energies

MDPI

Article

Prediction of Voltage Sag Relative Location with Data-Driven Algorithms in Distribution Grid

Yunus Yalman [1], Tayfun Uyanık [2], İbrahim Atlı [3], Adnan Tan [4], Kamil Çağatay Bayındır [1], Ömer Karal [1], Saeed Golestan [5,*] and Josep M. Guerrero [5]

[1] Department of Electrical and Electronic Engineering, Ankara Yıldırım Beyazıt University, Ankara 06010, Turkey
[2] Maritime Faculty, Istanbul Technical University, Istanbul 34940, Turkey
[3] Department of Computer Engineering, Ankara Yıldırım Beyazıt University, Ankara 06010, Turkey
[4] Department of Electrical and Electronics Engineering, Çukurova University, Adana 01250, Turkey
[5] Center for Research on Microgrids, AAU Energy, Aalborg University, 9220 Aalborg, Denmark
* Correspondence: sgd@energy.aau.dk

Abstract: Power quality (PQ) problems, including voltage sag, flicker, and harmonics, are the main concerns for the grid operator. Among these disturbances, voltage sag, which affects the sensitive loads in the interconnected system, is a crucial problem in the transmission and distribution systems. The determination of the voltage sag relative location as a downstream (DS) and upstream (US) is an important issue that should be considered when mitigating the sag problem. Therefore, this paper proposes a novel approach to determine the voltage sag relative location based on voltage sag event records of the power quality monitoring system (PQMS) in the real distribution system. By this method, the relative location of voltage sag is defined by Gaussian naive Bayes (Gaussian NB) and K-nearest neighbors (K-NN) algorithms. The proposed methods are compared with support vector machine (SVM) and artificial neural network (ANN). The results indicate that K-NN and Gaussian NB algorithms define the relative location of a voltage sag with 98.75% and 97.34% accuracy, respectively.

Keywords: artificial intelligence; distribution system; power quality; voltage sag

check for updates

Citation: Yalman, Y.; Uyanık, T.; Atlı, İ.; Tan, A.; Bayındır, K.Ç.; Karal, Ö.; Golestan, S.; Guerrero, J.M. Prediction of Voltage Sag Relative Location with Data-Driven Algorithms in Distribution Grid. *Energies* **2022**, *15*, 6641. https://doi.org/10.3390/en15186641

Academic Editor: Julio Barros

Received: 16 August 2022
Accepted: 8 September 2022
Published: 11 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Electrical power quality problems are defined as changes in the amplitude, frequency, and waveform of the grid voltage and current [1]. Integration of renewable power plants (RPP) and distributed generation (DG) affects the stability and power quality of the distribution system and transmission system [2–4]. Therefore, the importance of power quality is gradually increasing with the increase in RPPs integration and distributed generation in recent years [5]. The filters [6], flexible ac transmission system (FACTs) devices [7], smart impedance [8], energy storage [9,10], and multifunctional DGs [11] are used to improve the power quality in the power grid.

Voltage sag, which can be identified as the decrease in voltage amplitude between 0.1 and 0.9 pu in a time from 10 milliseconds to a minute, is one of the important power quality problems encountered in transmission and distribution systems [12]. The main reason for the voltage sag is phase-to-ground and phase-to-phase faults, and it causes shutdown, incorrect operation or malfunction of power electronics-based systems, and undesired operation due to low voltage in relay and contactor coils [13,14]. Researchers have usually focused on the classification, detection, and evaluation of voltage sag. However, it is crucial to determine the relative location of the sag, as it causes many disagreements between the company and the customers (transmission system-distribution system, distribution system-customer) in deciding on the responsibilities and determining the financial penalties [14].

Many papers have been published in the literature on the subject of voltage sag relative location [15–26]. These papers have been categorized under six different head-

ings [15]. These categories are based on the variation of energy and power [16,17], current changes [18–20], impedance changes [21,22], only voltage measurement [23,24], only current measurement [25,26], and artificial intelligence [27–36]. In the first category, the voltage sag relative location is determined using changes in the instantaneous active and reactive power and energy. During the DS events, power flow at the monitoring point (MP) increases, while during the US, it decreases [16,17]. The second category is based on the slope of voltage and active current change. If the current at MP increases, the voltage sag event originates from DS. If the current at MP decreases, it originates from the US [18,19]. In the third category, the relative location of voltage sag is detected using a change in impedance magnitude and phase angle calculated at MP [21,22]. The criteria, which is the change in only voltage magnitude or phase angle jump, is considered to define the voltage sag relative location in the fourth category [23,24]. For the fifth category, the only change in current magnitude and phase angle is observed to define the sag location [25,26]. The advantages and disadvantages of the generalized method are given in Table 1.

Table 1. Advantages and disadvantages of the generalized method [15].

Methods	Advantages	Disadvantages
Depend on power and energy [16]	The application of this method is simple	It requires a threshold for disturbance energy
Current change [20]	It shows suitable accuracy for symmetrical fault	It needs extra decision parameters for unsymmetrical faults
Impedance change [21,22]	It shows suitable accuracy for symmetrical and unsymmetrical fault	It depends on the voltage and current data cycle
Only voltage measurement [23,24]	It needs only the voltage measurement	The threshold is required for the voltage ratio before and during the voltage sag
Only current measurement [25,26]	It needs only current measurement	It needs phasor

In the sixth category, different methods related to artificial intelligence are implemented to find the voltage sag relative location. In [27], pattern recognition based on an SVM is applied to define the relative location of voltage sag using optimal hyperplane for a nonlinear binary classifier. In [28], the robust (SVM) with kernel, linear, and radial basis function is used to determine the relative location of voltage sag location. Moreover, optimal genetic search is applied to find the optimal parameter for each kernel configuration in SVM and to prevent the over-fitting k-fold cross validation is implemented. In [29], the multivariable regression method is proposed to detect voltage sag relative location. In [30], the method proposed in [29] is improved. In this method, the optimal power quality monitoring placement technique is applied based on the genetic algorithm (GA) and Mallow's Cp. In [31], SVM and ensemble techniques are applied to classify the relative location of the voltage sag source proposing a novel analytical rule based on the maximum wins strategy for classification. Moreover, the optimal feature selection is proposed to maximize the SVM accuracy and keeps the accuracy of ensemble techniques with fewer features. In [32], a new method is proposed to define the voltage sag relative location depending on matrix theory and power and energy method using the ring theorem and M-P law of random matrix theory. In [33], the fault location based on voltage sag data recorded by the PQMS is detected using the particle swarm optimization (PSO) algorithm. In [34], a new approach based on the multi-label random forest is proposed to define the identification and relative location of voltage sag sources. Compared with a similar algorithm, the proposed method has high accuracy. In [35], a robust neural network with the AdaBoost algorithm, which is a type of ensemble method, is developed to determine the area where voltage sag is located. Compared with ANN, decision tree, and AdaBoost with a decision tree, the proposed

approach has high accuracy. The deep learning technique based on an attention-based independently recurrent neural network is proposed to define the voltage sag relative location and type recognition at the same time [36].

In light of the studies in the literature, it has been determined that data-driven methods will be an effective way to define the voltage sag relative location. The Gaussian NB and KNN are fast and easy to implement, and these methods can be preferred for real-time application. Therefore, this paper proposes K-NN and GaussianNB algorithms to determine the relative location of voltage sag using real voltage sag data recorded in the distribution system with high accuracy regardless of uncertainty related to power system parameters, as these algorithms are fast and easy to implement. After detecting the voltage sag relative location as US or DS, the fault type that is the main reason for voltage sag is determined as a single-phase ground fault (SLGP), two-phase ground fault (LLGF), three-phase ground fault (LLLGF), and two-phase fault (LLF). The proposed algorithms are compared with common artificial intelligence algorithms, which are SVM and ANN. The results show that the proposed algorithms can identify the voltage sag relation location and fault type with high accuracy and validity. This is the first study for determining the voltage sag relative location and also fault type types that are the main root cause of voltage sag based on K-NN and GaussianNB algorithms using actual data provided by PQMS in the real distribution system. The proposed method can be integrated into PQMS, and this makes it valuable for engineering applications.

This paper is organized as follows: In Section 1, the introduction is given. In Section 2, the background is given, the system overview is explained in Section 3, simulation results are discussed in Section 4, and the discussion and conclusion are given in Sections 5 and 6, respectively.

