Identification of Type of a Fault in Distribution System Using Shallow Neural Network with Distributed Generation

Saurabh Awasthi, Gagan Singh and Nafees Ahamad

Department of Electrical and Electronics & Communication Engineering, DIT University, Dehradun, 248001, India

*Corresponding Author: Saurabh Awasthi. Email: saurabh.awasthi@dituniversity.edu.in

Received: 29 September 2022   Accepted: 12 December 2022

ABSTRACT

A distributed generation system (DG) has several benefits over a traditional centralized power system. However, the protection area in the case of the distributed generator requires special attention as it encounters stability loss, failure re-closure, fluctuations in voltage, etc. And thereby, it demands immediate attention in identifying the location & type of a fault without delay especially when occurred in a small, distributed generation system, as it would adversely affect the overall system and its operation. In the past, several methods were proposed for classification and localisation of a fault in a distributed generation system. Many of those methods were accurate in identifying location, but the accuracy in identifying the type of fault was not up to the acceptable mark. The proposed work here uses a shallow artificial neural network (sANN) model for identifying a particular type of fault that could happen in a specific distribution network when used in conjunction with distributed generators. Firstly, a distribution network consisting of two similar distributed generators (DG1 and DG2), one grid, and a 100 Km distribution line is modeled. Thereafter, different voltages and currents corresponding to various faults (line to line, line to ground) at different locations are tabulated, resulting in a matrix of $500 \times 18$ inputs. Secondly, the sANN is formulated for identifying the types of faults in the system in which the above-obtained data is used to train, validate, and test the neural network. The overall result shows an unprecedented almost zero percent error in identifying the type of the faults.

KEYWORDS

Distribution network; distributed generation; power system modeling; fault identification; neural network; renewable energy systems

1 Introduction

There exist two ways of producing electrical power, i.e., via centralized generation and distributed generation (DG) system [1]. In the former one, electrical energy is generated at some large power stations whereas in distributed generation system (DG) the generation is primarily done through
renewable energy resources near the load centres via small units [2]. A distributed generator offers many advantages when compared to a conventional system like enhancement in rural electrification, reduced Transmission and distribution expenses, improved energy efficiency, better reliability & security, ease in installation, and lesser environmental constraints [3]. However, one of the major issues that originates due to DG system is that it alters the radial nature of the existing distribution network and cause power to flow in reverse direction. Further DG system also increases the fault current level and affects the relay operation in the event of a fault. This problem becomes even more challenging if the fault prevails significantly for some duration. Thus, identifying the type of a fault and its location becomes highly important as any delay will not only propose a threat to the healthy components but also results in reduced power generation. Beside it can also imposes hindrance while carrying out switch-gear schemes [4].

In recent years, many authors have proposed several optimization algorithms and artificial intelligence tools like graph theory, genetic algorithm, artificial neural network (ANN), problem and fuzzy logic for solving optimization problems and identifying the fault type and its location in a distribution network when incorporated with DG [5–27]. Even so, the problem still requires extensive and reliable methods for better results as many of the methods were either conducted with lot many assumptions or under a particular set of conditions which cannot be generalized to all types of distributed generation system. The need of extensive and reliable methods is important to obtain better result as it is extremely critical to clear the fault as soon as it occurs in a distribution network. This will require an accurate data regarding the type of the fault and its location. An incorrect location or an incorrect type of the fault will deteriorate the conditions further and may result in power outage as well as hefty financial loss.

A method wherein the direction of current is taken as input for fault identification was proposed by Calderaro et al. [8]. However, the method requires umpteen number of sensors for localizing fault even for smaller sections. Javadian et al. [9] proposed a multi-layer perceptron orthodox method of splitting the system into different radial zones in which the protection of each such zone is to be carried out using a circuit breaker. Hence the scheme requires major alteration for better results. Most of such methods focus on measurements of current and voltage corresponding to different DGs, Pérez et al. [14] and Dashtdar et al. [15] proposed different methodologies to identify the fault location using support vector mechanics, Neural networks, smart meters, etc. Both these methods showed good accuracy in identifying the location of faults. However, the proposed methods were not able to identify the type of fault accurately.

