methods, even higher than sigmoid-SVM methods for any noise level and polynomial-SVM methods for 30 and 20 dB noises. It is worth mentioning that our unsupervised method does not need labeled training data compared to all supervised learnings. Since the input labels and labels obtained from the method are available (an unsupervised method by having the labels but not using them for training) in our study, the confusion matrixes are also given in Table 16 to show the efficiency of unsupervised methods for each class 0 and 1. In order to measure the performances, CR and FAR were calculated and given in Table 17. Unlike the analytical and supervised methods, which were more accurate for classifying the voltage sags belonging to class 0, the proposed unsupervised methods show better CR for class 1 samples except for 30 dB datasets. Although the analytical methods have CR higher than the proposed methods, the much lower FAR of unsupervised methods makes them more accurate, such as 1.57% for the noiseless model of class 0 compared to 19% for the best analytical method.

## 4. Field-testing results

This section introduces field testing developed to verify the performance of the ML methods to locate voltage sag sources. From Table 3, on average, the RF method showed the best overall accuracy for noisy and noiseless voltage sags. Hence, the results of the noiseless RF method (Total, PB, and IB features) (Fig. 1b), the proposed unsupervised method (Fig. 2b), and the best analytical

**Table 11**
Confusion matrix for best PB/IB analytical and developed supervised methods under 30 dB data.

| PB-RCC(8) | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2760 | 168 |
| | 1 | 325 | 1344 |

(a)

| PB-RF | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2923 | 5 |
| | 1 | 17 | 1652 |

(b)

| IB-DPE(3) | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2861 | 67 |
| | 1 | 643 | 102 |

(c)

| IB-RF | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2887 | 41 |
| | 1 | 70 | 1599 |

(d)

Colors: sorted from lowest value (light blue) to highest value (dark blue).

**Table 12**
Confusion matrix for best PB/IB analytical and developed supervised methods under 20 dB data.

| PB-CBM(20) | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2852 | 148 |
| | 1 | 588 | 1194 |

(a)

| PB-RF | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2984 | 16 |
| | 1 | 48 | 1734 |

(b)

| IB-DPE(3) | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2794 | 206 |
| | 1 | 697 | 1085 |

(c)

| IB-RF | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2909 | 91 |
| | 1 | 200 | 1582 |

(d)

Colors: sorted from lowest value (light blue) to highest value (dark blue).

ones are shown here. As given in Table 18, 22 field tests were applied to the methods. Cases 1–10 are related to a Slovenian power system at different voltage levels with different sampling frequencies, while cases 11–20 belong to a 20 kV Slovenian distribution network by monitors M1-M4 installed in a loop topology (Fig. 10). In order to highlight the capability of the developed RF and proposed unsupervised methods, tests 2, 9, 17, and 22 of the field-testing results are shown in Fig. 11 (current and voltage measurements attached by true classes) and Table 19 (extracted features and methods' output). From Table 19, while the best analytical methods of PB-RS (16) and IB-RS (18) failed

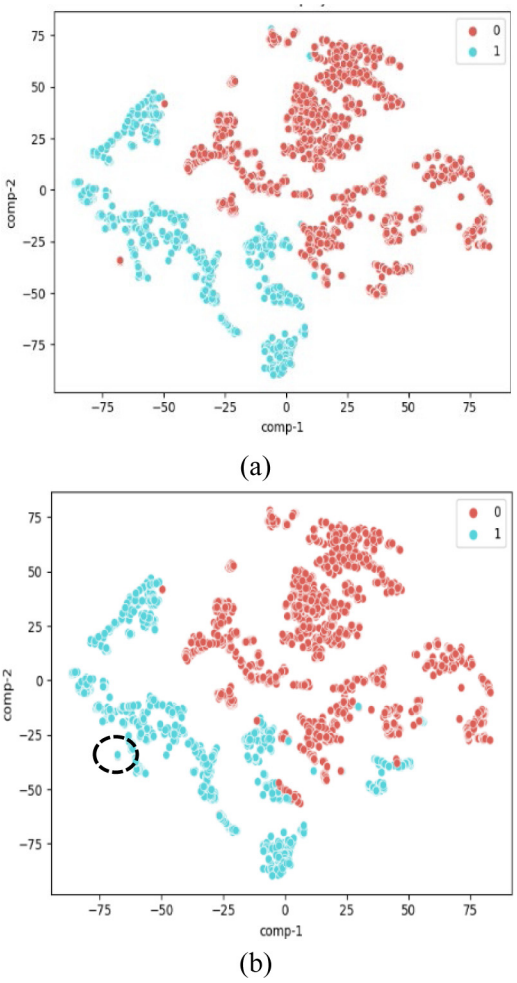


**Fig. 6.** Visualization by t-SNE for noiseless dataset, unknown class (dashed circles) (a) Original features; (b) Clustered features.

in two cases, the ML methods correctly classified the sag source's location.

All field-testing results obtained by the discussed methods are given in Table 20. Features with wrong performances from the viewpoint of analytical methods are shown in the 3rd column. As expected, the developed supervised Total-RF method shows the best performance, 22/22 cases (no incorrect results). The IB-RF method also shows no incorrect results when compared to IB-RF, which failed in two cases. It shows the importance of obtaining supervised methods separately for PB and IB features. The accuracy of three considered scenarios is shown for the proposed unsupervised method. The overall model (last column) has shown a good performance, 20/22 (90,9%) better than PB-RS (16), 18/22 (81,8%) and PB-RS (18), 19/22 (86,4%) as the best analytical methods. Using only k-means clustering (7th column) or even adding an internal SPCA (8th column) in knowing true classes has given the correct numbers of only 8 and 16 of 22 cases, respectively, which confirms the effectiveness of the proposed unsupervised overall model using an SPCA following a k-means clustering with initializing due to an internal SPCA. This shows the importance of using SPCA in our proposed schema. The results of developed supervised RF and proposed unsupervised methods for cases 11 to 22 related to a loop topology show that the methods may be a good candidate for looped networks.

**Table 13**
Performance of best PB/IB analytical and developed supervised methods on each class: (CR, FAR) in (%).

| Feature type | Best method | Noiseless | | 40 dB | | 30 dB | | 20 dB | |
|---|---|---|---|---|---|---|---|---|---|
| | | Class 0 | Class 1 | Class 0 | Class 1 | Class 0 | Class 1 | Class 0 | Class 1 |
| PB | Analytical | (100, 19.02) | (80.97, 0) | (97.09, 19) | (81, 2.07) | (94.26, 19.47) | (80.52, 5.73) | (95.07, 33) | (67, 4.93) |
| | Dev. supervised | (100, 0.5) | (99.5, 0) | (99.93, 0.18) | (99.81, 0.07) | (99.83, 1.02) | (98.98, 0.17) | (99.46, 2.7) | (97.03, 0.53) |
| IB | Analytical | (98.53, 23.48) | (76,5, 1.47) | (96.76, 28.58) | (71.42, 3.24) | (97.71, 38.5) | (61.47, 2.29) | (93.13, 39.11) | (60.88, 6.86) |
| | Dev. supervised | (99.78, 1.44) | (98.55, 0.21) | (99.4, 1.84) | (98.15, 0.6) | (98.56, 4.2) | (95.08, 1.04) | (96.97, 11.22) | (88.78, 3.03) |

**Table 14**
Parameters used in the proposed unsupervised overall models, required time, and known class accuracy under Total input features.

| Noise level | Inertia[a] | Homogeneity[b] | Completeness[c] | V-measure[d] | ARI[e] | AMI[f] | Silhouette[g] | L1[h] | Required total time (s) | Known class ACC (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Noiseless | 63 324 | 0.592 | 0.567 | 0.579 | 0.656 | 0.579 | 0.232 | 1/10, 1 | 0.013 | 97, 17 |
| 40 dB | 81 101 | 0.608 | 0.587 | 0.597 | 0.691 | 0.597 | 0.151 | 1/15, 1 | 0.023 | 95, 82 |
| 30 dB | 51 655 | 0.273 | 0.267 | 0.270 | 0.205 | 0.270 | 0.147 | 1/12, 1 | 0.024 | 92, 54 |
| 20 dB | 7446 | 0.149 | 0.305 | 0.201 | 0.165 | 0.200 | 0.683 | 1/10, 1 | 0.027 | 89, 13 |

[a]Within-cluster sum of squares criterion, a measure of how internally coherent clusters are (low values are better).
[b]A score to show each cluster contains only members of a single class.
[c]A score to show all members of a given class are assigned to the same cluster.
[d]A score which shows the harmonic mean of 2 and 3.
[e]Adjusted rand index.
[f]Adjusted mutual information.
[g]A measure of how close each point in one cluster is to points in the neighboring clusters.
[h]Regularization coefficient for SPCAs.

