IOSR Journal of Mobile Computing & Application (IOSR-JMCA)
e- ISSN: 2394-0050, P-ISSN: 2394-0042.Volume 7, Issue 6 (Nov - Dec 2020), PP 01-08
www.iosrjournals.org

Analysis of Data Aggregation Schemes for Smart Grid Communications

Usman Mohammed[1], G. M. Wajiga[2], M.K. Ahmed[3]
[1, 2]*(Department of Computer Science, ModibboAdama University of Technology, Yola, Nigeria)*
[3]*(Department of Mathematics, Gombe State University, Gombe, Nigeria)*
Corresponding Author: Usman Mohammed

***Abstract**: A lot has been said about different techniques of aggregating data in a smart grid communications. What is yet to be extensively explored is the concept of smart grid application in Nigeria's power sector. The smart grid can exchange data about current electricity status, pricing data and control commands in real-time. Due to these specific characteristics, the process of electricity's generation, transmission and distribution in the smart grid environment can be managed efficiently and reliably. However, smart grid technology may not be right for all power networks due to requirement for substantial resources. This work has studied the data aggregation scheme in smart grid communication, which is characterized by the review of previous studies that have proposed data aggregation structures for smart grid communications, and was evaluated using the benefit of proposed data aggregation scheme and its impact on smart grid structure as well as its potential limitations. The usage of neural networks was employed for the detection, classification and location of faults on transmission lines. The method employed made usage of the phase voltages and phase currents (scaled with respect to their pre-fault values) as inputs to the neural networks. To simulate the various faults model and to obtain the training data set, MATLAB R2015a was used along with the SimPowerSystems toolbox in Simulink. The performance of the model was analyzed using Mean Square Error (MSE). It was observed that the configuration for the chosen ANN was 6 – 10 – 5 – 3 – 1 and the number of iterations required for the training process were 37. It can be seen that the mean square error in fault detection achieved by the end of the training process was 9.43e-5 and that the number of validation check fails were zero by the end of the training process. More so, the configuration for the chosen ANN for fault location was 6-7-1 with 5 iterations required for the training process.*

I. Introduction

The world's electricity systems face a number of challenges, including ageing infrastructure, continued growth in demand, the integration of increasing numbers of variable renewable energy sources and electric vehicles, and the need to improve the security of supply and reduce CO2 emissions. Advancement in computing technology has led to the improvement of the performance and security of the power grid. Smart grid technology has been receiving a lot of attention from both industry and academia over the last few years. By providing bidirectional communication channels, the smart grid can exchange data about current electricity status, pricing data and control commands in real-time. Due to these specific characteristics, the process of electricity's generation, transmission and distribution in the smart grid environment can be managed efficiently and reliably. Consequently, the smart grid technology can reduce power consumption, lower energy costs, and bring much convenience to our work and daily life. Data aggregation usually involves the fusion of data from multiple intermediate nodes and transmission to the base station (sink)[9]. It attempts to collect the most critical data from the nodes and make it available to the base station in an energy efficient manner. A data aggregation scheme is energy efficient if it maximizes the functionality of the network. However, the design of efficient data aggregation algorithms is an inherently challenging task[10].

The aim of this research is to critically analyse the current data aggregation schemes for Smart Grid Communications. The work intends to identify the barriers to adoption of smart grid technologies in Nigeria.The Objectives of the Study are:

i. Identify the barriers to adoption of smart grid technologies in Nigeria.
ii. Compare the requirements for application of smart grid technology and existing network in Nigeria.
iii. Develop an artificial neural network model to detect fault occurrences.
iv. Monitor faults at various locations to improve distribution efficiency.
v. To detect and prevent meter bye-pass and energy theft.