2. Background

The background principle of the data-driven method and error metrics are described to show the theory behind the proposed method to determine voltage sag relative location.

2.1. ANN

Artificial neural networks (ANNs) are a method for classification, regression, and image recognition that was developed by replicating the operational structure of the human brain [37]. Digital modeling of actual neuron cells and their connections is provided in a simple artificial neural network. This gives the network the ability to learn new information, make judgments, and forecast [38]. The basic ANN construction is shown in Figure 1.

The input variables of the system are imported into the network through the first layer, the input layer, of a simple artificial neural network. These variables are multiplied by weights and stored in a hidden layer [39]. This structure is represented in the system via black dashed line arrows. Specific activations transfer the processed variables collected in the hidden layer to the output layer. These activation functions largely depend on the operation or variable to be taken from the output [40].

2.2. SVM

As a supervised learning method, the SVM, which is frequently used in classification, regression, and image recognition applications, is a powerful and effective method [42]. In this algorithm, two support vectors drawn parallel to each other are used to determine the boundary in classification problems [43] and to separate successful predictions from unsuccessful predictions in regression problems [44]. In this study, it was used to determine the relative location of voltage sag and type of fault since it is a useful classifier and an effective method. Equations (1)–(3) describe the SVM algorithm.

$$f(x) = w^{\mathrm{T}}x + \mathrm{w}_0 \tag{1}$$

$$H(w, \mathrm{w}_0) = \sum_{i=1}^{N}(y_i - f(x_i)) + \frac{\lambda}{2}\|w\|^2 \tag{2}$$

$$V_\epsilon(r) = \begin{cases} 0 \; if \; |r| < \epsilon \\ |r| - \epsilon \; otherwise \end{cases} \tag{3}$$

where w is the normal vector, and x is an independent variable, w_0 is coefficient, λ is the regularization parameter, V is the error function, r is the error, and ϵ is the error margin [45]. Figure 2 shows the main components of the SVM. The orange dots represent successful predictions within the permissible margin borders (support vectors), and the blue dots show unsuccessful results or another cluster [46].

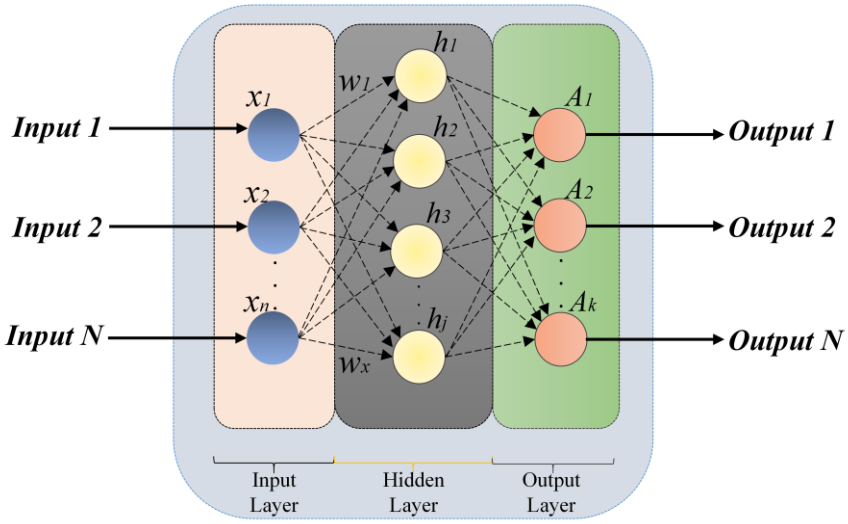


Figure 1. Artificial neural network [41].

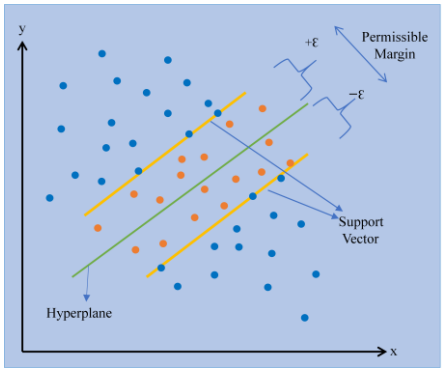


Figure 2. SVM [46].

2.3. Gaussian Naive Bayes

The Gaussian naive Bayes algorithm, a method obtained by applying Bayes' theorem in statistics, is a methodology that makes probabilistic classification based on the relationships between the features in the data set [47]. In this method, when a specific class variable is given to the algorithm, the algorithm takes this variable, processes it, and continues the operations, assuming that the value of this property is independent of any property in the data set [48]. Such a classifier assumes that each of the properties of the variables contributes to the output value of the variable independently of the correlation between them [49]. In this process, the data are first separated according to the classes, and the variance and mean values of the x (continuous attribute) value are found for these classes. The probabilistic distribution of the Gaussian NB model is given in Equation (4).

$$p(x = v|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}} \tag{4}$$

where x is the continuous attribute, the mean of the x values associated with the C_k class is μ_k, and the Bessel adjusted variance of the x values associated with the C_k class is σ_k^2, the observation value is v [50]. Gaussian NB is fast, and implementation is easy. Moreover, it does not require as much training data [51].

2.4. K-Nearest Neighbors

The K-nearest neighbors algorithm, first introduced by Evelyn Fix and Josep Hodges, is a nonparametric supervised learning model in statistics [52]. In regression problems, the average of the k-nearest neighbors is the object's property value, that is, the system's output. In classification problems, the output can be defined as a class membership. The knowledge of belonging to a class is determined by the majority vote of the k-nearest neighbors. In this process, if k = 1, this object should be assigned to the nearest neighbor class [53]. Since distance is considered as a criterion while classifying in the K-NN algorithm, normalizing the other features of the input in the training data while determining which class an input belongs to can significantly increase the accuracy of the estimation process [54]. The Minkowski distance is widely used in the K-NN algorithm, and it is described in Equation (5) [55].

$$L^{\mathrm{P}}(x_j, x_q) = \left(\sum_i |x_{j,i} - x_{q,i}|^{\mathrm{P}} \right)^{1/\mathrm{P}} \quad (5)$$

where L^{P} is Minkowski distance, x_j and x_q are given any points, and P is the coefficient. P = 1 is Manhattan distance and P = 2 is Euclidean distance [56]. It is too easy to implement as only two parameters (k and distance function) need to be specified. Moroever, there is no training period for KNN. It means that KNN stores training data, and it learns from the dataset only at the time of the prediction process [57].

2.5. Error Metrics

The structure that shows the realized estimations and the errors made in more detail than the error metrics is the confusion matrix. In this matrix, TN represents negative samples that are correctly classified, and TP represents positive samples that are correctly classified. FP and FN represent incorrectly classified positive and misclassified negative samples, respectively [58]. A typical confusion matrix is described in Figure 3.

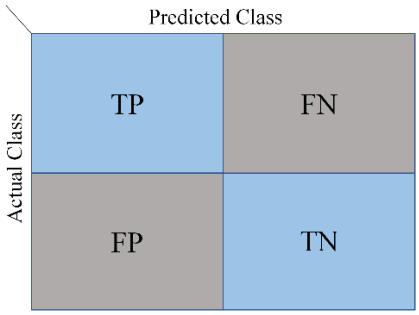


Figure 3. Typical structure of confusion matrix.

This study examined the success of predictive models using four different error metrics. The first of these methods is accuracy. Accuracy is the ratio of correctly predicted output variables to the total test data expressed in Equation (6) [59]. In cases where the variables in the data set do not have equal proportions, the accuracy metric is insufficient to measure the model achievement. An indication of how accurately the predictions are made can also be quantified with precision [60]. The calculation of the precision metric is shown in Equation (7). While the recall [61] metric shows how many values that need to be predicted correctly are estimated correctly (Equation (8)), the F1 score metric expresses the harmonic mean of the precision and recall metrics (Equations (9)). The advantage of using the F1 score

value is to detect whether there is a misleading success rate in datasets whose variables are not evenly distributed and to select the most successful model [62].

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

where TP is true positive predictions, FP is false positive predictions, and FN is false negative values, TN is true negative predictions.