As most of the methods proposed in the past, showed good accuracy in identifying the location of the fault, however the reliability in identifying the type of fault still requires a major improvement. The proposed method emphasizes more on the fault type identification, as for clearing the fault determining its type is equally important as its location. A small delay in identifying the type can cause the whole system to malfunction and ultimately power outage for a longer duration which means no further generation within that timeline. This would have a negative impact on the economy and growth. The major advantage with the proposed scheme is not only it helps in clearing the fault precisely but also ensures quicker response that improve the power optimization and thereby the economic growth.

In this paper, the emphasis has been laid out on identifying the type of the fault only using a shallow Neural Network. The proposed scheme uses a real-time distributed generation system with two distributed generators and a utility grid. An equivalent model is then made using MATLAB Simulink. Such systems are difficult to protect in the event of a fault as the conventional relaying cannot be easily adapted for a distributed generation system in a distribution network. The shallow Artificial Neural
Network requires data for training, validation, and testing which is to be collected through artificially creating short circuit faults at various locations at 2 km each within the distribution network. The data so recorded would then be formulated in a desired matrix form for ANN training and validation. Out of the total recorded data, around 70% would be used for ANN training, while 15% each is kept for validation, and testing. The results corresponding to training, validation and testing and relative percentage error between the actual values and calculated values would then be compared to depict the effectiveness of the scheme. The results of the conducted study would also be compared with [20] to give a better insight into the successful training, testing & validation.

Table 1 discuss the various important techniques and their pros and cons for identifying fault and their location.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Author</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive techniques</td>
<td>El Halabi et al. [16]</td>
<td>The proposed scheme was based on the current phase jump that can quickly determine the fault direction</td>
<td>No information regarding the type of the fault can be found out</td>
</tr>
<tr>
<td></td>
<td>Ma et al. [17]</td>
<td>A novel adaptive non-pilot current protection scheme utilizing steady state fault current is presented</td>
<td>It focusses mainly on the Primary and backup protection scheme with no information about fault type and location</td>
</tr>
<tr>
<td>Impedance based</td>
<td>Adwole et al. [18]</td>
<td>Provides accurate result for fault section identification</td>
<td>Cannot be used for identifying fault type</td>
</tr>
<tr>
<td></td>
<td>Hosseinimoghadam et al. [19]</td>
<td>Accuracy has improved over other methods of locating fault</td>
<td>The effect of changing fault impedance is ignored</td>
</tr>
<tr>
<td>Artificial neural network</td>
<td>Rezaei et al. [20]</td>
<td>Able to locate fault accurately when occurred in a particular zone. The accuracy in locating the fault within a section is close to almost 99%</td>
<td>Manual cross checking is required if fault occurs in a different zone. No information provided on fault type identification</td>
</tr>
<tr>
<td></td>
<td>Zayandehroodi et al. [21]</td>
<td>The method uses two stage radial basis function neural network</td>
<td>The method does not focus on fault type identification</td>
</tr>
<tr>
<td>Support vector</td>
<td>Agrawal et al. [22]</td>
<td>Faster response and high accuracy with higher number of Distributed generators</td>
<td>The training and testing for less distributed generators requires improvement</td>
</tr>
</tbody>
</table>

(Continued)
1.1 Types of the Faults

In a distribution system, faults are the unexpected deviations from the standard operating conditions. There may exist different reasons for their occurrence like a short circuit between physical contact, trees falling on the overhead lines, physical contact of birds or animals with the lines, etc. These faults may exist for a while or may remain until diagnosed and cleared. Accordingly, they are termed temporary faults and permanent faults, respectively. The temporary faults may eventually become permanent faults if not cleared in time. The permanent fault may occur mainly due to the failure of cable insulation failure, wires falling on the earth or objects coming in contact with the lines.