II. Related Works

Several data aggregation paradigms have been proposed within a smart grid environment, while some were implemented, others remain as theoretical inputs in the field of computing. As a result, we are motivated to investigate these schemes and provide a systematic review of the recent studies. Some of these research are; Homogeneous tree aggregation with home gateways that was proposed by[8] which enables a resident to send a service request to nearby homes under the identity and the location privacy. It involved the proximity score calculation and communication phase of users. The major advantage of this scheme is its efficiency in service rate and obtained bandwidth. However, Insider attacks are not considered. Similarly, Grid aggregation within home area network (HAN) was proposed by [5] with emphasis on data privacy and security which involved an aggregator, users, and an off-line trusted third party. The major advantage of this scheme is the fact that it is efficient in terms of aggregation and batch verification, and also in ensuring data integrity. However, Energy cost, Identity privacy and location privacy are not considered. Also, Data aggregation with a key management centre (KMC) was proposed by [11] to achieve privacy of users (residential users). The major goal is providing error-detection, and fault tolerance using a secret key or certificate to carry out the authentication. The major advantages of this scheme include; Accuracy of aggregation, efficient in communication and computation overheads. Another approach by [1] presented a Multi-agent System platform that will allow agents to detect and report fault in Power Distribution System (PDS) with the minimum time delay is presented. Multi-agent System Engineering (MaSE) methodology was used to demonstrate how agents will be able to manage the complexity of PDS. The simulation of the framework was done using Java Agent Development Framework (JADE). However, many assumptions needed to understand its implementation.

The table 2.1 shows a summary of the data aggregation paradigms have been proposed by different researchers.

Table 2.1 Analysis of data aggregation paradigms

S/N	AGGREGATION SCHEME	BENEFITS	IMPACT ON GRID STRUCTURE	LIMITATIONS
1.	**Homogeneous Tree aggregation with home gateways:** Proposed by [8]	• Efficiency in service rate. • Communication between users.	Improve energy consumption, and latency	• Time synchronization between the nodes increases. • Data confidentiality issue.
2.	**Grid Aggregation within home area network (HAN)**: Proposed by [5]	• Efficient batch verification. • Ensures data integrity.	Timely, and accurate meter reading reporting.	• High energy cost. • Identity privacy not considered.
3.	**Chain aggregation with a key management centre (KMC)**: Proposed by [11]	• Ensures data accuracy. • Dynamic join and leave.	Low communication and computation overheads.	• Difficulty in implementation. • Data attacks are not given priority.
4.	**One-way ANOVA Aggregation:** Proposed by [2]	Guarantees accountability, confidentiality, and integrity	High scalability	Low resilience to link failures. High overhead to setup/maintain the aggregation structure
5.	**Cluster aggregators with smart meters**: Proposed by [4]	• Consider human factor on data. • Data security.	Energy efficiency and measurement using smart meters.	• No comparison with other schemes. • High energy cost.
6.	**Private stream aggregation:** Proposed by [11]	• Accurate and timely billing. • Enable nearby homes communication.	Distributed access control. Efficient aggregation throughput.	• Energy cost is not considered. • No comparison with other schemes
7.	**Grid Aggregation with the advanced metering infrastructure (AMI):** Proposed by [6]	Guarantees privacy, integrity, and availability.	Metering and querying Processes and Settlement process accuracy.	• No threat model presented. • No comparison with other schemes.
9.	**Linear Aggregation:** Proposed by [7]	Guarantees accountability, and integrity of data	Resilience in case of node mobility.	Low resilience to link failures.
10.	**Variance aggregation with authentication:** Proposed by [3]	High resilience to link failures.	• Periodic timing strategy. High scalability.	Low resilience in case of node mobility.

III. Methodology

The data collection was carried out to identify the load records including the voltage and current values at the TCN 132/33 kV, Gombe. The data sources include validated summary of suppressed demand data in the fault record, Load participant invoice, daily load MYTO allocation, fault level remedial measure and fault analysis data supplied by the Nigerian Electricity Regulatory Commission (NERC).

After data collection, it has been compiled, in order to receive the required information, such as:

i. Determination and classification of various faults locations.
ii. Aggregated load centres and population clusters with high fault rate.
iii. Shortcomings in existing power supplies and network limitations.

Data Pre-Processing

Voltage and current waveforms have been generated and were sampled at a frequency of 50 Hertz. The voltage and current samples of all the three phases are noted along with the corresponding pre-fault values. The samples after the occurrence of the fault on the phases are noted and samples before the occurrence of the fault, corresponding to the post-fault sample considered. Once this is done, the inputs to the neural network are the ratios of the voltages and currents in each of the phases before and after the occurrence of fault.