3. System Overview

The relative location of the voltage sag is determined according to the busbar to which the power quality device is connected. Here, the position of the voltage sag is determined as US or DS by reference to the direction of the active power flow, as shown in Figure 4. If an F1 fault has occurred, the position of the voltage sag is the US; if the F2 fault has occurred, the position of the voltage sag is DS [15].

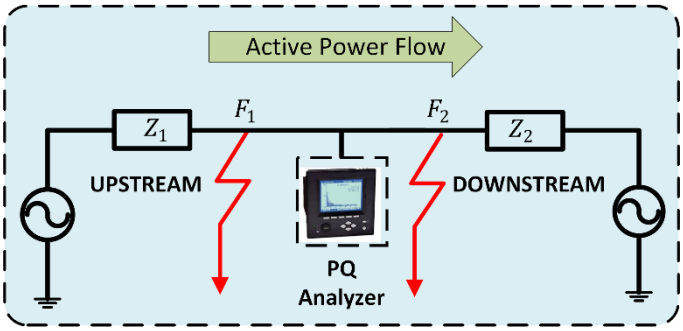


Figure 4. Equivalent circuit model of the voltage sag.

Distribution system operators (DSOs) are required to have a number of power quality analyzers determined by the Energy Market Regulatory Authority (EMRA), which regulates electricity, natural gas, and petroleum markets in Turkey. The number of power quality analyzers is calculated using Equation (10) [63]. The data received from these power quality analyzers are monitored by DSO with the help of PQMS, which is called Inavitas [64]. This software can perform monitoring and management of distribution grids for innovative grid solutions.

The PQ analyzer records the current and voltage waveform to the Comtrade file if the voltage magnitude of the busbar to which it connects drops below 0.9 pu. This Comtrade file is read via PQMS. After that, the PQMS system provides event data during voltage and before voltage sag, and the algorithm results in DSO. These data provided by PQMS are shown in Figure 5.

$$\text{NPQA} = \left(\frac{(\text{MS} \times 0.6) + (\text{FS}_{\text{OG}} \times 0.3) + (\text{DTS} \times 0.2)}{100} \right) + 35, \tag{10}$$

where NPQA is the number of power quality analyzers to be installed for the first time in the measurement year, MS is the substation number, FS_{OG} is the feeder number, and DTS is the number of distribution transformers.

Date and Time	Nominal voltage of bus	$V_{1\ max}$ p.u	$V_{a\ min}$ p.u	$V_{b\ max}$ p.u	$V_{b\ min}$ p.u	$V_{c\ max}$ p.u	$V_{c\ min}$ p.u	Voltage sag location
26.08.2017 01:52:01	19.9 kV	0.89	0.89	0.904	0.904	0.901	0.901	Upstream
Voltage sag Duration (s)	Phase	$I_{a\ max}$ A	$I_{a\ min}$ A	$I_{b\ max}$ A	$I_{b\ min}$ A	$I_{c\ max}$ A	$I_{c\ min}$ A	Fault Type
0.059	A	51.5	51.5	51.5	51.5	52.4	52.4	----------

Figure 5. Example data provided by the PQMS during voltage sag and before voltage sag.

Figure 6 shows the general structure of the test system. Firstly, voltage sag data are recorded by the power quality analyzer, and these data are sent to the operators with the help of PQMS. The relative location of the voltage sag is estimated using the K-NN, Gaussian NB, SVM, and ANN. After that, fault types that are the main reason for voltage sag are determined for DS events as SLGF, LLGF, LLLGF, and LLF.

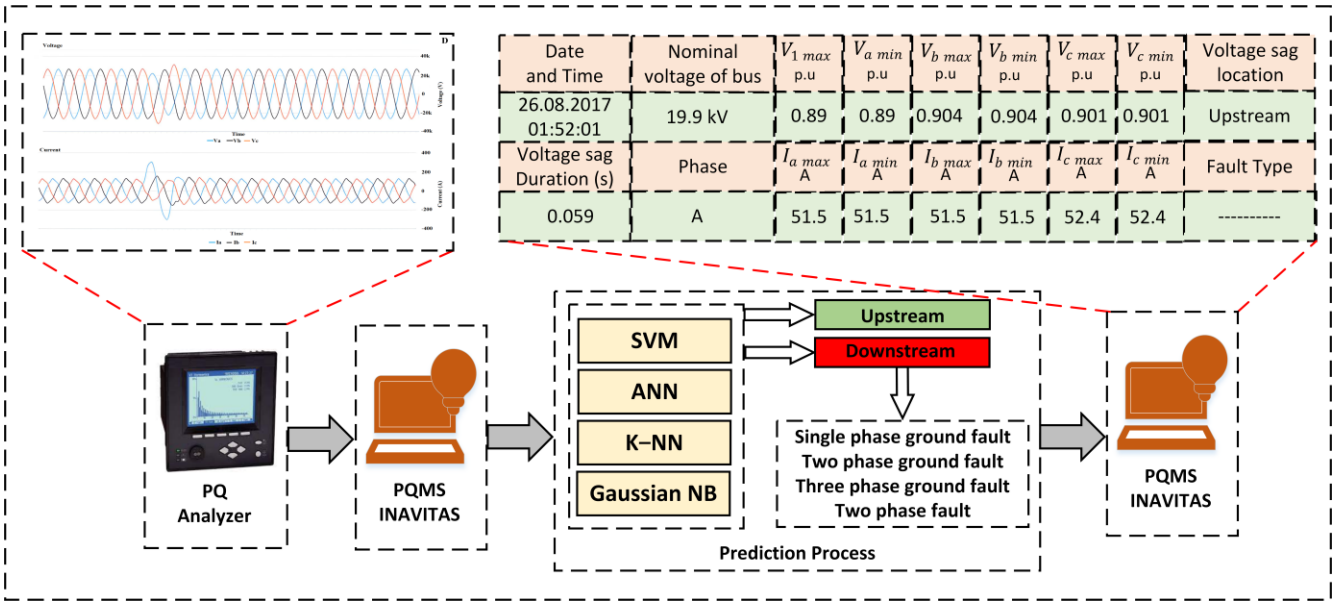


Figure 6. The general structure of the test system.

Event Data

Power quality devices record current and voltage waveforms during voltage sag. The checking process of the accuracy for the prediction of the voltage sag relative direction is determined by these current and voltage waveforms. Some samples of the voltage sag events are shown in Table 2 by giving information such as the location, time, and fault type, and current and voltage waveforms are represented in Appendix A. Figures A1–A8 show the three-phase voltage and current waveforms recorded by the PQ analyzer during the voltage sag. While the proposed method classifies the voltage sag relative location as DS for the event, which is shown in Figures A1–A6, it is determined as the US for the voltage sag event shown in Figures A7 and A8. As can be seen from Figures A1–A4, while the voltage level of only one phase drops below 0.9 pu, at the same time the current of this phase increases because of SLGF. For Figure A5, the voltage level of phase-A and phase-B drops below 0.9 pu; at the same time, the current of these phases increases due to LLGF. When Figure A6 is examined, it is observed that the voltage level of the three-phase drops below 0.9 pu, and meanwhile current of the three phases increases due to LLLGF. Since the voltage level of the two phases drops below 0.9 pu, and the current of these phases decreases, the relative location of the voltage sag given in Figures A7 and A8 is determined as the US.

Table 2. Power quality analyzer data related to the events.

Event ID	Event Direction	Fault Type	V-I Data
1	DS	SLGF	Figure A1
2	DS	SLGF	Figure A2
3	DS	SLGF	Figure A3
4	DS	SLGF	Figure A4
5	DS	LLGF	Figure A5
6	DS	LLLGF	Figure A6
7	US	-	Figure A7
8	US	-	Figure A8

In this dataset, each event datum contains the magnitude of voltage and current during the voltage sag and before voltage sag per each phase. In this study, 6118 event data recorded in the distribution system are used. Because there are some missing data during the events, these missing data are removed from the dataset, and a dataset containing 6028 samples in total is created. A total of 80% of these data are reserved for training, and the remaining 20% is for testing the success of algorithms. The training and test dataset are randomly selected by the computer. Four fault types are determined as output variables in the data set. The ratio of these fault types to the whole output data set is 0.0343, 0.013, 0.005, and 0.948 for DS and SLGF, DS and LLGF, DS and LLLGF, and the US, respectively.