**Table 15**
A comparison between overall accuracy (%) of supervised SVM-based, proposed unsupervised, and best analytical methods fed by Total features.

| Type | Method | Noiseless | 40 dB | 30 dB | 20 dB |
|---|---|---|---|---|---|
| Analytical | Best | 93,09[a] | 91,77[b] | 89,28[b] | 84,61[c] |
| Supervised learning | Sigmoid-SVM | 96,44 | 91,99 | 92,25 | 79,57 |
| | Polynomial-SVM | 98,77 | 99,22 | 64,48 | 86,99 |
| | Linear-SVM | 98,88 | 98,10 | 98,46 | 90,09 |
| | RBF-SVM | 98,95 | 98,46 | 97,59 | 92,85 |
| Prop. unsupervised learning | Overall model | 97,17 | 95,82 | 92,54 | 89,13 |

[a]RS (16).
[b]RCC (8).
[c]CBM (20).

**Table 16**
Confusion matrix for unsupervised proposed overall method fed by Total features (a) noiseless; (b) 40 dB; (c) 30 dB; (d) 20 dB.

| | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2693 | 99 |
| | 1 | 25 | 1568 |

(a)

| | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2714 | 128 |
| | 1 | 59 | 1568 |

(b)

| | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2743 | 185 |
| | 1 | 158 | 1511 |

(c)

| | | Predicted class | |
|---|---|---|---|
| | | 0 | 1 |
| True class | 0 | 2535 | 465 |
| | 1 | 55 | 1727 |

(d)

Colors: sorted from lowest value (light blue) to highest value (dark blue).

**Table 17**
Performance of best analytical methods and proposed unsupervised methods in each class: (CR, FAR) in (%).

| Method | Noiseless | | 40 dB | | 30 dB | | 20 dB | |
|---|---|---|---|---|---|---|---|---|
| | Class 0 | Class 1 | Class 0 | Class 1 | Class 0 | Class 1 | Class 0 | Class 1 |
| Best analytical | (100, 19.02)[a] | (80.97, 0)[a] | (97.09, 19)[b] | (81, 2.07)[b] | (94.26, 19.47)[b] | (80.52, 5.73)[b] | (95.07, 33)[c] | (37, 4.93)[c] |
| Prop. unsupervised learning | (96.24, 1.57) | (98.43, 3.54) | (95.49, 3.62) | (96.37, 4.5) | (93.68, 9.46) | (90.53, 6.32) | (84.5, 3.08) | (96.91, 15.5) |

(,): (CR, FAR).
[a]RS (16).
[b]RCC (8).
[c]CBM (20).

## 5. Discussion

### 5.1. Comparing the methods and a recommendation for choosing practical methods

According to the existing features and the SNR obtained from measurement devices, the best-developed supervised models (shown by a tick in Table 3) can be recommended for real applications. The candidate methods for overall voltage sag cases are as follows:

- ***For Total features***: KNN (99.84%)/RF (99.84%)-noiseless, and RF-(99.82%) 20, (99.46%) 30 and (99.81%) 40 dB. However, the KNN is more recommended for real-time applications than RF (Table 2).

**Table 18**
Field-testing cases.

| Test no. | Sag source | True class | Monitor (kV) | Sampling freq. (kHz) | Vsag (pu) | Sag duration (s) |
|---|---|---|---|---|---|---|
| 1 | LG-a | 1 | 20 | 5 | 0.22 | 0.3 |
| 2 | LG-a | 1 | 20 | 10 | 0.34 | 0.22 |
| 3 | LL-bc | 1 | 20 | 10 | 0.9 | 0.16 |
| 4 | LLL | 1 | 20 | 10 | 0.9 | 0.15 |
| 5 | LG-a | 1 | 400 | 1 | 0.65 | 0.3 |
| 6 | LG-c | 0 | 220 | 1 | 0.35 | 0.06 |
| 7 | LLL | 0 | 110 | 1 | 0.07 | 0.07 |
| 8 | LL-ac | 0 | 110 | 6.4 | 0.82 | 0.06 |
| 9 | LL-ab + LLL | 1 | 20 | 6.4 | 0.02 | 0.04 + 0.03 |
| 10 | TE | 1 | 20 | 1.6 | 0.8 | 0.06 |
| 11 | LG-a | 1 | M1–20 | 1 | 0.1 | 0.08 |
| 12 | LG-a | 1 | M1–20 | 1 | 0.09 | 0.08 |
| 13 | LL-ac | 1 | M1–20 | 1 | 0.64 | 0.08 |
| 14 | LG-a | 0 | M2–20 | 1 | 0.15 | 0.08 |
| 15 | LG-a | 0 | M2–20 | 1 | 0.05 | 0.08 |
| 16 | LL-ac | 0 | M2–20 | 1 | 0.47 | 0.08 |
| 17 | LG-c | 0 | M3–20 | 1 | 0.11 | 0.08 |
| 18 | LG-c | 0 | M3–20 | 1 | 0.05 | 0.08 |
| 19 | LL-bc | 0 | M3–20 | 1 | 0.49 | 0.08 |
| 20 | LG-c | 0 | M4–20 | 1 | 0.07 | 0.08 |
| 21 | LG-c | 0 | M4–20 | 1 | 0.06 | 0.08 |
| 22 | LL-bc | 0 | M4–20 | 1 | 0.5 | 0.08 |

Case 9: a transformer energizing developed by fault; Cases 11…22: test cases concerning Fig. 10; Cases 11, 14, 17, 20: substation transformer is grounded by Petersen coil; Cases 12, 15, 18, 21: substation transformer is grounded by 80-ohm resistance.

**Table 19a**
Field-testing features ($f_1$−$f_{15}$).

| Test no. | DPE(1:3) | | | RP(4:6) | | | | | | RCC(7:9) | | | SST(10:12) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | $f_{12}$ | $f_{13}$ | $f_{14}$ | $f_{15}$ |
| 2 | 0,011 | 0,024 | 0,007 | 0,091 | 0,223 | **−0,050** | 0,118 | **−0,001** | **−0,009** | 0,083 | 0,057 | 0,002 | −0,689 | **0,005** | **0,015** |
| 9 | 0,337 | 0,443 | 0,045 | **−29,512** | 1096 | 20,027 | 4103 | 1081 | 1167 | 6413 | 1533 | 0,036 | −0,011 | −0,008 | −0,011 |
| 17 | −0,021 | −0,023 | −0,009 | −0,285 | **0,748** | −0,621 | −0,063 | −0,010 | **0,128** | −0,011 | −0,002 | −0,0001 | **−0,015** | **−0,292** | **−0,033** |
| 22 | −2,640 | −2,405 | −0,531 | −82,034 | **0,615** | −54,143 | **0,802** | −0,949 | **0,034** | −0,009 | −0,496 | −0,009 | **−0,004** | **−0,002** | **−0,001** |

Underlined bold text: wrong performance of a feature.

**Table 19b**
Field-testing features ($f_{16}$−$f_{28}$).

| Test no. | DR(13:15) | | | | | RS(16:19) | | | | CBM(20,21)-PCSC(22,23) | | | | RS (16) | RS (18) | T-RF | P-RF | I-RF | Unsup. method |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $f_{16}$ | $f_{17}$ | $f_{18}$ | $f_{19}$ | $f_{20}$ | $f_{21}$ | $f_{22}$ | $f_{23}$ | $f_{24}$ | $f_{25}$ | $f_{26}$ | $f_{27}$ | $f_{28}$ | | | | | | |
| 2 | −0,27 | 0,12 | **−0,09** | −0,02 | **−0,01** | **0,04** | **0,001** | 0,04 | **0,001** | 0,056 | **0,110** | 0,002 | **0,01** | 0✘ | 1 | 1 | 1 | 1 | 1 |
| 9 | −1,82 | 1,12 | **−0,45** | −0,11 | 0,08 | **0,07** | **0,070** | **−0,01** | −0,077 | 6068 | **0,255** | 0,151 | −0,01 | 0✘ | 0✘ | 1 | 1 | 1 | 1 |
| 17 | 0,58 | −0,05 | −0,75 | **−0,03** | −0,02 | 0,42 | 0,053 | −0,05 | **−0,020** | −0,002 | 0,762 | **0,001** | **−0,01** | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | **−1,08** | −2,46 | −3,07 | **−0,05** | −0,03 | 0,03 | 0,002 | **4,26** | **−0,011** | **0,614** | 2,995 | **0,020** | 0,02 | 0 | 1✘ | 0 | 0 | 0 | 0 |

Underlined bold text: wrong performance of a feature; ✘: wrong performance of a method; T: Total: P: PB; I: IB; Unsup.: Unsupervised.

- ***For PB features***: KNN (99.82%)/RF (99.79%)-noiseless, and noisy RF-(99.80%) 20, (99.52%) 30, and (98.66%) 40 dB. However, for real-time applications, the KNN is more recommended than RF.
- ***For IB features***: LR (99.34%)/RF (99.22%)-noiseless, and RF-(98.95%) 20, (97.59%) 30, and (93.96%) 40 dB. However, the LR is more recommended than RF for real-time application.