There are mainly four types of short circuit faults that may occur in a distribution system which are shown in Table 2. All these faults are characterized by an increase in current, reduction in voltage, frequency, and power factor.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Type of the fault</th>
<th>Causes</th>
<th>Occurrence/Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Single line to ground (SLGF)</td>
<td>Line falling on to ground, trees falling on to line.</td>
<td>70%-Least severe</td>
</tr>
<tr>
<td>2</td>
<td>Line to line fault (LLF)</td>
<td>Short Circuiting between lines due to wind, animal contact.</td>
<td>15%-Severe</td>
</tr>
<tr>
<td>3</td>
<td>Double line to ground fault (DLGF)</td>
<td>Tree falling on to line and cause short circuiting between two phases.</td>
<td>10%-Severe than SLGF &amp; LLF</td>
</tr>
<tr>
<td>4</td>
<td>Three-phase to ground fault (LLLGF)</td>
<td>Conductor comes in contact with other phases, failure of the equipment, Falling of towers, trees onto conductor.</td>
<td>5%-Most severe</td>
</tr>
</tbody>
</table>

Of all these faults, SLGF is the most common but least severe and LLLF is the least common but most severe fault and requires immediate attention. Due to the occurrence of these faults, the generators must be disconnected from the system via circuit breaker which in turn causes a power outage and deteriorates the power quality.
1.2 Artificial Neural Network for Fault Classification

Once the data gets stored and transformed into the required matrix form, the next step is to train the artificial neural network. This calculation will be performed by an artificial neural network program. To identify the type of fault, it becomes necessary to successfully train the neural network through collected data and once the training is completed, testing and validation are to be carried out.

1.2.1 Artificial Neural Network (ANN)

It is defined as an information processing epitome that works the same way as our biological nervous systems. It includes millions of neurons that are complicatedly inter-connected and operate in a coalition with one another to solve specific problems. The major application of an artificial neural network is primarily in the field of pattern classification. Fig. 1 shows an elementary artificial neuron.

![Figure 1: An artificial neuron](image)

As learning in organic systems includes alterations in the synaptic network which can be realized among different neurons, the include synapses like input which upon multiplication by signal strength or weight derives a mathematical relationship that further causes neurons to activate. The other identity functions determined the output of the ANN which many times depends upon a particular threshold.

One of the most important internal functions of an ANN network is the activation function which is defined as a weighted sum, i.e., Sum of inputs x when multiplied by their proportionate weights W_{ji} is represented in Eq. (1) as:

\[ A_j(\bar{x}, \bar{w}) = \sum_{i=0}^{n} x_i w_{ji} \]  

Similarly, the output function is a sigmoidal function that is given by Eq. (2):

\[ O_j(\bar{x}, \bar{w}) = \frac{1}{1 + e^{A_j(\bar{x}, \bar{w})}} \]  

The Backpropagation algorithm is mainly used for this purpose. There is one hidden layer used with as many as 34 neurons in it. The number of neurons in the hidden layer is selected based on the minimum RMS error encountered simultaneously in all three phases during validation and testing.

The parameters which affect the overall accuracy of an ANN are inter-relation among different neurons, activation functions, and the method via which weights can be determined. The network accepts inputs through input layer neurons whereas the network output is delineated through output layer neurons.

Some intermediate hidden layers are also present within the structure. Since this algorithmic program uses the concept of supervised learning, hence input-output samples are required to be analysed one after another. A simple neural network diagram is shown in Fig. 2.
The content of the paper is organized as:

**Section 1:** Introduction.

**Section 2:** Methodology.

**Section 3:** System configuration and modelling.

**Section 4:** Neural Network designing for fault classification.

**Section 5:** Result and Discussion.

**Section 6:** Conclusion.

2 **Methodology**

The protection procedure mainly consists of three major steps, System Modelling, Data recording, and ANN designing for training and testing as shown in Fig. 3.

![System modelling](image)

**Figure 3:** Protection procedure block diagram

2.1 **System Modelling**

It involves collecting data on the proposed system for which analysis is to be carried out. The next step is to define the input variables and their rated value which is then followed by making a corresponding model of the system on MATLAB Simulink as shown in Fig. 4.

![System Modelling](image)

**Figure 4:** System modelling
The modeling can also be done on other platforms such as ETAP. In the proposed scheme an elementary distribution network is modeled using two distributed generators and a utility grid on MATLAB Simulink only. The rating of all the sources and corresponding transformers matches a real-time system.

