An Intelligent Fault Detection and Classification Scheme for Distribution Lines Using Machine Learning

Balamurali Krishna Ponukumati  
School of Electrical Engineering  
KIIT Deemed to be University  
Bhubaneswar, India and  
Department of EEE  
Madanapalle Institute of Technology and Science  
Madanapalle, India  
pbmk20031993@gmail.com

Pampa Sinha  
School of Electrical Engineering  
KIIT Deemed to be University  
Bhubaneswar, India  
pampa.sinhafel@kiit.ac.in

Manoj Kumar Maharana  
School of Electrical Engineering  
KIIT Deemed to be University  
Bhubaneswar, India  
mkmfel@kiit.ac.in

A. V. Pavan Kumar  
Department of EEE  
Madanapalle Institute of Technology and Science  
Madanapalle, India  
pawanrao82@gmail.com

Akkenaguntla Karthik  
Department of EEE  
Madanapalle Institute of Technology and Science  
Madanapalle, India  
akarthik160@gmail.com

Received: 28 May 2022 | Revised: 8 June 2022 | Accepted: 9 June 2022

Abstract-The current paper focuses on the development and deployment of Machine Learning (ML) based algorithms for the classification and detection of different faults in the electrical distribution system. The methodology adapted using ML has higher computational accuracy than traditional computational algorithms. The parameters involved in developing ML for fault detection/classification are fundamental frequency, fault voltage, and current components at fault situations. During faults, the current and voltage waveforms consist of high-frequency transient signals. The Wavelet Decomposition (WD) technique is used to break down transient signals to obtain the required information. To investigate the performance of the ML-based algorithms, an IEEE 33 bus system is utilized, and a fault is generated in Matlab/Simulink environment. The methodologies used for fault detection and classification are K Nearest Neighbor (KNN), Decision Tree (DT), and Support Vector Machine (SVM). The performance of the designed algorithm is assessed by employing a confusion matrix, and the results demonstrated extraordinarily high accuracy.

Keywords-machine learning; wavelet decomposition; k nearest neighbor; decision tree; support vector machine

I. INTRODUCTION

The intricacy of power system networks presents many difficulties in distributing and preserving electricity. The primary aim of the protection system is safeguarding the overall system by identifying faults and preventing malfunctions to attain adequate power supply continuity [1]. Numerous domestic and commercial clients possess equipment that is susceptible to electrical interference [2]. A good protection system must ensure system stability by protecting against all types of failures at all network locations. A system built with the current technology must be trustworthy and dependable [3]. Phasor Measurement Units (PMUs) can detect a variety of typical distribution system events, including load and equipment switching transients and short circuit faults [4, 5].

Convolutional Neural Networks (CNNs) are implemented in a variety of fields, including object recognition and language recognition. CNNs are getting increasingly popular, notably in power system management and control, due to their capacity and feasibility in identifying and classifying issues that arise in
power systems [6-11]. The authors of [12] employed Support Vector Regression (SVR) to enhance the effectiveness of short-term estimation techniques based on historical electricity demand information. Authors in [13] suggest fault categorization in a distribution network by implementing an Adaptive Neuro-Fuzzy Inference System (ANFIS), and authors in [14] discuss rapid fault recognition and categorization in electrical distribution networks employing a hybrid of fuzzy logic and singular entropy theory. The application of ML is not restricted to limited fields. It is also applied to the agriculture sector in detecting plant diseases [15]. The necessity for a highly precise and resilient waveform deviation tracking system in high-voltage power systems makes research into its categorization and identification using deep learning techniques a popular subject [11, 16, 17]. In [18], a hybrid strategy of Gated Recurrent Unit (GRU) with Discrete Wavelet Transform (DWT) involved features employed as inputs. Fault recognition and categorization in microgrid employing CNNs was proposed in [19].