Therefore, regardless of the sag sources, input feature subset, and noise levels, the RF models have the best performance from the accuracy perspective. However, since the features are time-consuming in practice, a reasonable model is PB-RF, which uses 17 input features.

Moreover, a practical and intelligent model applicable for different SNRs can be an EL model utilizing the best models in noiseless and noisy conditions (Tables A.2a and A.2b) to classify the sag source location as follows:

- ***For Total features***: An EL using learners of noiseless KNN/RF and noisy RF (20 and 40 dB) (can be a future work).
- ***For PB features***: An EL using learners of noiseless KNN/RF and noisy RF (20, 30, and 40 dB) (can be a future work).
- ***For IB features***: Only a noiseless RF model or an EL using learners of noiseless RF and LR is enough.

As said, since the RF models fed by all different types of features were selected as highly effective for noisy and noiseless signals; hence the RF models may be used in all scenarios mentioned above.

The same methodology can be used for the four proposed unsupervised models. Hence, an EL model using the noiseless and noisy models (Table 14) can be used as a practical model (can be future work).

**Table 20**
Field-testing results for noiseless analytical, developed supervised and proposed unsupervised methods.

| Test no. | True class | Features with wrong performance | Best analytical | | Best developed supervised learning | | | Proposed unsupervised method | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | RS (16) | RS (18) | RF | RF | RF | Only k-means | Internal SPCA + k-means | Overall model |
| | | | PB-$f_{21}$ | IB-$f_{23}$ | Total | PB | IB | Total | Total | Total |
| 1 | 1 | [$f_4$–$f_8$, $f_{17}$, $f_{21}$, $f_{22}$] | ✘ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 2[a] | 1 | [$f_6$, $f_8$, $f_9$, $f_{14}$, $f_{15}$, $f_{18}$, $f_{20}$ – $f_{22}$, $f_{24}$, $f_{26}$, $f_{28}$] | ✘ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 3 | 1 | – | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 4 | 1 | [$f_{24}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 5 | 1 | [$f_1$, $f_{22}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✘ | ✓ |
| 6 | 0 | [$f_7$, $f_9$, $f_{13}$–$f_{16}$, $f_{19}$, $f_{25}$, $f_{27}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ |
| 7 | 0 | [$f_5$, $f_7$, $f_9$, $f_{13}$–$f_{16}$, $f_{19}$, $f_{27}$, $f_{27}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ |
| 8 | 0 | [$f_{13}$–$f_{16}$, $f_{19}$, $f_{25}$, $f_{27}$] | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ | ✓ | ✓ |
| 9[a] | 1 | [$f_4$, $f_{18}$, $f_{21}$–$f_{23}$, $f_{26}$] | ✘ | ✘ | ✓ | ✓ | ✓ | ✘ | ✘ | ✓ |
| 10 | 1 | [$f_{10}$, $f_{24}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 11 | 1 | [$f_4$, $f_5$, $f_9$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✘ | ✓ |
| 12 | 1 | [$f_4$, $f_5$, $f_{14}$, $f_{18}$, $f_{21}$, $f_{22}$, $f_{26}$] | ✘ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 13 | 1 | – | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 14 | 0 | [$f_5$, $f_7$, $f_9$, $f_{17}$, $f_{19}$, $f_{20}$, $f_{22}$, $f_{24}$, $f_{27}$, $f_{28}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 15 | 0 | [$f_4$, $f_{10}$, $f_{13}$, $f_{16}$–$f_{19}$, $f_{27}$, $f_{28}$] | ✓ | ✓ | ✓ | ✘ | ✓ | ✘ | ✘ | ✘ |
| 16 | 0 | [$f_5$, $f_7$, $f_{13}$–$f_{16}$, $f_{18}$, $f_{19}$, $f_{25}$–$f_{28}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 17[a] | 0 | [$f_5$, $f_9$, $f_{13}$–$f_{15}$, $f_{19}$, $f_{24}$, $f_{27}$, $f_{28}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 18 | 0 | [$f_5$, $f_7$, $f_{10}$, $f_{13}$–$f_{16}$, $f_{18}$, $f_{19}$, $f_{22}$, $f_{25}$–$f_{28}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 19 | 0 | [$f_5$, $f_7$, $f_{10}$, $f_{13}$–$f_{16}$, $f_{19}$, $f_{24}$, $f_{25}$, $f_{27}$, $f_{28}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| 20 | 0 | [$f_5$, $f_{13}$–$f_{15}$, $f_{19}$, $f_{23}$, $f_{24}$, $f_{27}$] | ✓ | ✘ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 21 | 0 | [$f_7$, $f_{13}$, $f_{16}$, $f_{18}$, $f_{19}$, $f_{25}$–$f_{28}$] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✘ |
| 22[a] | 0 | [$f_5$, $f_7$, $f_9$, $f_{13}$–$f_{16}$, $f_{19}$, $f_{24}$, $f_{25}$, $f_{27}$] | ✓ | ✘ | ✓ | ✓ | ✓ | ✘ | ✓ | ✓ |
| Correctly classified out of 22 | | | 18 (81,8%) | 19 (86.4%) | 22 (100%) | 20 (90,9%) | 22 (100%) | 8 (36,4%) | 16 (72,7%) | 20 (90,9%) |

✘: incorrect; ✓: correct.
[a]Examples shown in Fig. 11, Tables 19a and 19b.

### 5.2. Number of clusters in the proposed unsupervised method

The k-means clustering method requires the user to select the number of clusters in advance. In our study, $K$ is pre-determined as 2, and Table 14 gives the inertia and silhouette score for this number of $K$. This is because the investigated problem in this study is the finding relative location of the sag sources as upstream (class/label 0) and downstream (class/label 1). Although the authors have checked the $K = 3$ and the accuracy of our proposed unsupervised method reached about 98.5%; however, having three groups of labels for the problem investigated in this study is meaningless.

### 5.3. Impact of changing the number of folds in the cross-validation process of the learning methods

In order to check the performance of developed supervised methods for different numbers of training and testing data, the number of folds was changed to 3, 5, and 10 in the cross-validation process for noiseless input data of $A^{m\times 28}$. Fig. 12 shows that increasing the number of folds from 3 to 5 improves the performance of methods. But some of the methods show a similar or a bit lower accuracy for the increase from 5 to 10. On average, the impact of the number of folds is as follows:

- For SVM-based and non-SVM-based methods: $ACC_{3-fold} < ACC_{10-fold} \leq ACC_{5-fold}$
- For EL method: $ACC_{3-fold} \approx ACC_{5-fold} \approx ACC_{10-fold}$

Therefore, the 5-fold cross-validation was selected in the training process of the developed methods in this paper.

### 5.4. Relation between input features to the methods

The Pearson Correlation Coefficients (cross-correlation) between all 28 input features of noiseless $A^{m\times 28}$ is shown in Fig. 13. As shown in this figure, maximum correlations are only between {$f_1$, $f_2$, $f_3$} (0.99), {$f_4$, $f_6$} (0.93), {$f_{25}$, $f_{27}$} (0.99), {$f_{11}$, $f_{12}$, $f_{23}$} ($\approx$ 1, the values are rounded), {$f_{18}$, $f_{26}$} (−0.94), and {$f_{20}$, $f_{28}$} (−0.91). However, three different types of input feature sets are investigated in this study, and one of the obtained models can be used according to the availability of the features. Moreover, there is a difference between the highly correlated features. For example, $f_1$ is a PB feature obtained from line phasors, while $f_2$ is calculated from positive-sequence phasors. Meanwhile, $f_3$ is an IB feature based on positive-sequence instantaneous values. Feature $f_{18}$ needs voltage and current measurements, while $f_{26}$ is based only on current information.

The authors of this paper used the optimal feature selection in [40]. Five optimal features were obtained as [$f_3$/$f_6$, $f_{17}$, $f_{21}$, $f_{26}$, $f_{27}$], applied to an EL method with DT learners, which reached an accuracy of a maximum of 99.2% for a dataset of only 672 samples of voltage sags due to just short-circuit faults (and not TEs). All the optimal features were PB except $f_{27}$.