4. Simulation Results

The 3.7.7 version of the Python programming language and the TensorFlow environment on the Spyder 4 interface are used as the simulation software. The parameters of the models and simulation results are shown in Table 3.

Table 3. Model Parameters and simulation results.

Model	Parameter	Accuracy	Recall	F1 Score	Precision
ANN	Optimizer = Adam, iteration = 2000, input layer size = 2048, hidden layer size = 1024, output layer size = 4, activation function (input layer) = ReLU, activation function (hidden layer) = ReLU, activation function (output layer) = Softmax, batch_size = 2000, dropout ratio = 0.5.	0.9819	0.99	0.98	0.99
SVM	Kernel = Rbf, degree = 3, penalty = l2, iteration = 30, class weight = {0:25, 1:197., 2:65.5, 3:1}.	0.9759	0.98	0.98	0.97
K-NN	n_neighbors = 4, metric = 'minkowski'.	0.9875	0.99	0.99	0.99
Gaussian NB	var_smoothing = 0.1	0.9734	0.97	0.97	0.97

Figure 7 shows the confusion matrix of the proposed methods. D1, D2, D3, and UP shown in Figure 7 indicate downstream and single-phase ground fault, downstream and two-phase ground fault, downstream and three-phase ground fault, and upstream, respectively. When the results are examined, it is observed that the KNN model is a more successful solution with an accuracy rate of 0.9875, a recall rate of 0.99, an F1 score of 0.99, and a precision score of 0.99 compared to the other models. Although the accuracy rate of the ANN is lower than the KNN model, it was able to detect the fault more successfully than SVM and Gaussian NB algorithms. Estimation results are analyzed with four different metrics. A more detailed examination of the success of the algorithms according to the type of fault is made in Figure 7 with the help of confusion matrices.

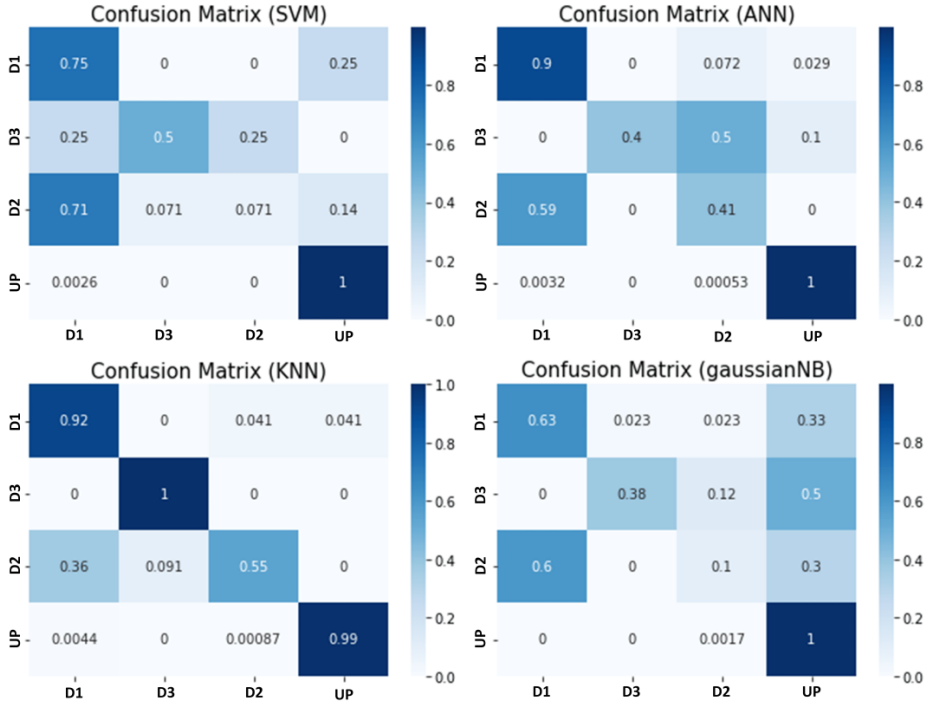


Figure 7. Confusion matrices of the algorithms.

5. Discussion

The proposed algorithms classify the voltage sag relative location with high accuracy, and the results of the algorithms are compared with ANN and SVM, which are common state-art methods. A detailed explanation of the results, which are shown in Figure 7, is given below for each method.

SVM performed the predictions with high accuracy for the US event. In problems in which the data set is not properly distributed, the correct estimation of the variables with the low ratio in the data set will determine the most efficient algorithm. Therefore, if DS and SLGF, DS and LLGF, and DS and LLLGF faults are evaluated, respectively:

- When the DS and SLGF prediction results are examined, which include approximately 3.43% of the data set, it can be said that the SVM method predicts the fault type correctly with a rate of 75%. In addition, 25% of what was expected to be DS and SLGF was predicted as the US;
- When the DS and LLGF estimation results are examined, it is seen that this fault type, which has a rate of 1.31% in the data set, was predicted correctly by SVM at a rate of 7.1%. With this score, it can be said that the SVM algorithm is quite unsuccessful in predicting DS and LLGF. In the DS and LLGF estimation process, SVM incorrectly predicted the type of fault as DS and SLGF in 71% of the cases. In addition, it estimated the fault type as UP with a rate of 14% and DS and LLLGF with a rate of 7.1%. It can be said that the SVM algorithm fails to predict the DS and LLGF;
- When the DS and LLLGF failure is examined, it can be seen that it constitutes approximately 0.48% of the data set. When the prediction performance of SVM in this fault type is examined, it is seen that it predicts the fault type correctly at 50%. In addition, SVM estimated 25% as DS and SLGF, and the remaining 25% as DS and LLGF, which should have been estimated as DS and LLLGF.

The ANN algorithm, like the SVM algorithm, predicted the US with a high accuracy rate. If the prediction success of the algorithm is examined in more detail:

- When the DS and SLGF prediction performance is examined, it can be seen that the ANN algorithm reaches 90% accuracy. In addition, ANN estimated about 7.2% of the output data as DS and LLGF, which it should have estimated as DS and SLGF. In the

remaining 2.9% estimation of DS and SLGF, ANN predicted the voltage sag relative location and fault type as DS and LLGF;

- When the results obtained by the ANN algorithm in the DS and LLGF estimation process are examined, it can be said that the algorithm can predict this error type correctly at a rate of 41%. Moreover, 59% of what was expected to estimate as DS and LLGF were DS and SLGF. Therefore, the algorithm was ineffective in estimating DS and LLGF;
- If the DS and LLLGF prediction performance is examined, it is seen that the ANN algorithm reaches 40% accuracy. This value shows that the algorithm is ineffective in the DS and LLLGF estimation process. On the other hand, ANN, DS, and LLLGF estimated 50% of the outputs as DS and LLGF and 10% as the US.

When the fault type prediction results of the Gaussian NB algorithm are examined, it can be said that it predicts the US event with a high accuracy rate, such as ANN and SVM methods. If the estimations made by this method are examined in detail:

- When the DS and SLGF estimation performance of the Gaussian NB algorithm is examined, it can be seen that it reaches 63% accuracy. It was determined that this method estimated 2.3% as DS and LLLGF and 2.3% as DS and LLGF, which should be estimated as DS and SLGF. The algorithm estimated the remaining 33% as the US, while it should have predicted DS and SLGF;
- When the DS and LLGF estimation results were examined, the Gaussian NB algorithm, which showed a success rate of 10%, was inefficient in estimating this fault type. The algorithm estimated 60% as DS and SLGF and 30% as US, which it should have estimated as DS and LLGF;
- If the algorithm's DS and LLLGF estimation performance is analyzed, it can be observed that it reaches an accuracy of 38%. In addition, 12% of the output values that should be estimated as DS and LLLGF were estimated as DS and LLGF, and 50% as the US.