### 2.2 Fault Data Collection

The purpose of this step is to run the load flow and short circuit test and record the corresponding data in matrix form [20]. Normally during steady-state operation, the net current via each source must be corresponding to the existing load only. However, during faulty conditions, this current exceeds the load current value and eventually becomes several times that of its rated value. To differentiate between the healthy and faulty conditions and to measure the degree of severity of these faults, the net current through all the sources is compared. The block diagram for offline calculation is shown in Fig. 5.

![Figure 5: Fault data collection](image)

Identification of a given fault is attributed on account of the high value of pre- and post-current before and after the fault has occurred. The normalization value of current for ten different types of faults can be calculated using Eq. (3):

\[
I_{normal} = \frac{2I - I_{max} - I_{min}}{I_{max} - I_{min}}
\]

(3)

where \(I_{min}\), \(I_{max}\) are the minimum & maximum value of fault current for a specific type of fault.

### 2.3 ANN Formulation

After successfully recording the steady state and fault data, the next step is to define a suitable Neural Network structure followed by training, validation and testing as shown in Fig. 6. To open the correct unhealthy section, verification of the direction of input current via breaker is obligatory.

![Figure 6: ANN formulation](image)

### 3 System Configuration and Modeling

The system considered here is a distribution network incorporated with two DGs along and a grid which is equivalently modeled in MATLAB Simulink as shown in Fig. 7. Here two distributed generators having star-grounded configuration is interconnected to the grid via delta-grounded star distribution transformers.
The grid is rated at the voltage level of 132 kV at the power frequency of 50 Hz. The SC MVA level of the grid is taken at 100 MVA with reactance to resistance ratio of 7. A transformer of rating 100 MVA 132/66 kV is stepping down this voltage at the distribution level.

A V-I measurement toolbox has been connected with this transformer to measure and record the voltage & current parameters. The distribution network of 100 km distance is terminated through two distributed generators of rating 100 kVA at 3.3 kV. The entire distribution network is split into 50 zones of 2 km each.

The corresponding Simulink model on MATLAB is shown in Fig. 8. A common scope to all three sources ensures measurement of three-phase current and voltage. A 10 kW 3-phase load is connected for extracting power. Since it is not practically possible to carry out the calculations on a real system, hence modeling of the system becomes mandatory.

3.1 Fault Data Collection

Once modelling of the system is done, the next step is to carry out load flow and short circuit tests at various locations. The reference point chosen here is the utility grid. In recent years, many authors have proposed several techniques to conduct load flow and short circuit tests [28–30]. To carry out these tests, fault data corresponding to each 2 km is taken. In this case with generators 1 and 2, a total of 50 locations have been taken into consideration for carrying out fault analysis. A 3-phase fault box is used to conduct and record data corresponding to the load flow and short circuit tests. The load flow and magnitude of voltage and current of all three sources can be recorded via scope data. The simulation and sample time are selected as 0.2 & 0.00002 s, respectively. This will ensure 500 values of voltage & current for all the three sources to be recorded.

The instantaneous value of all these quantities can be checked from the scope results which indicates all the quantities are equally displaced from each other by 120 degrees as shown in Fig. 9.
Figure 8: Simulink model of the test system

Figure 9: Scope results: steady state
The magnitude and phase angle of current and voltage for all three sources under steady state are already stored in the workspace and have eventually been used for training purpose. A similar approach has been adopted to carry out the short circuit tests and likewise, the already saved data has been extracted from the workspace.

Similarly, short circuit analysis can be done at this location using the same procedure and data can be collected from the scopedata1 stored in the workspace. For example, three phase fault (ABC-G) can be performed by selecting all three phases-A,B,C and Ground variable in the 3-phase fault box. Fig. 10 shows the scope results corresponding to a 3-phase fault at DG2.

![Figure 10: Scope results for a 3-phase fault at DG2](image)

The short circuit test for other marked locations such as 2, 4, 6, 8 km, and so on can be executed. The data corresponding to all such locations for all types of the fault has been recorded and transformed into a matrix form for training purposes.