Although, a feature selection might be made between the highly correlated features in practical cases, in this work, we tried to investigate some other aspects of applied features depending on their availability in reality, in this work, we tried to investigate some other aspects of applied features depending on their availability in reality. The subsets were Total, PB and IB. Even with employing a bigger set of features, the accuracy of the developed supervised methods was higher than that [40] for noiseless conditions:

***For Total features***: RF/KNN (99.84%)
***For PB features***: KNN (99.82%)
***For IB features***: LR (99.34%)

This confirms that no redundant features were being used in this study and that learning models considered some weights for the features. Nevertheless, an optimal feature selection can be formed between the highly correlated features in practical cases. For the proposed unsupervised method, an SPCA reduced the Total feature size from 28 and generated the principal features

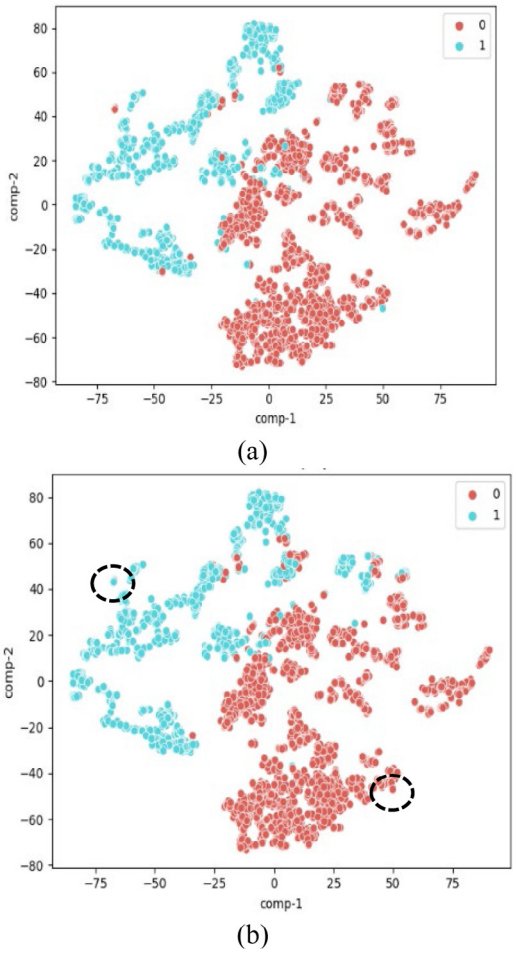


**Fig. 7.** Visualization by t-SNE for 40 dB dataset, unknown class (dashed circles) (a) Original features; (b) Clustered features.

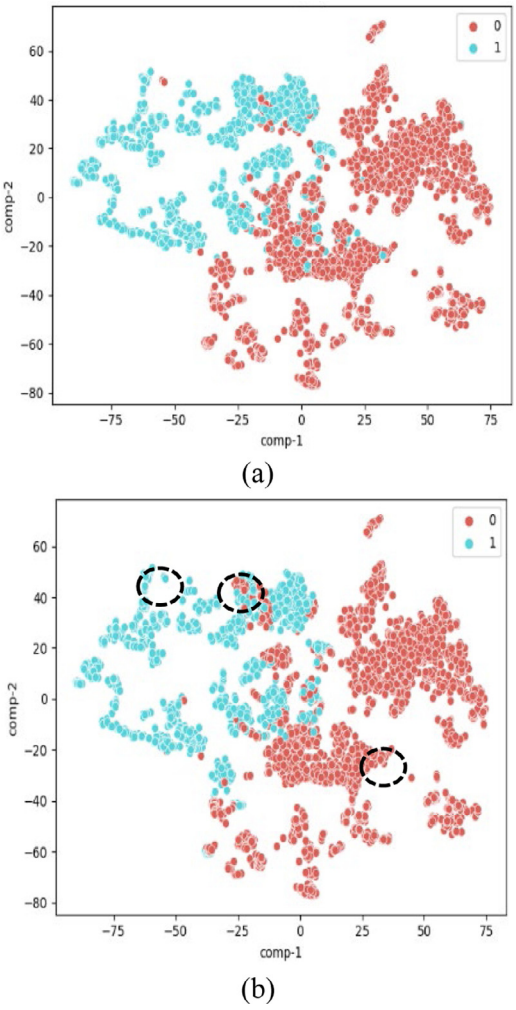


**Fig. 8.** Visualization by t-SNE for 30 dB dataset, unknown class (dashed circles) (a) Original features; (b) Clustered features.

with the numbers 17, 24, 13, and 2 for the noiseless, 40 dB, 30 dB, and 20 dB, respectively.

### 5.5. Impact of real data on the methods

In a real case, voltage and current signals may present noises, monitoring system errors, and underdetermined changes due to inconstant fault conditions. Therefore, some of the input features may be affected. Hence, in this paper, the noise effect was highly considered by adding white noises, and the different fault cases were applied on many buses over the network. However, making synthetic noises to show the real behavior of existing noises in real waveforms [65] and applying the faults on the lines could also be considered a future work to make the methods more and more practical. Nevertheless, since the intelligent strategies utilize all the advantages of input features, the accuracy of the intelligent strategies will be higher than that of the analytical methods. The case study used in this research is also well-known in the literature [3,6,7,18,19,21,32,39,40] by updated parameters, and the dataset extracted from the real system is credible.

### 5.6. Effect of changing characteristics of the case study on the methods

In general, training data should cover all complexities of the problem and any change in the parameters by considering various scenarios; therefore, the trained models can handle new unknown cases. However, in this paper, while generating the datasets, it has been tried to consider all the possible scenarios of a power system, such as changing the load sizes, line impedances, different typologies of a system (PQMs installed at different locations), different combinations of faults, TEs, etc. On the other hand, the 28 features explaining source location considered in this paper have two characteristics of sign and magnitude to classify the sag source location. The effect of sign characteristics (+/-) was highly reflected compared to another characteristic during the data normalization step. The changes in the case study parameters will affect the magnitude characteristic of the features only and not the sign characteristic. Therefore, the changes in the system parameters are not expected to affect the performance of the models.

### 5.7. Testing the models in a different power system

The tuned models of the developed supervised methods and the proposed unsupervised one are employed for real applications (part b of Figs. 1 and 2). In order to obtain a similar accuracy in a new system, a retuning of models' parameters (developed Total/PB/IB-RF and proposed unsupervised methods) is recommended. Nevertheless, the obtained models should remain unchanged. Also, as a key point, to avoid re-training the models in another power system, the active learning and fine-tuning of

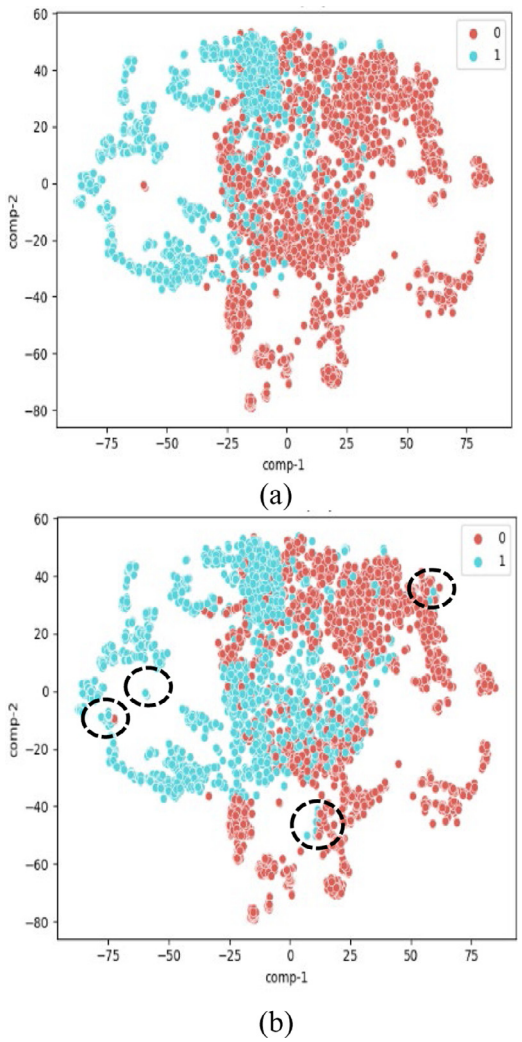


**Fig. 9.** Visualization by t-SNE for 20 dB dataset, unknown class (dashed circles) (a) Original features; (b) Clustered features.

pre-trained models based on certain new samples, which may be selected either randomly or by the active learning strategies, would be effective.

The methods studied in this paper are capable of the case of multi-stage sags like case 9 in Table 18 and Fig. 11b and f. Each of the needed features can be calculated by registering a voltage sag in each stage and considering differences between variables before and after the sag. Since the best-developed supervised methods and the proposed unsupervised method have shown high accuracy against both fault and TE events, the simultaneous events on one side (downstream/upstream) will be detected by the methods. This is because the harmonic current due to TE will be riding on the fault current but still showing the downstream/upstream features needed for the methods. However, since the input features of intelligent methods come from analytical methods, the analytical methods would already fail in the failing cases of intelligent methods. By the way, we have considered a TE and fault as two different samples in the datasets.

### 5.8. Speed of the ML methods

In terms of accuracy, the developed supervised methods and the proposed unsupervised one (utilizing a subset of features) showed higher accuracy when compared to analytical methods (utilizing one/two features).