If the fault type prediction results of the KNN method are examined, it can be seen that it predicts the US event with high accuracy, like the other algorithms discussed before. If the fault type prediction results of the KNN method are discussed in more detail:

- In DS and SLGF estimation, it can be observed that the algorithm is more successful than other algorithms, with an accuracy of 92%. The algorithm estimated 4.1% as DS and LLGF and 4.1% as the US of the output values it should have predicted as DS and SLGF;
- When the DS and LLGF estimation performance of the KNN method was examined, it was observed that it reached 55% accuracy. Again, this value is higher than other algorithms' DS and LLGF prediction scores. KNN estimated 36% as DS and SLGF and 9.1% as DS and LLLGF, which should have estimated DS and LLGF;
- It has been observed that the algorithm predicts this fault type exactly in the DS and LLLGF process, where the KNN algorithm differs significantly from other algorithms in the prediction process. When compared with other algorithms used in voltage sag relative estimation, it is seen that the KNN algorithm has a limited advantage over other algorithms in general accuracy. However, for DS and SLGF, DS and LLGF, and DS and LLLGF, which make up about 6% of the output values of the dataset and are more difficult to predict than the US event, it can be said significantly more effective than the other methods.

6. Conclusions

In this paper, KNN and Gaussian NB algorithms are proposed to classify the relative location of the voltage sag and the type of fault causing the voltage sag. The real voltage sag data provided by PMQS in the distribution system are used as a dataset for realistic, useful, and meaningful results. The results of the study are listed:

- The KNN algorithms classify the voltage sag relative location with high rates of 0.9875. The KNN shows more accurate results, as can be seen from the error metrics, which are accuracy, precision, F1 score, and recall. Moreover, according to the confusion matrix created to examine the estimation results of the algorithms in more detail, KNN is more successful than other algorithms in terms of classification of fault type despite limited data for some events;
- Gaussian NB algorithms determine the voltage sag relative location with high rates of 0.9734. As can be seen from the confusion matrix, the Gaussian NB shows lower accuracy like SVM and ANN when it determines the fault type, especially in D3 and D2 regions due to limited data regarding these events;
- Distribution grid operators can take much faster action to mitigate voltage sag problems in the grid, thanks to a highly accurate estimation of the voltage sag relative location and type of fault. In this way, potential damages to industrial users will be prevented;
- As different companies operate the distribution and transmission systems in Turkey, this study helps to understand whether the voltage sag event is caused by the transmission or the distribution grid. In addition to determining the relative position of the voltage sag, the proposed method can also help to identify the primary source of the voltage sag if the grid model and PQ analyzer measurement points are integrated into the PQMS. After the proposed algorithms determine the voltage sag relative location, this information should be visualized on the grid model. DSO operator can define the exact location of voltage sag by following the change of direction (US to DS or DS to US). Moreover, if the proposed algorithm is revised by adding grid topology and rules for finding the exact location, they can automatically determine the exact location of voltage sag.

Author Contributions: Conceptualization, Y.Y. and T.U.; methodology, Y.Y., T.U. and İ.A.; software, Y.Y., T.U. and İ.A.; validation, Y.Y., T.U. and K.Ç.B.; formal analysis, Y.Y.; investigation, A.T. and K.Ç.B.; writing—original draft preparation, Y.Y.; writing—review and editing, A.T., K.Ç.B., Ö.K. and S.G.; visualization, Y.Y. and T.U.; supervision, A.T., K.Ç.B. and J.M.G.; project administration, K.Ç.B., S.G. and J.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by The Scientific and Technological Research Council of Turkey BIDEB-2214-A International Research Fellowship Programme for PhD Students.

Acknowledgments: The authors would like to thank ENDOKS and KCETAŞ company.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

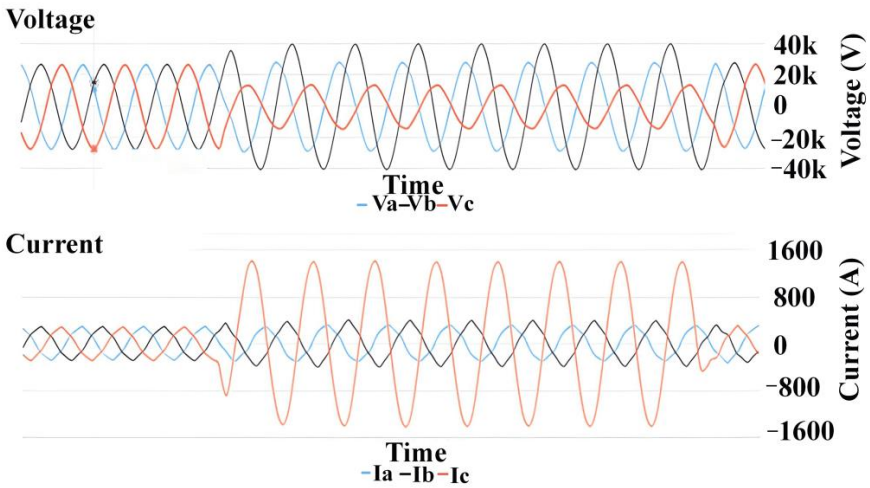


Figure A1. Event data recorded by PQ analyzer for SLGF and DS.

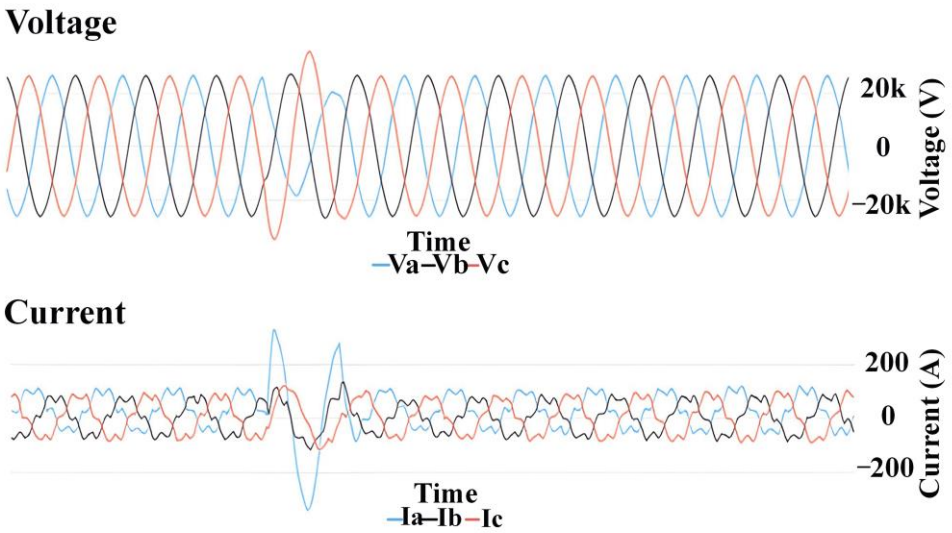


Figure A2. Event data recorded by PQ analyzer for SLGF and DS.

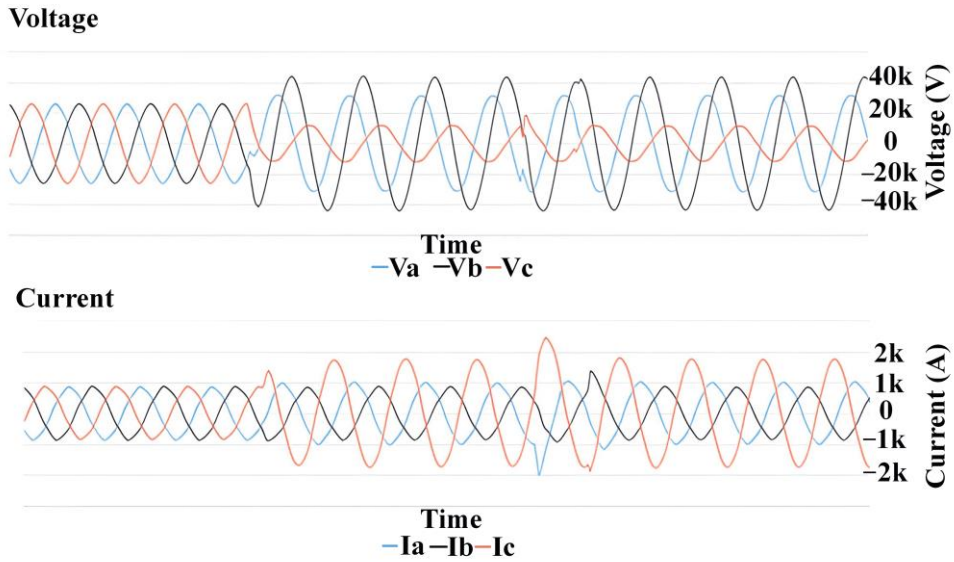


Figure A3. Event data recorded by PQ analyzer for SLGF and DS.