The Table 3.1 shows the voltage and current values that are scaled with respect to their pre-fault values and used as a part of the training set. In Table 3.1, Vx, Vy and Vz are the post fault voltage and current sample values and Va, Vb and Vc are the corresponding pre-fault values as illustrated earlier. The given table depicts the values for all the various types of faults and also during the no fault case. The fault has been simulated on a 10 km long transmission line at a distance of 2 km from the terminal A.

Table 3.1 V/I values with respect to Faults

Case	Fault Type	V_x/V_a	V_y/V_b	Vz/Vc	I_x/I_a	I_y/I_b	I_z/I_c
1.	A to Ground	0.62	0.97	1.04	1.68	0.50	0.87
2.	B to Ground	0.65	0.73	0.82	0.40	5.68	1.74
3.	C to Ground	1.25	0.91	0.79	1.49	4.50	1.70
4.	A to B	0.18	0.60	1.00	0.81	8.71	3.24
5.	B to C	1.00	0.55	0.34	1.00	7.81	1.11
6.	C to A	1.15	1.00	0.92	1.03	8.25	2.10
7.	A, B to Ground	0.27	0.58	0.90	2.16	6.81	3.47
8.	B, C to Ground	0.93	0.51	0.98	0.38	5.30	1.75
9.	A, C to Ground	0.31	0.43	0.49	1.86	5.90	4.57
10.	No Fault	1.00	1.00	1.00	1.00	1.00	0.99

More so, for each of the three phases, faults have been simulated at every 2 Km on a 10 Km long transmission line. Along with the fault distance, the fault resistance has been varied as mentioned earlier for each of the three phases with each of the different fault resistances as 0.25, 0.5, 0.75, 1, 5, 10, 25 and 50 ohms respectively). In each of these cases, the voltage and current samples for all three phases (scaled with respect to their pre-fault values) are given as inputs to the neural network. The output of the neural network is the distance to the fault from terminal A as shown in the table 3.2;

Table 3.2 Fault distance and fault resistance for Fault Location ANN.

Case	Fault Type	Fault Resistance = 20Ω		Fault Resistance = 60Ω	
		Fault Resistance (Ω)	Fault Location (KM)	Fault Resistance (Ω)	Fault Location (KM)
1.	A to Ground	5	5.49	5	6.56
2.	B to Ground	7	5.58	7	5.68
3.	C to Ground	10	7.12	10	4.50
4.	A to B	11	7.12	11	5.15
5.	B to C	12	7.30	12	5.40
6.	C to A	14	7.50	14	6.20
7.	A, B to Ground	16	8.20	16	8.50
8.	B, C to Ground	18	9.10	18	9.20
9.	A, C to Ground	20	9.50	20	9.50

IV. Programming Language Used

MATLAB (Matrix Laboratory) was chosen to carry out the simulation and analysis of the network and fault detection scheme. MATLAB is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. The work therefore, employs the use of MATLAB R2015a GUIDE, and Simulink from SimPowerSystem Packages.

The Studied Model Architecture

The system consist of a three phase source labelled A, B, C with a frequency of 50Hz and 11KV, a three phase breaker with 4 inputs and 3outputs, which is activated by an external source called a relay, and the V-I measurement unit which specify the type of fault and from which phases. The three phase RLC load control the load changes in the transmission line. The three phase fault activate the step function whenever the load changes with respect to step time. The relay has 3 inputs and one output which are compared using a Boolean operator AND, the output goes to the three phase breaker as faults occur. The Figure 3.1shows model of three phase transmission line.

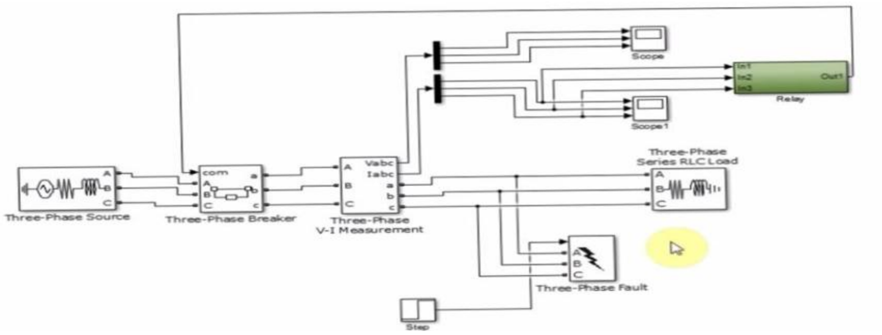

Figure 3.1 Model of three phase transmission line.