Like many kinds of literature, the dataset used in this study (considering many scenarios) is trained offline. When the parameters of models are tuned, they would be employed for real applications (Part b of Figs. 1 and 2). So, the training phase of the learning methods would not affect the speed of the methods when used in online/real-time applications. However, the methods are based on a group of scaler's phasor or instantaneous (PB and IB) features extracted from analytical methods for detecting voltage sag sources. One of the features, for instance, was the magnitude changes of positive sequence current phasor ($f_{25}$). Nevertheless, preparing the feature subsets might be time-consuming for the online application of the methods for the protection goals (Section 1.3, application of the methods), when they would work as a directional function in relays or fault indicators (backup protection process always has some delays, so, the high/low speed of the proposed methods is not an issue, here). Although, the features are only using some simple formulations based on voltage and current

Hence, two recommendations are given to increase the speed of the methods:

1- Using the proper features extracted from the transient period of voltage sags [49,66] (half/one cycle after starting sags).
2- Developing the intelligent methods directly, in which time sampled inputs can be directly applied to them. The inputs may include time sampled root mean square (RMS) values of line voltages and currents, adding the time sampled angle of voltages and currents, considering time sampled positive sequence (phasor or instantaneous) values, etc. This idea is more applicable for the time-sampled phasors obtained from the measurements of PMUs. This is the future research that will be investigated in our subsequent work.

As a step further, the good performance of our unsupervised method indicates that it may be possible to train an unsupervised method using only recorded (unlabeled) waveforms. We believe thousands of sag recordings available worldwide could be used for our unsupervised model training to achieve high performance.

## 6. Conclusion

An evaluation of existing phasor-based analytical methods for detecting the sag sources' location showed a maximum accuracy of 93% and 84% for noiseless and high-level noise signals, respectively. The accuracy for instantaneous-based methods was about 90% and 80% in the cases. Hence, in this paper, firstly supervised ML methods as SVM-based, tree-based (DT and RF), others (KNN and LR), and a proposed EL by using PB, IB, and Total features were developed for noiseless and noisy input features (12 models/each of nine methods). Among the nine methods, the best models were obtained as noiseless KNN/RF and noisy RF fed by Total or PB features and a noiseless RF/LR fed by IB features for any noise level. Then, a new unsupervised method was proposed using an SPCA applied to a k-means clustering (which used an internal initializing SPCA) to handle unlabeled datasets. The results of best developed supervised methods (a, b, and c below) and proposed unsupervised method (d below) through extensive simulated voltage sags due to fault and transformer energizing on a Brazilian regional network showed:

(a) Total/PB-KNN/RF by overall accuracy of about 99.8%-noiseless and Total/PB-RF by 98.7%-high-level noises; (b) IB-LR/RF by the accuracy of about 99.2%-noiseless and 94%-high-level noises; (c) Low false alarm rate for PB-KNN (0.5%) and IB-LR

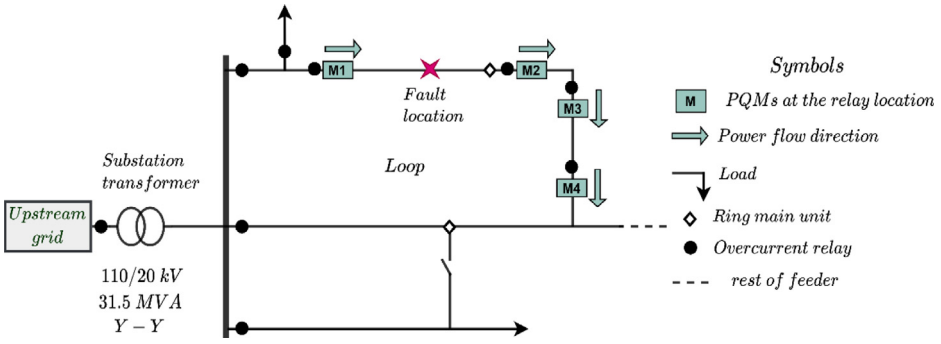


**Fig. 10.** Part of 20 kV Slovenian distribution network (Electrical distribution company at Celje).

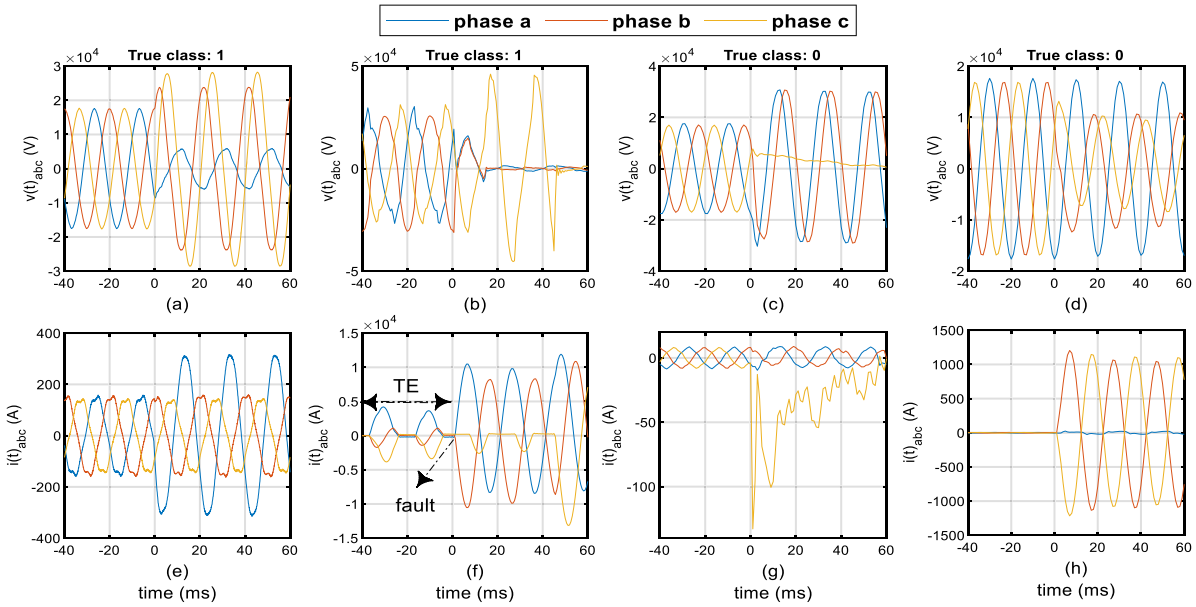


**Fig. 11.** Line voltages and currents for field testing (a), (e) Test 2, a sag of 0.34 pu due to downstream LG-a fault; (b), (f) Test 9, a sag of 0.02 pu due to DS LL-ab fault (c), (g) Test 17, a sag of 0.11 pu due to a US LG-c (d), (h) Test 22, a sag of 0.5 pu due to a US LL-bc fault.

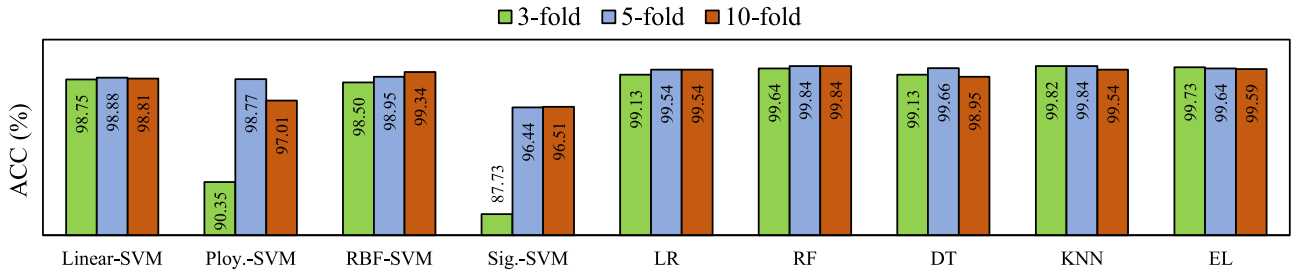


**Fig. 12.** Overall accuracy (%) of Total developed supervised methods for different the number of folds of the CV under a noiseless dataset.

(1.44%) under noiseless signals, and a high classification rate for PB-RF (99.46%) and IB-RF (96.97%) under high-level noises, all for an upstream class (which has more sag samples); (d) Overall accuracy of 99.17%-noiseless and 89.1%-high-level noises, which are higher than analytical methods, very close to SVM-based methods, and it does not need labeled training data, besides showing a classification rate of 98.43% and 96.91 in downstream class much higher than analytical methods.

Moreover, the results of Slovenian field measurements confirmed the efficiency of best developed supervised and proposed unsupervised methods. However, re-training our unsupervised method using sag records worldwide can make it high-performance using no label for the datasets. A practical model can be applying an ensemble learning on the obtained noiseless and noisy RF supervised models as well as on the concluded noiseless and noisy unsupervised models.

The obtained methods can be used in directional overcurrent relays and fault indicators of medium voltage networks. They can also be used to assign penalties to distribution networks with high penetration of distributed energy resources propagating voltage sags into transmission systems, which is in the interest of transmission operators.