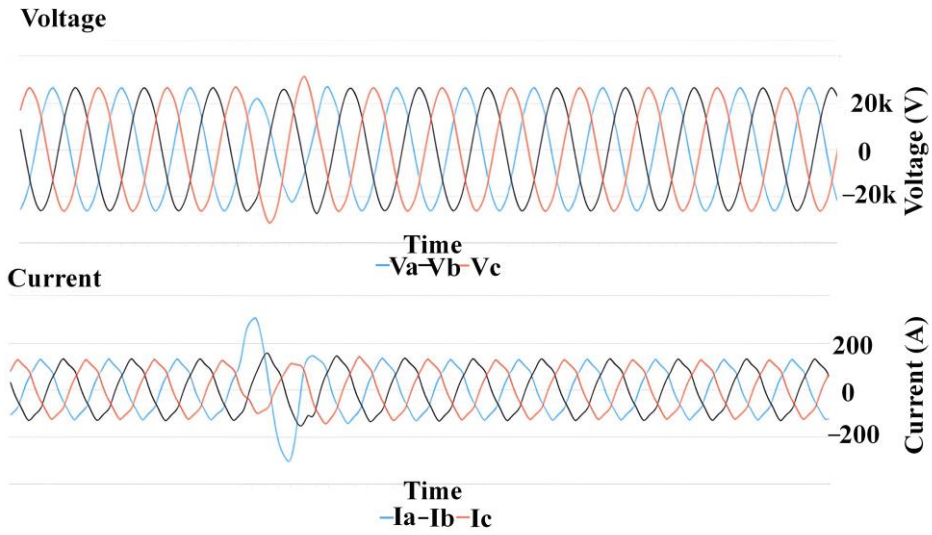


Figure A4. Event data recorded by PQ analyzer for SLGF and DS.

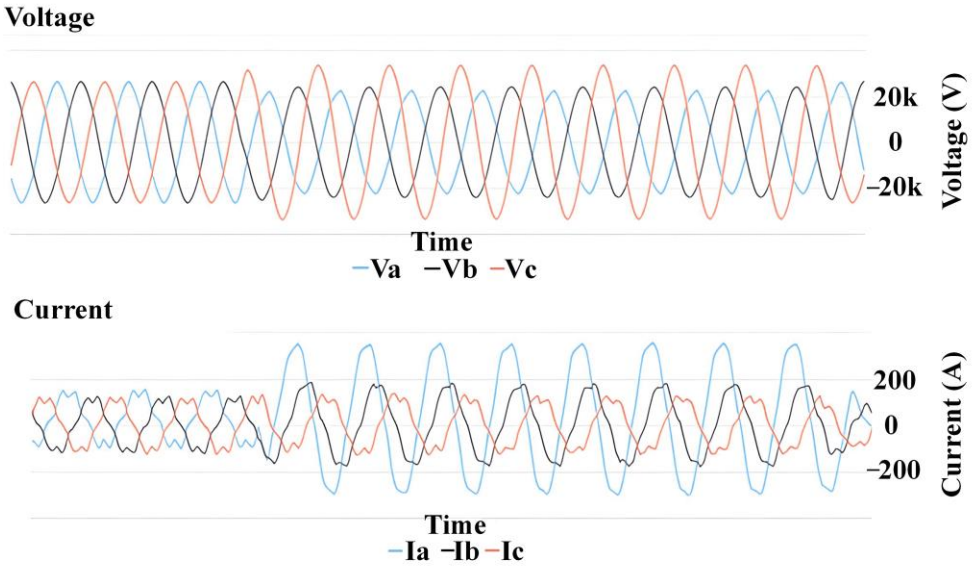


Figure A5. Event data recorded by PQ analyzer for LLGF and DS.

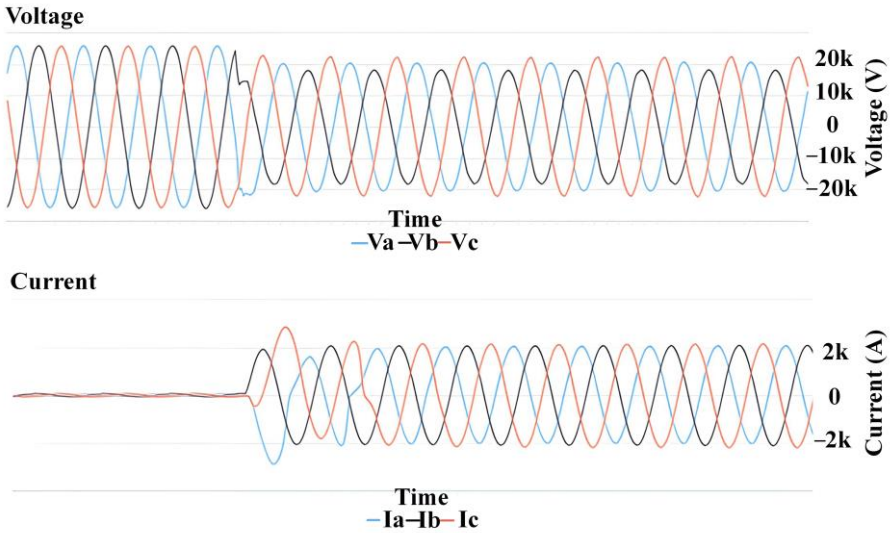


Figure A6. Event data recorded by PQ analyzer for LLLGF and DS.

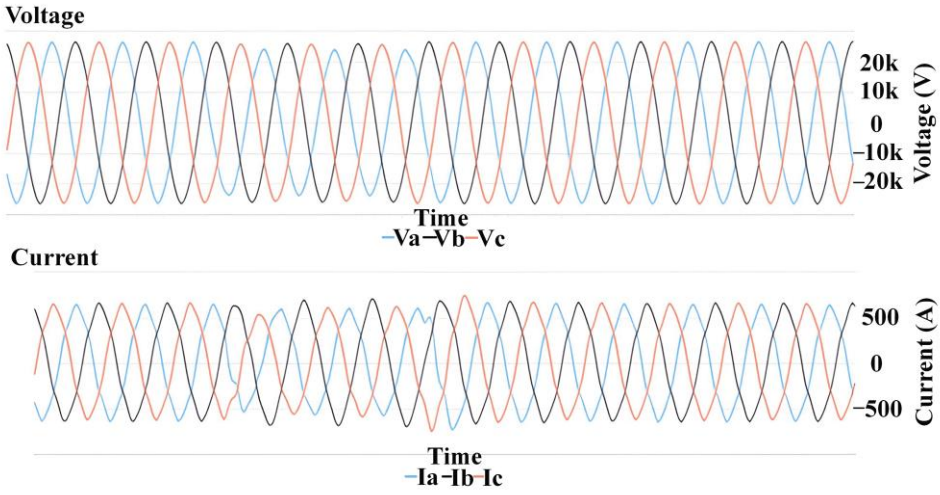


Figure A7. Event data recorded by PQ analyzer for the US.

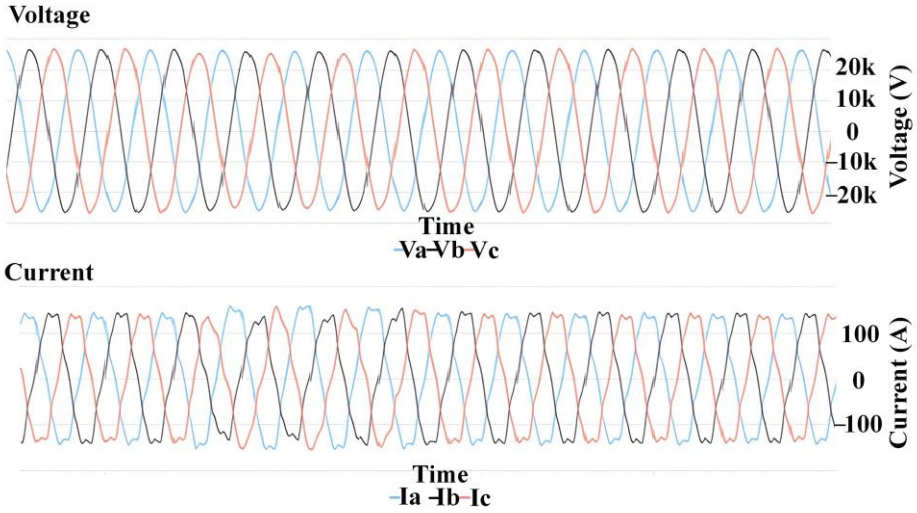


Figure A8. Event data recorded by PQ analyzer for the US.