Fast Fourier Transform (FFT), DWT, S-Transform (ST) and statistical techniques are some examples [20-23] of ML algorithms used for fault detection and classification in distributions lines and microgrids. Usually, the extraction of the feature components is specific to the system setup and might involve continued attempts, limiting generalization [24]. DWT is widely used in the feature extraction step [25]. However, the feature extraction section and functioning approaches are unique throughout respective deployments. For example, as a functional technique for defect identification, [26] employs DWT calculated singular entropy including a set of parameters. This approach works perfectly, however, the margin to greater fault impedance wasn’t addressed. In [27], DWT is employed to collect features from current signals, while Particle Swarm Optimization (PSO) is used to select the best wavelet configuration. For fault categorization and diagnosis in a microgrid, authors in [10, 25] propose ML and in [18] propose deep learning-based methods. To categorize the problem, the wavelet transform is utilized to extract data from current signals, which are then classified using SVM [28] and DT [12].

Techniques such as ANFIS, Fuzzy Interference Systems (FIS), Artificial Neural Networks (ANNs), Feedforward Neural Networks (FNNs), CNNs, and conventional fault identification with relay systems are some efficient ML techniques. ML focuses on extracting knowledge from a large set of information and provides conventional ways for problem-solving. It really can deal with a wide range of information and be employed in various situations. Furthermore, by integrating more data for training, the accuracy of ML algorithms improves. Controllers like Fuzzy logic and FIS are straightforward and use “if-then” relationships to handle problems with uncertainty. They are a generally strong candidate even though they require huge training data sets. Incorporating ANFIS assists with tunable characteristics, decreases computational time, gives quick convergence and responsiveness, but the difficulty arises when large computation is expected, which is removed by adopting ML. The ANNs, FNNs, and CNNs help predict the specific fault type and are simple to apply. Their use is simple, involving merely a few characteristics to be adjusted, and have many applications in real-world situations, but the training procedure is rather hard for high-dimension challenges. The gradient-based back propagation approach for non-linear discrete classification tasks provides a locally optimal solution. Slow convergence is provided by ANNs in the back propagation method. The convergence is determined by the initial value of weight restrictions linked to the system. Conventional relay systems use mechanical parts to detect a fault. These mechanical parts need periodical maintenance, otherwise their performance will be degraded and it should match the ratings of the transmission or distribution line.

Figure 1 depicts the flowchart of the suggested approach, where \( V_a, V_b, \) and \( V_c \) are the three-phase voltages. \( I_a, I_b, \) and \( I_c \) are the three-phase currents. \( V_{\text{COFF}} \) and \( I_{\text{COFF}} \) are Voltage and current coefficients, and the angles between \( V \) and \( I \) are considered as inputs. These inputs are fed to the developed ML detection and classification algorithm which continuously monitors the status of inputs until a fault occurs.

![Flow chart of the proposed methodology.](image)

**II. FAULT DETECTION ANALYSIS**

WD approach is considered to detect the fault and classify it. The WD method was also utilized in [29, 30] to extract the waveform but that’s a 3-level decomposition, whereas this approach uses a 6-level WD as shown in Figure 2. The fault/disturbance signal is routed via Low Pass Filters (LPFs) and High Pass Filters (HPFs) until the necessary signal for analysis is achieved. The fault/disturbance signal is decomposed at 6 levels by passing via an LPF D and an HPF A. At level 1, the signal is decimated by 2 and D1 and A1 signals are obtained. D2 and A2 are obtained by passing the A1 signal through the second stage of LPF and HPF. Again, A2 passes through filters until level 6 and the final obtained signals are D6 and A6. As a result, significant information is derived from the initial signal which has to be evaluated.