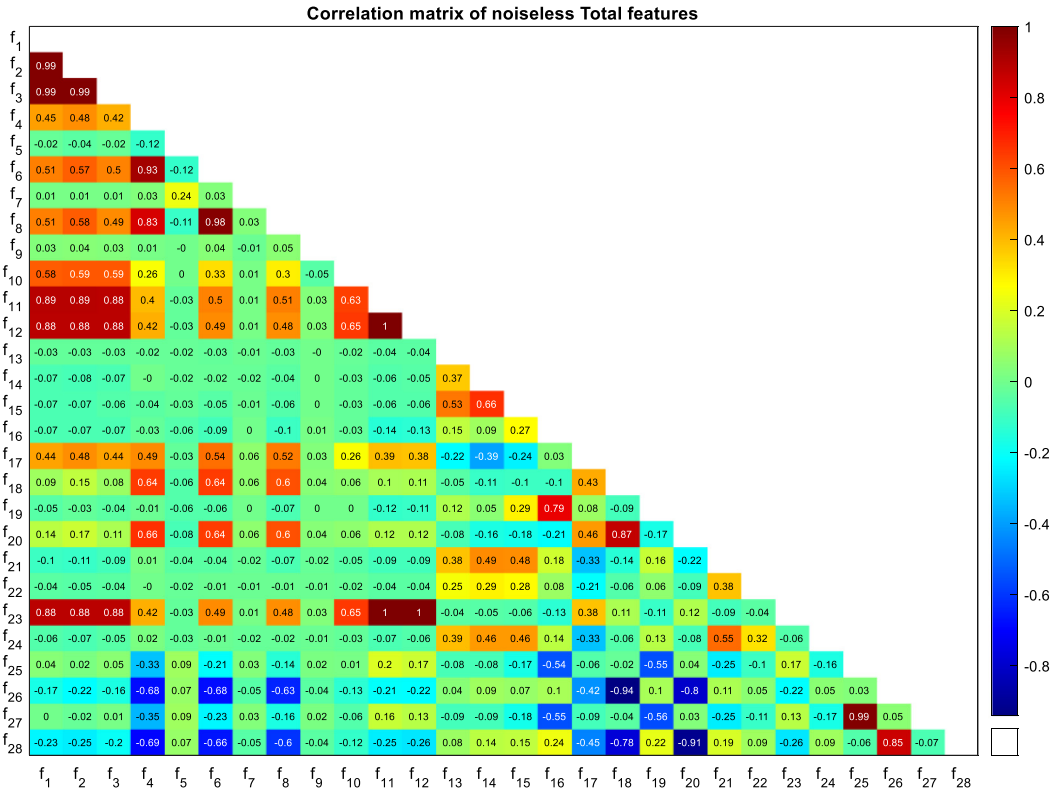


**Fig. 13.** Correlation matrix of noiseless input Total features, −0 and 0 mean a very small negative and positive correlation (somehow no correlation).

**Table A.1**
Description of the analytical methods for voltage-sag source location classification.

| Method | Category | [Ref.] | Basic rule for class 1 (else 0) | Method | Category | [Ref.] | Basic rule for class 1 (else 0) |
|---|---|---|---|---|---|---|---|
| DPE (1) | PB | [5] | $f_1 > 0$ | DR (13) | PB | [12] | $f_{16} < 0$ and $f_{17} > 0$ |
| DPE (2) | PB | [27] | $f_2 > 0$ | DR (14) | PB | [13] | $f_{16} < 0$ and $f_{18} > 0$ |
| DPE (3) | IB | [32] | $f_3 > 0$ | DR (15) | IB | [32] | $f_{19} < 0$ and $f_{20} > 0$ |
| RP (4) | PB | [6] | $f_4 > 0$ and $f_5 > 0$ | RS (16) | PB | [10] | $f_{21} < 0$ |
| RP (5) | PB | [21] | $f_6 > 0$ and $f_7 > 0$ | RS (17) | IB | [3] | $f_{22} < 0$ |
| RP (6) | IB | [32] | $f_8 > 0$ and $f_9 > 0$ | RS (18) | IB | [28,29] | $f_{23} > 0$ |
| RCC (7) | PB | [26] | $f_{10} > 0$ | RS (19) | PB | [11] | $f_{24} < 0$ |
| RCC (8) | PB | [27] | $f_{11} > 0$ | CBM (20) | PB | [21] | $f_{25} > 0$ and $f_{26} < 0$ |
| RCC (9) | IB | [32] | $f_{12} > 0$ | CBM (21) | IB | [31] | $f_{27} > 0$ and $f_{28} < 0$ |
| SST (10) | PB | [26] | $f_{13} < 0$ | PCSC (22) | PB | [20] | $f_{26} < 0$ |
| SST (11) | PB | [27] | $f_{14} < 0$ | PCSC (23) | IB | [31] | $f_{28} < 0$ |
| SST (12) | IB | [32] | $f_{15} < 0$ | | | | |

## CRediT authorship contribution statement

**Younes Mohammadi:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Revise R1. **Seyed Mahdi Miraftabzadeh:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing, Revise R1. **Math H.J. Bollen:** Investigation, Writing – review & editing. **Michela Longo:** Investigation, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

The authors acknowledge Prof. Boštjan Polajžer, head of the institute of electrical power engineering at the University of Maribor, Maribor, Slovenia, for providing the field measurements in this research.

## Appendix

### A.1. Description of analytical methods (feature extractor methods)

The description of the 23 analytical methods (feature extractor methods) related to Table 1 is tabulated in Table A.1. The features are extracted for an interval between half and three cycles after sag starting.

### A.2. Optimum parameters of best supervised models

The optimum values of various parameters of best supervised models indicated by tick in Table 3 are given in Tables A.2a and A.2b. Noiseless and noisy similar models are placed in one column together (highlighted by gray).

**Table A.2a**
Optimum parameters of supervised RF method under noiseless and noisy data fed by Total, PB, and IB features.

| Parameters | | Total features | | | PB features | | | | IB features |
|---|---|---|---|---|---|---|---|---|---|
| | | Noiseless, 30 dB | 20 dB | 40 dB | Noiseless | 20 dB | 30 dB | 40 dB | Noiseless, 20 dB, 30 dB, 40 dB |
| max_features[1] | | log2 | sqrt | sqrt | sqrt | log2 | sqrt | sqrt | log2 |
| n_estimators[2] | | 100 | 100 | 1000 | 10 | 100 | 1000 | 100 | 100 |
| For any type of models | Bootstrap[3]/ccp_alpha[4]/class_weight[5] | True/0/None | | | | | | | |
| | Criterion[6]/max_depth[7]/max_leaf_nodes[8] | gini/None/None | | | | | | | |
| | max_samples[9]/min_impurity_decrease[10]/min_impurity_split[11] | None/0/None | | | | | | | |
| | min_samples_leaf[12]/min_samples_split[13]/min_weight_fraction_leaf[14] | 1/2/0 | | | | | | | |
| | oob_score[15]/verbose[16]/warm_start[17] | False/0/False | | | | | | | |

1: a function on the number of features when looking for the best split; 2: the number of trees in the forest; 3: true when bootstrap samples are used to make trees and false when whole dataset is used; 4: Complexity parameter used for Minimal Cost-Complexity Pruning; 5: weight for label of each class; 6: function to measure the quality of a split; 7: maximum depth of the tree; 8: grow tress with this value in best-first fashion; 9: number of samples to draw from input samples to train each base estimator; 10: a node will be split if this split induces a decrease of the impurity greater than or equal to this value; 11: Threshold for early stopping in tree growth; 12: minimum number of samples required to be at a leaf node; 13: minimum number of samples required to split an internal node; 14: minimum wight fraction of the sum total of weights of all input samples to be at a leaf node; 15: Whether to use out-of-bag samples to estimate the generalization score; 16: controls the verbosity when fitting and predicting; 17: reusing or erasing the solution of previous call to fit initialization.

**Table A.2b**
Optimum parameters of supervised methods of Total-KNN, PB-KNN and IB-LR under noiseless data.

| Method | Parameters | Value |
|---|---|---|
| Total-KNN | Algorithm[a]/leaf size[b]/metric[c]/metric_params[d]/n_neighbors[d]/p[f]/weights[g] | Auto/30/Manhattan/None/2/2/distance |
| PB-KNN | Algorithm/leaf_size/metric/metric/params/n_neighbors/p/wights | Auto/30/Manhattan/None/5/2/distance |
| IB-LR | C[h]/class_weight[i]/dual[j]/fit_intercept[k]/intercept_scaling[l] | 1 000 000/None/False/True/1 |
| | l1_ratio[m]/max_iter[n]/Penalty[o] | None/100/l2 |
| | solver[p]/tol[q]/verbose[r]/warm_start[s] | newton-cg/0.0001/0/False |

[a]Computes the nearest neighbors.
[b]Speed of structure and memory used to store the tree.
[c]Distance metric to use for the tree.
[d]Additional arguments for the metric function number of nearest neighbors to classify a sample.
[f]Power parameter for used metric.
[g]A weight for the points in each neighborhood.
[h]Inverse of regularization strength.
[i]Weight for label of each class.
[j]Primal formulation which is false when number of samples are higher than features.
[k]Specify if a constant should be added to the decision function.
[l]Constant value for the feature used in 11.
[m]Elastic-net mixing parameter.
[n]Maximum number of iterations taken for the solvers to converge.
[o]Specify the norm used in the penalization.
[p]Algorithm to use in the optimization problem.
[q]Tolerance for stopping criteria.
[r]Controls the verbosity when fitting and predicting.
[s]Reusing or erasing the solution of previous call to fit initialization.