References

1. Ucar, M.; Ozdemir, S.; Ozdemir, E. A Four-Leg Unified Series-Parallel Active Filter System for Periodic and Non-Periodic Disturbance Compensation. *Electr. Power Syst. Res.* **2011**, *81*, 1132–1143. [CrossRef]
2. Ghaffarianfar, M.; Hajizadeh, A. Voltage Stability of Low-Voltage Distribution Grid with High Penetration of Photovoltaic Power Units. *Energies* **2018**, *11*, 1960. [CrossRef]
3. Kreishan, M.Z.; Fotis, G.P.; Vita, V.; Ekonomou, L. Distributed Generation Islanding Effect on Distribution Networks and End User Loads Using the Load Sharing Islanding Method. *Energies* **2016**, *9*, 956. [CrossRef]
4. Sultan, H.M.; Zaki Diab, A.A.; Kuznetsov, O.N.; Ali, Z.M.; Abdalla, O. Evaluation of the Impact of High Penetration Levels of PV Power Plants on the Capacity, Frequency and Voltage Stability of Egypt's Unified Grid. *Energies* **2019**, *12*, 552. [CrossRef]
5. Saribulut, L.; Teke, A.; Latran, M.B. Dağitim Sistemleri Için Çok Fonksiyonlu Statik Senkron Kompanzatör. *J. Fac. Eng. Archit. Gazi Univ.* **2016**, *31*, 727–736. [CrossRef]
6. Lee, T.L.; Hu, S.H. An Active Filter with Resonant Current Control to Suppress Harmonic Resonance in a Distribution Power System. *IEEE J. Emerg. Sel. Top. Power Electron.* **2016**, *4*, 198–209. [CrossRef]
7. Jiang, F.; Tu, C.; Shuai, Z.; Cheng, M.; Lan, Z.; Xiao, F. Multilevel Cascaded-Type Dynamic Voltage Restorer with Fault Current-Limiting Function. *IEEE Trans. Power Deliv.* **2016**, *31*, 1261–1269. [CrossRef]
8. Da Silva, C.H.; Pereira, R.R.; Borges Da Silva, L.E.; Lambert-Torres, G.; Gonzatti, R.B.; Ferreira, S.C.; Fernandez Silva, L.G. Smart Impedance: Expanding the Hybrid Active Series Power Filter Concept. In Proceedings of the IECON 2012—38th Annual Conference on IEEE Industrial Electronics Society, Montreal, QC, Canada, 25–28 October 2012; pp. 1416–1421. [CrossRef]
9. Adewumi, O.B.; Fotis, G.; Vita, V.; Nankoo, D.; Ekonomou, L. The Impact of Distributed Energy Storage on Distribution and Transmission Networks' Power Quality. *Appl. Sci.* **2022**, *12*, 6466. [CrossRef]
10. Nieto, A.; Vita, V.; Maris, T.I. Power Quality Improvement in Power Grids with the Integration of Energy Storage Systems. *Int. J. Eng. Res. Technol.* **2016**, *5*, 438–443.
11. He, J.; Li, Y.W.; Munir, M.S. A Flexible Harmonic Control Approach through Voltage-Controlled DG-Grid Interfacing Converters. *IEEE Trans. Ind. Electron.* **2012**, *59*, 444–455. [CrossRef]
12. Latran, M.B.; Teke, A. Investigation of Inverter Based Shunt Compensators for Mitigation of Power Quality Problems in Power Distribution System. *J. Fac. Eng. Archit. Gazi Univ.* **2014**, *29*, 793–805.
13. Inci, M.; Bayindir, K.Ç.; Tümay, M. Dinamik Gerilim Iyileştiricilerde Gerilim Problemlerinin Tespiti Için Yeni Yöntem Geliştirilmesi. *J. Fac. Eng. Archit. Gazi Univ.* **2016**, *31*, 997–1006. [CrossRef]
14. Moradi, M.H.; Mohammadi, Y. Voltage Sag Source Location: A Review with Introduction of a New Method. *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 29–39. [CrossRef]
15. Mohammadi, Y.; Moradi, M.H.; Chouhy Leborgne, R. Locating the Source of Voltage Sags: Full Review, Introduction of Generalized Methods and Numerical Simulations. *Renew. Sustain. Energy Rev.* **2017**, *77*, 821–844. [CrossRef]
16. Parsons, A.C.; Grady, W.M.; Powers, E.J.; Soward, J.C. A Direction Finder for Power Quality Disturbances Based upon Disturbance Power and Energy. *IEEE Trans. Power Deliv.* **2000**, *15*, 1081–1086. [CrossRef]
17. Kong, W.; Dong, X.; Chen, Z. Voltage Sag Source Location Based on Instantaneous Energy Detection. *Electr. Power Syst. Res.* **2008**, *78*, 1889–1898. [CrossRef]
18. Li, C.; Tayjasanant, T.; Xu, W.; Liu, X. Method for Voltage-Sag-Source Detection by Investigating Slope of the System Trajectory. *IEE Proc. Commun.* **2003**, *150*, 367–372. [CrossRef]
19. Hamzah, N.; Mohamed, A.; Hussain, A. A New Approach to Locate the Voltage Sag Source Using Real Current Component. *Electr. Power Syst. Res.* **2004**, *72*, 113–123. [CrossRef]