4 Neural Network Designing for Fault Classification

After recording the short circuit test data, the next step is to design a shallow ANN that can be trained and validated using the faults data at different locations in the desired form. The steps involved are as follows:

- Designing the structure of the Neural Network in terms of number of Neurons in the hidden layers, Input & Output Layers.
- Artificial Neural Network training through the data collected earlier.
- Validation & testing once the training is completed.
- Identification of fault and its location using the validation process.

The input variables are phase voltage and current of the Utility Grid, Distributed generators 1 & 2 and hence there are 18 inputs (9 voltage & 9 current). As a result, the size of the input matrix is 500 \( \times \) 18.
The output corresponding to the recorded input can be any one of the four possible outcomes (SLGF, LLF, LLGF, and LLLGF) which makes the target matrix size $500 \times 4$. Out of the total recorded data, around 70% has been used for ANN training, while 15% each is kept for validation and testing. Thus, the total number of input and output are 18 & 4, respectively.

The toolbox itself splits the recorded data for validation, training, and testing. By default, around 70% (samples) have been used for training purposes. Around 15% which are samples of the total have been kept reserved for deriving generalized values. Once the generalization does not improve anymore, the training would stop. As it does not have any impact on training, thus the result obtained would have independent characteristics. Here around 15% (samples) have been taken for this purpose.

It may be noted that the allocation for training, validation, and testing can be altered as per the need. Based on the minimum error convergence criteria. The number of neurons in the hidden layer is selected as 34 as shown in Fig. 11.

![Figure 11: Neural network structure for test system](image)

The identification of a given type of fault can be done using the coding as shown in Table 3.

<table>
<thead>
<tr>
<th>Fault type</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steady-state</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>3-phase fault</td>
<td>1 1 1 0</td>
</tr>
<tr>
<td>A to ground</td>
<td>1 0 0 1</td>
</tr>
<tr>
<td>B to ground</td>
<td>0 1 0 1</td>
</tr>
<tr>
<td>C to ground</td>
<td>0 0 1 1</td>
</tr>
<tr>
<td>A-B (LL fault)</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>A-C (LL fault)</td>
<td>1 0 1 0</td>
</tr>
<tr>
<td>B-C (LL fault)</td>
<td>0 1 1 0</td>
</tr>
<tr>
<td>A-B to ground</td>
<td>1 1 0 1</td>
</tr>
<tr>
<td>A-C to ground</td>
<td>1 0 1 1</td>
</tr>
<tr>
<td>B-C to ground</td>
<td>0 1 1 1</td>
</tr>
</tbody>
</table>
5 Results and Discussions

The outcomes of the proposed methodology on the test system involving two DGs and a grid together forming a distribution network can be studied in Figs. 12–19.

![Figure 12: Predicted vs. actual training & validation-Phase A](image1)

![Figure 13: Predicted vs. actual training & validation-Phase B](image2)

It can be seen from Fig. 12 that the predicted vs. actual training and validation for phase A is a straight-line characteristic showing linearity in prediction and actual results.
Although there are very few outliers, the characteristics are linear showing excellent relation between the predicted value and actual value.

A similar type of results was obtained for phases B, C, and ground also with very few outliers which can be studied through Figs. 13–15, respectively.
Figure 16: RMSE vs. neurons for phase A

Figure 17: RMSE vs. neurons for phase B
Figure 18: RMSE vs. neurons for phase C

Figure 19: RMSE vs. neurons for Ground

Overall, the accuracy of the scheme is close to 100 percent for all four outputs. Except for one of the values, the curve is a perfectly straight line showing the linearity between predicted value and actual value.
The root means square error for phases A, B, C and ground indicate close proximity between the error during training and validation.

These errors are minimum at various neurons in the hidden layers such as at 10 neurons, 15 neurons, 22 neurons, 34 neurons and so on. However, the error is minimum for all the phases at 34 neurons in the hidden layer. And hence the number of neurons in the hidden layer is chosen as 34.

The RMSE for training and validation for all the phases and ground comes out to be approximately less than 0.02 percent. These errors against different neurons for phases A, B, C, and D can be seen in Figs. 16–19, respectively.