## References


[1] M.H.J. Bollen, I.Y.H. Gu, Signal processing of power quality disturbances, 2005, http://dx.doi.org/10.1002/0471931314.
[2] M. Bollen, Understanding Power Quality Problems: Voltage Sags & Interruptions, IEEE Press, 2000.
[3] Y. Mohammadi, M.H. Moradi, R. Chouhy Leborgne, Locating the source of voltage sags: Full review, introduction of generalized methods and numerical simulations, Renew. Sustain. Energy Rev. 77 (2017) 821–844, http://dx.doi.org/10.1016/J.RSER.2017.04.017.
[4] M.V. Costa, N.B. Pereira, J.M.C. Filho, R.C. Leborgne, A novel methodology for determining the voltage sag impact factor, Electr. Power Syst. Res. 174 (2019) 105865, http://dx.doi.org/10.1016/j.epsr.2019.105865.
[5] A.C. Parsons, W.M. Grady, E.J. Powers, J.C. Soward, A direction finder for power quality disturbances based upon disturbance power and energy, IEEE Trans. Power Deliv. 15 (2000) 1081–1086, http://dx.doi.org/10.1109/61.871378.
[6] R.C. Leborgne, R. Makaliki, Voltage sag source location at grid interconnections: a case study in the Zambian system, in: 2007 IEEE Lausanne Power Tech, 2007, pp. 1852–1857, http://dx.doi.org/10.1109/PCT.2007.4538599.
[7] F.O. Passos, J.M. de Carvalho Filho, R.C. Leborgne, P.M. da Silveira, P.F. Ribeiro, An alternative approach to locating voltage sag source side at the point of common coupling based on power-flow information, J. Control Autom. Electr. Syst. 26 (2015) 579–587, http://dx.doi.org/10.1007/s40313-015-0199-x.
[8] N. Hamzah, A. Mohamed, A. Hussain, A new approach to locate the voltage sag source using real current component, Electr. Power Syst. Res. 72 (2004) 113–123, http://dx.doi.org/10.1016/j.epsr.2004.03.010.
[9] C. Li, T. Tayjasanant, W. Xu, X. Liu, Method for voltage-sag-source detection by investigating slope of the system trajectory, Gener. Transm. Distrib. IEE Proc.- 150 (2003) 367–372, http://dx.doi.org/10.1049/ip-gtd:20030214.
[10] T. Tayjasanant, C. Li, W. Xu, A resistance sign-based method for voltage sag source detection, IEEE Trans. Power Deliv. 20 (2005) 2544–2551, http://dx.doi.org/10.1109/TPWRD.2005.852341.
[11] W. Kanokbannakorn, T. Saengsuwan, S. Sirisukprasert, Unbalanced voltage sag source location identification based on superimposed quantities and negative sequence, in: 8th Electr. Eng. Electron. Comput. Telecommun. Inf. Technol. Assoc. Thail. - Conf. 2011, 2011, pp. 617–620, http://dx.doi.org/10.1109/ECTICON.2011.5947915.
[12] A.K. Pradhan, A. Routray, Applying distance relay for voltage sag source detection, IEEE Trans. Power Deliv. 20 (2005) 529–531, http://dx.doi.org/10.1109/TPWRD.2004.839741.
[13] Z. Shao, J. Peng, J. Kang, Locating voltage sag source with impedance measurement, in: 2010 Int. Conf. Power Syst. Technol, 2010, pp. 1–6, http://dx.doi.org/10.1109/POWERCON.2010.5666557.
[14] C. Liu, J. Zhao, W. Wu, Z. Lu, B. Zhang, A method of voltage sag source location considering source load characteristics, in: 2019 12th Int. Conf. Intell. Comput. Technol. Autom, 2019, pp. 566–569, http://dx.doi.org/10.1109/ICICTA49267.2019.00127.