20. Leborgne, R.C.; Makaliki, R. Voltage Sag Source Location at Grid Interconnections: A Case Study in the Zambian System. In Proceedings of the 2007 IEEE Lausanne Power Tech, Lausanne, Switzerland, 1–5 July 2007; pp. 1852–1857. [CrossRef]
21. Tayjasanant, T.; Li, C.; Xu, W. A Resistance Sign-Based Method for Voltage Sag Source Detection. *IEEE Trans. Power Deliv.* **2005**, *20*, 2544–2551. [CrossRef]
22. Pradhan, A.K.; Routray, A. Applying Distance Relay for Voltage Sag Source Detection. *IEEE Trans. Power Deliv.* **2005**, *20*, 529–531. [CrossRef]
23. Leborgne, R.C.; Karlsson, D. Voltage sag source location based on voltage measurements only. *Electr. Power Qual. Util. J.* **2008**, *14*, 25–30.
24. Blanco-Solano, J.; Petit-Suárez, J.F.; Ordóñez-Plata, G. Methodology for Relative Location of Voltage Sag Source Using Voltage Measurements Only. *DYNA* **2015**, *82*, 94–100. [CrossRef]
25. Pradhan, A.K.; Routray, A.; Gudipalli, S.M. Fault Direction Estimation in Radial Distribution System Using Phase Change in Sequence Current. *IEEE Trans. Power Deliv.* **2007**, *22*, 2065–2071. [CrossRef]
26. Polajžer, B.; Štumberger, G.; Dolinar, D. 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.* **2015**, *64*, 967–976. [CrossRef]
27. Lv, G.; Sun, W. Voltage Sag Source Location Based on Pattern Recognition. *J. Energy Eng.* **2013**, *139*, 136–141. [CrossRef]
28. Mohammadi, Y.; Moradi, M.H.; Chouhy Leborgne, R. A Novel Method for Voltage-Sag Source Location Using a Robust Machine Learning Approach. *Electr. Power Syst. Res.* **2017**, *145*, 122–136. [CrossRef]
29. Kazemi, A.; Mohamed, A.; Shareef, H. Tracking the Voltage Sag Source Location Using Multivariable Regression Model. *Int. Rev. Electr. Eng.* **2011**, *6*, 1853–1861.
30. Kazemi, A.; Mohamed, A.; Shareef, H.; Raihi, H. Accurate Voltage Sag-Source Location Technique for Power Systems Using GACp and Multivariable Regression Methods. *Int. J. Electr. Power Energy Syst.* **2014**, *56*, 97–109. [CrossRef]
31. Mohammadi, Y.; Salarpour, A.; Chouhy Leborgne, R. 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.* **2020**, *124*, 106363. [CrossRef]
32. Wu, L.; Zhang, Y.; Hao, X.; Chen, W. Research on a Location Method for Complex Voltage Sag Sources Based on Random Matrix Theory. *Math. Probl. Eng.* **2020**, *2020*, 7870461. [CrossRef]
33. Zhao, C.; Tao, S.; Xiao, X. Fault Location Estimation Based on Voltage Sag Information of PQMS. *Dianwang Jishu/Power Syst. Technol.* **2016**, *40*, 642–648. [CrossRef]
34. Liu, J.; Song, H.; Zhou, L. Identification and Location of Voltage Sag Sources Based on Multi-Label Random Forest. In Proceedings of the 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 21–23 November 2019; pp. 2025–2030. [CrossRef]
35. Borges, F.A.S.; Rabelo, R.A.L.; Fernandes, R.A.S.; Araujo, M.A. Methodology Based on Adaboost Algorithm Combined with Neural Network for the Location of Voltage Sag Disturbance. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019. [CrossRef]
36. Deng, Y.; Liu, X.; Jia, R.; Huang, Q.; Xiao, G.; Wang, P. Sag Source Location and Type Recognition via Attention-Based Independently Recurrent Neural Network. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 1018–1031. [CrossRef]
37. Noronha Barros, F.G.; dos Fonseca, W.A.S.; Bezerra, U.H.; Nunes, M.V.A. Compression of Electrical Power Signals from Waveform Records Using Genetic Algorithm and Artificial Neural Network. *Electr. Power Syst. Res.* **2017**, *142*, 207–214. [CrossRef]
38. Moustafa, E.B.; Hammad, A.H.; Elsheikh, A.H. A New Optimized Artificial Neural Network Model to Predict Thermal Efficiency and Water Yield of Tubular Solar Still. *Case Stud. Therm. Eng.* **2022**, *30*, 101750. [CrossRef]
39. Bento, M.E.C. Load Margin Assessment of Power Systems Using Artificial Neural Network and Genetic Algorithms. *IFAC-PapersOnLine* **2022**, *55*, 944–948. [CrossRef]
40. Lin, W.; Chen, G. Large Memory Capacity in Chaotic Artificial Neural Networks: A View of the Anti-Integrable Limit. *IEEE Trans. Neural Netw.* **2009**, *20*, 1340–1351. [CrossRef]
41. Han, Y.; Xiu, L.; Wang, Z.; Chen, Q.; Member, S.; Tan, S.; Member, S. Artificial neural networks controlled fast valving in a power generation plant. *IEEE Trans. Neural Netw.* **1997**, *8*, 373–389. [PubMed]
42. Eke, C.S.; Jammeh, E.; Li, X.; Carroll, C.; Pearson, S.; Ifeachor, E. Early Detection of Alzheimer's Disease with Blood Plasma Proteins Using Support Vector Machines. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 218–226. [CrossRef] [PubMed]
43. Borges, F.; Pinto, A.; Ribeiro, D.; Barbosa, T.; Pereira, D.; Magalhaes, R.; Barbosa, B.; Ferreira, D. An Unsupervised Method Based on Support Vector Machines and Higher-Order Statistics for Mechanical Faults Detection. *IEEE Lat. Am. Trans.* **2020**, *18*, 1093–1101. [CrossRef]
44. Arias Velásquez, R.M. Support Vector Machine and Tree Models for Oil and Kraft Degradation in Power Transformers. *Eng. Fail. Anal.* **2021**, *127*, 105488. [CrossRef]
45. Steinhäuser, L.; Coumont, M.; Weck, S.; Hanson, J. Comparison of RMS and EMT Models of Converter-Interfaced Distributed Generation Units Regarding Analysis of Short-Term Voltage Stability. In Proceedings of the NEIS 2019, Conference on Sustainable Energy Supply and Energy Storage Systems, Hamburg, Germany, 19–20 September 2019; pp. 1–6.
46. Hafdaoui, H.; Boudjelthia, E.A.K.; Chahtou, A.; Bouchakour, S.; Belhaouas, N. Analyzing the Performance of Photovoltaic Systems Using Support Vector Machine Classifier. *Sustain. Energy Grids Netw.* **2022**, *29*, 100592. [CrossRef]

47. Venkata, P.; Pandya, V. Data Mining Model and Gaussian Naive Bayes Based Fault Diagnostic Analysis of Modern Power System Networks. *Mater. Today Proc.* **2022**, *62*, 7156–7161. [CrossRef]
48. Oladeji, I.; Zamora, R.; Lie, T.T. An Online Security Prediction and Control Framework for Modern Power Grids. *Energies* **2021**, *14*, 6639. [CrossRef]
49. Nam, S.; Hur, J. Probabilistic Forecasting Model of Solar Power Outputs Based on the Na ve Bayes Classifier and Kriging Models. *Energies* **2018**, *11*, 2982. [CrossRef]
50. Pérez, A.; Larrañaga, P.; Inza, I. Supervised Classification with Conditional Gaussian Networks: Increasing the Structure Complexity from Naive Bayes. *Int. J. Approx. Reason.* **2006**, *43*, 1–25. [CrossRef]
51. Naive Bayes Classifier—Machine Learning [Updated] | Simplilearn. Available online: https://www.simplilearn.com/tutorials/machine-learning-tutorial/naive-bayes-classifier (accessed on 29 August 2022).
52. Fix, E.; Hodges, J.L., Jr. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **1989**, *57*, 238–247. [CrossRef]
53. Sun, L.; Zhang, J.; Ding, W.; Xu, J. Feature Reduction for Imbalanced Data Classification Using Similarity-Based Feature Clustering with Adaptive Weighted K-Nearest Neighbors. *Inf. Sci.* **2022**, *593*, 591–613. [CrossRef]
54. Gao, L.; Li, D.; Liu, X.; Liu, G. Enhanced Chiller Faults Detection and Isolation Method Based on Independent Component Analysis and K-Nearest Neighbors Classifier. *Build. Environ.* **2022**, *216*, 109010. [CrossRef]
55. Luo, M.; Liu, B. Robustness of Interval-Valued Fuzzy Inference Triple I Algorithms Based on Normalized Minkowski Distance. *J. Log. Algebr. Methods Program.* **2017**, *86*, 298–307. [CrossRef]
56. Hu, L.Y.; Huang, M.W.; Ke, S.W.; Tsai, C.F. The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets. *Springerplus* **2016**, *5*, 1304. [CrossRef]
57. The Professionals Point: Advantages and Disadvantages of KNN Algorithm in Machine Learning. Available online: http://theprofessionalspoint.blogspot.com/2019/02/advantages-and-disadvantages-of-knn.html (accessed on 29 August 2022).
58. Theissler, A.; Thomas, M.; Burch, M.; Gerschner, F. Knowledge-Based Systems ConfusionVis: Comparative Evaluation and Selection of Multi-Class Classifiers Based on Confusion Matrices. *Knowl.-Based Syst.* **2022**, *247*, 108651. [CrossRef]
59. Molla, W.; Ambri, Z.; Karim, A.; Alemu, T. Investigation of Fault Detection and Isolation Accuracy of Different Machine Learning Techniques with Different Data Processing Methods for Gas Turbine. *Alex. Eng. J.* **2022**, *61*, 12635–12651. [CrossRef]
60. Oksuz, K.; Cam, B.C.; Kalkan, S.; Akbas, E. One Metric to Measure Them All: Localisation Recall Precision (LRP) for Evaluating Visual Detection Tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *8828*, 1–18. [CrossRef] [PubMed]
61. Seliya, N.; Van Hulse, J. Aggregating Performance Metrics for Classifier Evaluation. In Proceedings of the 2009 IEEE International Conference on Information Reuse & Integration, Las Vegas, NV, USA, 10–12 August 2009; pp. 35–40.
62. Sepúlveda, J.; Velastín, S.A. F1 Score Assesment of Gaussian Mixture Background Subtraction Algorithms Using the MuHAVi Dataset. *IET Semin. Dig.* **2015**, *2015*, 1–6. [CrossRef]
63. EMRA. *Procedures and Principles Regarding Technical Quality of the Electricty Distribution System*; EMRA: Ankara, Turkey, 2020.
64. Enerji Yönetim Sistemi | Inavitas (EMS). Available online: https://www.inavitas.com/tr/ (accessed on 15 August 2022).