The maximum error encountered from zero to 60 neurons is still well below 0.2% during starting transient period. Once it gets settled the error never exceeds 0.02% for A, and 0.05% for phases B, C, and ground.

RMSE is the main parameter that depicts the effectiveness of pattern recognition. The RMS error for different stages corresponding to phases A, B, C, and D are given in Table 4.

**Table 4: RMS error (%) in different stages**

<table>
<thead>
<tr>
<th>S.No.</th>
<th>State</th>
<th>Phase A</th>
<th>Phase B</th>
<th>Phase C</th>
<th>Ground</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Training</td>
<td>0.0029</td>
<td>0.0058</td>
<td>0.0046</td>
<td>0.0047</td>
</tr>
<tr>
<td>2</td>
<td>Validation</td>
<td>0.0059</td>
<td>0.0157</td>
<td>0.0158</td>
<td>0.0276</td>
</tr>
<tr>
<td>3</td>
<td>Testing</td>
<td>0.0045</td>
<td>0.0136</td>
<td>0.0115</td>
<td>0.0170</td>
</tr>
</tbody>
</table>

During training the error for phases A, B, C and ground is coming out as 0.0029%, 0.0058%, 0.0046%, and 0.0047%, respectively. Similar kinds of results are coming for the validation and testing stage as well.

The result of the conducted study have also been compared with one of the best fault identification methodologies as proposed by [20] for a similar distribution network incorporated with 2 distributed generators for a network of 50 km in Table 5 to depict the effectiveness of the study.

**Table 5: Result comparison**

<table>
<thead>
<tr>
<th>Fault type</th>
<th>MSE found in N rezaei experiment [20]</th>
<th>MSE found in the proposed work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 phase with Gen 1 &amp; Gen 2</td>
<td>0.0033</td>
<td>0.000031</td>
</tr>
<tr>
<td>Line to line between Gen 1 and Gen 2 and Rf = 5 P. U.</td>
<td>0.7755</td>
<td>0.000029</td>
</tr>
<tr>
<td>Double line to ground with Gen 1 and Gen 2</td>
<td>0.1691</td>
<td>0.00245</td>
</tr>
<tr>
<td>3 phase with Gen 1, Gen 2 and Gen 3</td>
<td>0.1394</td>
<td>0.0086</td>
</tr>
</tbody>
</table>

All the other methods explained earlier laid out emphasis only on clearing the fault without giving much importance to the speed and accuracy. The method proposed by [20] focuses mainly on how the system can be restored into its normal operation, however the accuracy of the scheme was merely 1/10th of the proposed work.

When training, testing, and validation are conducted, these errors all return 0, demonstrating great accuracy in recognising the type of issue.
6 Conclusion

This paper presented an artificial intelligence-based methodology to identify the type of fault. A Shallow Neural Network has been used for this purpose. The Neural Network have been trained through the fault data which has been recorded through MATLAB Simulink. To record fault data, a distribution network with two distributed generators is incorporated. The results of the proposed method as shown in Table 4 along with the linear characteristics between actual and calculated values (Figs. 12–15) concludes that the proposed work helps in mitigating the protection challenges encountered in distribution network systems when incorporated with distributed generation system.

The technical indicator for successful testing and validation is MSE (Mean Square Error). For a successful result, its value must be very close to zero. In this case, it comes out to be nearly zero percent as shown from the RMSE characteristics (Figs. 16–19) depicting 100% accuracy in identifying the type of the fault.

The results of the proposed scheme have also been compared with one of the past research works [20] in Table 5 and it has been found that the proposed methodology yields far better results irrespective of the size of the system. The accuracy of the proposed scheme comes out to be 0.000031, 0.000029, 0.0024, 0.0086 for single line to ground, line to line, Double line to ground, 3 phase fault which is very less in comparison to one of the schemes proposed earlier on a similar type of system.

Hence, it can be concluded that among all the artificial intelligence-based methods proposed to date, this scheme has shown the best results in identifying the type of fault in a distribution network with distributed generation system. The proposed scheme helps in restoring the system very quickly so that not only it prevents further damage to the healthy components but also ensures least interruption in power generation which would raise the per capita energy consumption and improves socioeconomic conditions of the country.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References