[15] Y. Yilin, X. Yonghai, Research of method for voltage sag source detection in power distribution network, in: 2011 6th IEEE Conf. Ind. Electron. Appl, 2011, pp. 485–488, http://dx.doi.org/10.1109/ICIEA.2011.5975634.
[16] R. Chouhy Leborgne, D. Karlsson, Voltage sag source location based on voltage measurements only, Electr. Power Qual. Util. 14 (2008) 25–30.
[17] J. Blanco, J.F. Petit, G. Ordóñez, Algorithm for relative location of voltage sags and capacitor switching transients based on voltage measurements only, in: 2014 16th Int. Conf. Harmon. Qual. Power, 2014, pp. 833–837, http://dx.doi.org/10.1109/ICHQP.2014.6842809.
[18] M.H. Moradi, Y. Mohammadi, A new current-based method for voltage sag source location using directional overcurrent relay information, Int. Trans. Electr. Energy Syst. 23 (2013) http://dx.doi.org/10.1002/etep.659.
[19] M.H. Moradi, Y. Mohammadi, M. Hoseyni Tayyebi, A novel method to locate the voltage sag source: A case study in the Brazilian power network (Mato Grosso), Prz. Elektrotechn. 88 (2012).
[20] A.K. Pradhan, A. Routray, S. Madhan Gudipalli, Fault direction estimation in radial distribution system using phase change in sequence current, IEEE Trans. Power Deliv. 22 (2007) 2065–2071, http://dx.doi.org/10.1109/TPWRD.2007.905340.
[21] M.H. Moradi, Y. Mohammadi, Voltage sag source location: A review with introduction of a new method, Int. J. Electr. Power Energy Syst. 43 (2012) http://dx.doi.org/10.1016/j.ijepes.2012.04.041.
[22] Y. Mohammadi, R.C. Leborgne, Improved DR and CBM methods for finding relative location of voltage sag source at the PCC of distributed energy resources, Int. J. Electr. Power Energy Syst. 117 (2020) 105664, http://dx.doi.org/10.1016/j.ijepes.2019.105664.
[23] Y. Mohammadi, R.C. Leborgne, A new approach for voltage sag source relative location in active distribution systems with the presence of inverter-based distributed generations, Electr. Power Syst. Res. 182 (2020) http://dx.doi.org/10.1016/j.epsr.2020.106222.
[24] S.-J. Ahn, D.-J. Won, I.-Y. Chung, S.-I. Moon, A new approach to determine the direction and cause of voltage sag, J. Electr. Eng. Technol. 3 (2008) 300–307, http://dx.doi.org/10.5370/jeet.2008.3.3.300.
[25] V. Barrera, B. López, J. Meléndez, J. Sánchez, Voltage sag source location from extracted rules using subgroup discovery, Front. Artif. Intell. Appl. 184 (2008) 225–235, http://dx.doi.org/10.3233/978-1-58603-925-7-225.
[26] V.B. Núñez, J.M.i. Frigola, S.H. Jaramillo, J.S. Losada, Evaluation of fault relative location algorithms using voltage sag data collected at 25-kV substations, Eur. Trans. Electr. Power 20 (2010) 34–51.
[27] B. Polajžer, D. Dolinar, Evaluation of different methods for voltage sag source detection based on positive-sequence components, 2014, http://dx.doi.org/10.24084/repqj07.270.
[28] B. Polajˇzer, G. Stumberger, S. Seme, D. Dolinar, Detection of voltage sag sources based on instantaneous voltage and current vectors and orthogonal Clarke's transformation, Gener. Transm. Distrib. IET 2 (2008) 219–226, http://dx.doi.org/10.1049/iet-gtd:20070114.
[29] B. Polajzer, G. Stumberger, S. Seme, D. Dolinar, Generalization of methods for voltage sag source detection using vector space approach, in: 2008 IEEE Ind. Appl. Soc. Annu. Meet, 2008, pp. 1–8, http://dx.doi.org/10.1109/08IAS.2008.370.
[30] B. Polajžr, G. Štumberger, D. Dolinar, Detection of voltage sag sources based on the angle and norm changes in the instantaneous current vector written in Clarke's components, Int. J. Electr. Power Energy Syst. 64 (2015) 967–976, http://dx.doi.org/10.1016/j.ijepes.2014.08.011.
[31] B. Polajžer, G. Štumberger, D. Dolinar, Instantaneous positive-sequence current applied for detecting voltage sag sources, IET Gener. Transm. Distrib. 9 (2015) 319–327, http://dx.doi.org/10.1049/iet-gtd.2014.0483.
[32] Y. Mohammadi, M.H. Moradi, R. Chouhy Leborgne, Employing instantaneous positive sequence symmetrical components for voltage sag source relative location, Electr. Power Syst. Res. 151 (2017) 186–196, http://dx.doi.org/10.1016/j.epsr.2017.05.030.
[33] B. Solak, T. Sikorski, Analysis of voltage dip source location methods, in: 2019 Mod. Electr. Power Syst, 2019, pp. 1–6, http://dx.doi.org/10.1109/MEPS46793.2019.9394971.
[34] Y. Mohammadi, R.C. Leborgne, Relative location of voltage sags source at the point of common coupling of constant power loads in distribution systems, Int. Trans. Electr. Energy Syst. 30 (2020) 1–15, http://dx.doi.org/10.1002/2050-7038.12516.
[35] W. Kong, X. Dong, Z. Chen, Voltage sag source location based on instantaneous energy detection, Electr. Power Syst. Res. 78 (2008) 1889–1898, http://dx.doi.org/10.1016/j.epsr.2008.03.016.
[36] W.Ling. Ai, H. Shareef, A single monitor method for voltage sag source location using Hilbert Huang transform, Res. J. Appl. Sci. Eng. Technol. 5 (2013) 192–202, http://dx.doi.org/10.1109/ASSCC.2012.6523296.
[37] X. Chen, D. Li, J. Zhang, K. Ning, Research on location and recognition method of voltage sag disturbance, IOP Conf. Ser. Earth Environ. Sci. 619 (2020) 12027, http://dx.doi.org/10.1088/1755-1315/619/1/012027.
[38] H. Shareef, A. Mohamed, A.A. Ibrahim, Identification of voltage sag source location using S and TT transformed disturbance power, J. Cent. South Univ. 20 (2013) 83–97, http://dx.doi.org/10.1007/s11771-013-1463-5.
[39] Y. Mohammadi, M.H. Moradi, R. Chouhy Leborgne, A novel method for voltage-sag source location using a robust machine learning approach, Electr. Power Syst. Res. 145 (2017) 122–136, http://dx.doi.org/10.1016/j.epsr.2016.12.028.
[40] Y. Mohammadi, A. Salarpour, R. Chouhy Leborgne, Comprehensive strategy for classification of voltage sags source location using optimal feature selection applied to support vector machine and ensemble techniques, Int. J. Electr. Power Energy Syst. 124 (2021) 106363, http://dx.doi.org/10.1016/j.ijepes.2020.106363.
[41] C. Liu, J. Zhao, W. Wu, Z. Lu, R. Zhang, Voltage sag source location based on comprehensive criterion and neural network method, {IOP} Conf. Ser. Earth Environ. Sci. 512 (2020) 12131, http://dx.doi.org/10.1088/1755-1315/512/1/012131.
[42] D. Kai, L. Wei, S. Jianfeng, X. Xianyong, W. Ying, Convolutional neural network for voltage sag source azimuth recognition in electrical internet of things, Wirel. Commun. Mob. Comput. 2021 (2021) 6656564, http://dx.doi.org/10.1155/2021/6656564.
[43] Y. Deng, X. Liu, R. Jia, Q. Huang, G. Xiao, P. Wang, Sag source location and type recognition via attention-based independently recurrent neural network, J. Mod. Power Syst. Clean Energy 9 (2021) 1018–1031, http://dx.doi.org/10.35833/MPCE.2020.000528.
[44] L. Wu, Y. Zhang, X. Hao, W. Chen, Research on a location method for complex voltage sag sources based on random matrix theory, Math. Probl. Eng. 2020 (2020) 7870461, http://dx.doi.org/10.1155/2020/7870461.
[45] J.C. Filho, F.A. Da Silva Borges, R. De Andrade Lira Rabelo, I.S. Silva, A.O. De Carvalho Filho, Optimized method for locating the source of voltage sags, J. Commun. Softw. Syst. 17 (2021) 203–204, http://dx.doi.org/10.24138/JCOMSS-2021-0070.
[46] R.A. de Oliveira, C. Chen, M.H.J. Bollen, R. Chouhy Leborgne, Comparative analysis of transformer-energizing and fault-caused voltage dips on the dynamic behavior of DFIG-based wind turbines, in: 2020 IEEE PES Innov. Smart Grid Technol. Eur, 2020, pp. 589–593, http://dx.doi.org/10.1109/ISGT-Europe47291.2020.9248983.
[47] R.A. de Oliveira, M.H.J. Bollen, Susceptibility of large wind power plants to voltage disturbances — Recommendations to Stakeholders, J. Mod. Power. Syst. Clean Energy XX (2021) 1–13, http://dx.doi.org/10.35833/MPCE.2020.000543.
[48] B. Polajžer, G. Štumberger, S. Seme, D. Dolinar, Impact of asymmetrical disturbance events on voltage sag source detection, Renew. Energy Power Qual. J. 1 (2007) 111–114, http://dx.doi.org/10.24084/repqj05.227.
[49] Y. Mohammadi, R.C. Leborgne, B. Polajžer, Modified methods for voltage-sag source detection using transient periods, Electr. Power Syst. Res. 207 (2022) 107857, http://dx.doi.org/10.1016/j.epsr.2022.107857.
[50] S.M. Miraftabzadeh, M. Longo, F. Foiadelli, M. Pasetti, R. Igual, Advances in the application of machine learning techniques for power system analytics: A survey †, Energies 14 (2021) http://dx.doi.org/10.3390/en14164776.
[51] Y. Sang, H. Zhang, L. Zuo, Least squares support vector machine classifiers using PCNNs, in: 2008 IEEE Int Conf Cybern Intell Syst CIS, Vol. 2008, 2008, pp. 290–295, http://dx.doi.org/10.1109/ICCIS.2008.4670890.
[52] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification And Regression Trees. n.d, http://dx.doi.org/10.1201/9781315139470.
[53] S.M. Miraftabzadeh, F. Foiadelli, M. Longo, M. Pasetti, A survey of machine learning applications for power system analytics, in: Proc - 2019 IEEE Int Conf Environ Electr Eng 2019 IEEE Ind Commer Power Syst Eur EEEIC/I CPS Eur 2019, 2019, http://dx.doi.org/10.1109/EEEIC.2019.8783340.
[54] M. Belgiu, L. Drăguţ, Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens. 114 (2016) 24–31, http://dx.doi.org/10.1016/j.isprsjprs.2016.01.011.
[55] S. Zhang, X. Li, M. Zong, X. Zhu, R. Wang, Efficient kNN classification with different numbers of nearest neighbors, IEEE Trans. Neural Netw. Learn. Syst. (2017) 1–12, http://dx.doi.org/10.1109/TNNLS.2017.2673241.
[56] A. De Caigny, K. Coussement, K.W. De Bock, A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees, Eur. J. Oper. Res. 269 (2018) 760–772.
[57] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52, http://dx.doi.org/10.1016/0169-7439(87)80084-9.
[58] N. Halko, P.G. Martinsson, J.A. Tropp, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev. 53 (2011) 217–288, http://dx.doi.org/10.1137/090771806.

[59] Y. Mohammadi, S.M. Miraftabzadeh, M.H.J. Bollen, M. Longo, An unsupervised learning schema for seeking patterns in rms voltage variations at the sub-10-minute time scale, Sustain Energy Grids Netw. 31 (2022) 100773, http://dx.doi.org/10.1016/j.segan.2022.100773.

[60] K. Van Deun, L. Thorrez, M. Coccia, D. Hasdemir, J.A. Westerhuis, A.K. Smilde, et al., Weighted sparse principal component analysis, Chemom. Intell. Lab. Syst. 195 (2019) 103875, http://dx.doi.org/10.1016/j.chemolab.2019.103875.

[61] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, in: Proc Annu ACM-SIAM Symp Discret Algorithms, 2007, pp. 1027–1035.

[62] D. Mavroeidis, E. Marchiori, Feature selection for k-means clustering stability: theoretical analysis and an algorithm, Data Min. Knowl. Discov. 28 (2014) 918–960, http://dx.doi.org/10.1007/s10618-013-0320-3.

[63] L.van.der. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

[64] A. Bagheri, I.Y.H. Gu, M.H.J. Bollen, E. Balouji, A robust transform-domain deep convolutional network for voltage dip classification, IEEE Trans. Power Deliv. 33 (2018) 2794–2802, http://dx.doi.org/10.1109/TPWRD.2018.2854677.

[65] A. Tjader, I.Y.H. Gu, M.H.J. Bollen, S.K. Ronnberg, Performance evaluation for frequency estimation of transients using the ESPRIT: Measured noise versus white noise, in: 2008 13th Int. Conf. Harmon. Qual. Power, 2008, pp. 1–8, http://dx.doi.org/10.1109/ICHQP.2008.4668780.

[66] Y. Mohammadi, M.H.J. Bollen, Voltage sag source location methods' performance during transient and steady-state periods, in: 2022 20th International Conference on Harmonics & Quality of Power (ICHQP), 2022, http://dx.doi.org/10.1109/ICHQP53011.2022.9808649, In this issue.