**III. PHASE LOCKED LOOP**

A Phase-Locked Loop (PLL) is a regulatory technique which produces a signal as output with a phase-matched to that of an input signal. Its block diagram is illustrated in Figure 3. In order to keep the inlet and outlet phases locked, the input and output frequencies must be steady. Using the input voltage, \( V_{\text{abc}}, V_{\text{dc}}, \) and \( V_{\text{q}} \) are obtained. \( V_{\text{d}} \) and \( V_{\text{q}} \) are voltage values near the Point of Common Coupling (PCC) in the d- and q- axes respectively. The output is an approximation of the voltage phase. The suggested PLL architecture for fault identification is depicted in Figure 3.
In comparison to the block diagram in Figure 2, the error signal \( e(t) \) relates to the waveform deviation, while the Proportional-Integral (PI) regulator and integrator correlate to the loop filter and Voltage Controlled Oscillator (VCO) respectively. The three-phase voltages \( V_a, V_b, \) and \( V_c \) that are sent into the PLL are first changed to \( \alpha\beta \) quantities \( V_\alpha \) and \( V_\beta \) using Clarke’s transformation. A schematic of a PQ-PLL block is demonstrated in Figure 4. The quantities are then transferred into a reference frame. Considering the proximity of the \( \alpha\beta \) quantities to the reference frame, an error signal is created. The error signal is sent into a PI regulator, which controls the error to zero. The input signals are in phase with the frame of reference once the error is zero. When the PLL’s output angle is in phase with the current of phase A, the error is exactly zero. When there is a transient in the system, the error signal deviates from zero. The magnitude and frequency of the deviation will vary depending on the features of the transient.

**IV. MACHINE LEARNING**

As the name implies, ML is a phrase that is used to define the field of learning algorithms. An ML model will generally have three different phases: training, validation, and real-time implementation. During the training process, the system is fed through test data with known correlations so that it can detect patterns and create predictions within the same set accurately. The validation stage is intended to assess the overall system performance utilizing similar but disjoint data, as well as previously identified responses. In terms of model training methods, ML is typically classified into four major methodologies: supervised, unsupervised, semi-supervised, and reinforced learning. Models which are trained on a sample with knowing labelled outcomes are referred to as supervised learning. A list of classification algorithms that have been frequently used in the literature follows.

**A. Support Vector Machine**

During appropriate post-fault diagnosis, a knowledge-based approach that is based on SVMs is offered. SVMs are used as an intelligence tool to detect the faulty line originating first from the substation and to determine the distance from that. SVMs are compared to radial-based neural networks in datasets that depict diverse transmission system challenges.

**B. Decision Tree**

The DT’s design is simple, and we can easily follow the tree structure to describe how to make a conclusion. The vast scope of power system DTs has lately been discovered to be extremely effective in applications such as online dynamic safety evaluation, stability to transients, and islanding identification.

**C. K Nearest Neighbor**

The KNN approach is a safe, supervised ML approach utilized to tackle classification and regression issues. Faults may be detected and recognized in distance protection using the KNN method.
D. Bayesian Learner

The Bayesian classifiers, also called Naïve Bayes are statistical classifiers that estimate the probability of class membership using supervised learning techniques.

E. Sequential Minimal Optimization (SMO)

Training an SVM necessitates the solving of a large quadratic programming optimization problem. SMO divides this massive quadratic programming issue into a bunch of smaller quadratic programming problems.

F. Logistic Regression

If the dependent variable is binary, logistic regression is the best regression approach to use. Like other regression studies, it is a statistical approach. The dependent variable Y in logistic regression has values of 1 and 0 for the outcomes of interest.

G. Linear Regression

The relation between both the input variables (x) and output variables (y) for linear regression is generally stated as a linear equation of the form y = a + bx. As a result, the aim of linear regression would be to determine the values of coefficients a and b, where a is the interception and b is the line’s slope.

V. RESULTS AND DISCUSSION

A. The Considered ML Techniques

To detect the fault and classify it in the distribution network, the testing data are evaluated using SVM and DT algorithms.

- If there are little training data (less than 1000), support vector methods are often good because they are highly regularized.
- SVMs have better production results than ANNs, while they avoid the problem of over-fitting. One can use the kernel function in the SVM to solve complex problems. They can effectively handle data with higher dimensions that are linear inseparable.

For data preprocessing, WD is employed. Matlab is used to obtain various operating and fault data. A total of 2746 input samples are considered, with 70% are used for training and 30% for testing the model’s performance. The training data are used to train a model that can predict how to respond to incoming new data. Table I shows the accuracy rates of the tested algorithms. The cubic SVM algorithm yielded the best accuracy rate of 87.1%.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine tree</td>
<td>82.20%</td>
</tr>
<tr>
<td>Medium tree</td>
<td>79.40%</td>
</tr>
<tr>
<td>Coarse tree</td>
<td>55.20%</td>
</tr>
<tr>
<td>Linear SVM</td>
<td>52.50%</td>
</tr>
<tr>
<td>Quadratic SVM</td>
<td>78.20%</td>
</tr>
<tr>
<td>Cubic SVM</td>
<td>87.10%</td>
</tr>
<tr>
<td>Fine Gaussian SVM</td>
<td>63.60%</td>
</tr>
<tr>
<td>Medium Gaussian SVM</td>
<td>67.20%</td>
</tr>
<tr>
<td>Coarse Gaussian SVM</td>
<td>51.80%</td>
</tr>
</tbody>
</table>

B. Case Study

The developed ML algorithms are tested on the IEEE 33 bus system (Figure 5). The suggested fault detection and classification technique was tested using the IEEE 33 radial distribution scheme.

A fault is developed to analyze the performance of the system. Figure 6 illustrates the occurrence of an L-G fault between 0.2s and 0.4s in a three-phase voltage waveform. WD is used to assess the fault voltage signal. Figure 7 depicts the L-G fault voltage signal of a multiresolution WD. The Db6 wavelet is used for producing a 6-step decomposition. At every frequency spectrum, the rebuilt variants of every element, along with the approximation and records of the unique signal, are represented.

![Fig. 5. The IEEE 33 bus system.](image)

![Fig. 6. Three-phase voltage waveforms under L-G fault occurrence.](image)

C. Confusion Matrix

The confusion matrix is the most often used method for evaluating the accuracy and correctness of a classification algorithm. It also helps identifying areas where the classifier has underperformed. It can also help understanding what the classification model is doing correctly and what sorts of errors it is producing. In Figure 8, the faulty classification confusion matrix is shown. It is evident that the suggested system can classify various faults, and the prediction rate is high. Figure 9 shows the True Positive Rate (TPR) and False Negative Rate (FNR) data and the TPR value is high, which says that the model can detect faults accurately.

Precision is a measure of how accurate a classifier is. A low accuracy might also be indicative of a massive amount of False Positives. Figure 10 shows the Positive Predictive Value (PPV)

![Fig. 7.](image)
and False Discovery Rate (FDR) values. It is clear that the PPV has higher value than FDR. The rows of the confusion matrix correspond to the anticipated class, while the columns are commensurate to the actual class. The diagonal cells are commensurate to suitably categorized observations. The off-diagonal cells are commensurate to observations that were wrongly categorized.

Fig. 7. WD analysis on LG fault.

Fig. 8. Faulty classification confusion matrix.

Fig. 9. TPR and FNR values of data.

VI. CONCLUSION

The utilization of ML algorithms for fault categorization and identification on a three-distribution network was extensively investigated in this study. The created systems used instantaneous current and voltage data as inputs for fault detection and categorization. This paper described a waveform-based fault classification method for electrical distribution systems. WD was employed for bringing out the features from recorded wave data. The massive volume of wave data processing activity was treated effectively by carefully spreading the computing burden. ML modules identified and classified the faults. The fault classification and detection results demonstrate the efficacy of ML in categorizing and identifying faults in less time than other traditional relaying systems. The systems make use of the natural capacity for pattern classification and identification. For distance protection, ML technology is strong, effective, reliable and precise.

REFERENCES