V. Results and Discussion

Evaluating the Performance of the Network

In the first stage which is the fault detection phase, the network takes in six inputs at a time, which are the voltages and currents for all the three phases (scaled with respect to the pre-fault values) for ten different faults and also no-fault case. Hence the training set consisted of a set of six inputs and one output in each input-output pair. The output of the neural network is just a yes or a no (1 or 0) depending on whether or not a fault has been detected. After simulations it has been decided that the desired network has one hidden layer with 10 neurons in the hidden layer. For illustration purposes, several neural networks (with varying number of hidden layers and neurons per hidden layer) that achieved satisfactory performance are tested and described. Figures 4.1 – 4.3 show the error performance plots of neural networks with 1 and 2 hidden layers respectively. The chosen network has been depicted in Fig 4.3

Fig 4.1 shows the training performance plot of the neural network 6-10-1 (6 neurons in the input layer, 1 hidden layer with ten neurons in it and one neuron in the output layer). It can be seen that the network did not achieve the desired Mean Square Error (MSE) goal by the end of the training process.

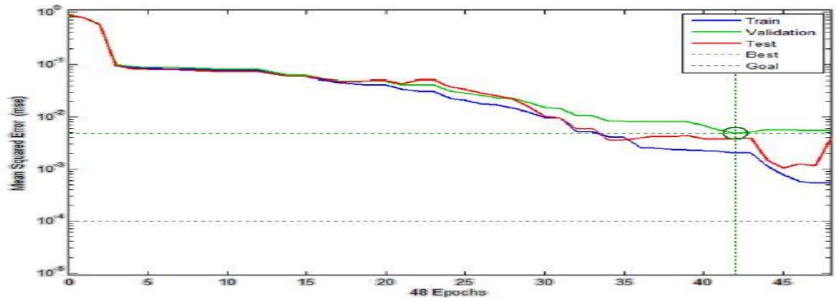

Figure 4.1 Mean Square error Performance (6-10-1)

Figure 4.2 shows the training performance plot of the neural network with 6-10-5-1 configuration (6 neurons in the input layer, two hidden layers with 10 and 5 neurons respectively and one neuron in the output layer). It is to be noted that the neural network could not achieve the MSE goal by the end of the training process.

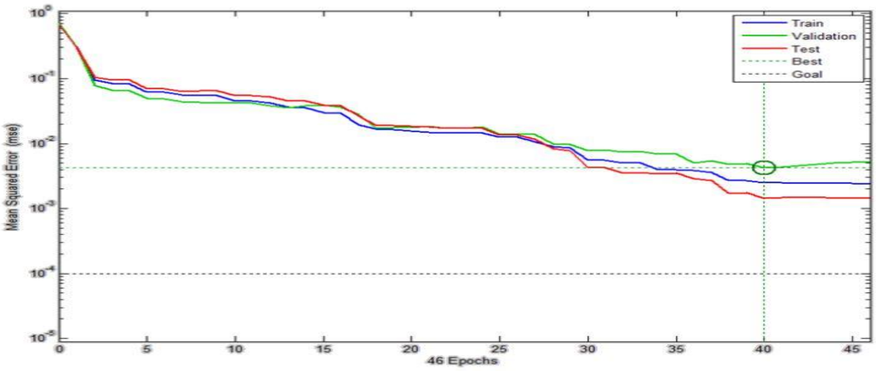

Figure 4.2 Mean-square error performance (6-10-5-1).

Figure 4.3 shows the training performance plot of the neural network with 6-10-5-3-1 configuration (6 neurons in the input layer, 3 hidden layers with 10, 5 and 3 neurons in them respectively and one neuron in the output layer).). It is to be noted that the neural network has achieved the MSE goal at Epoch 24 at the end of the training process.

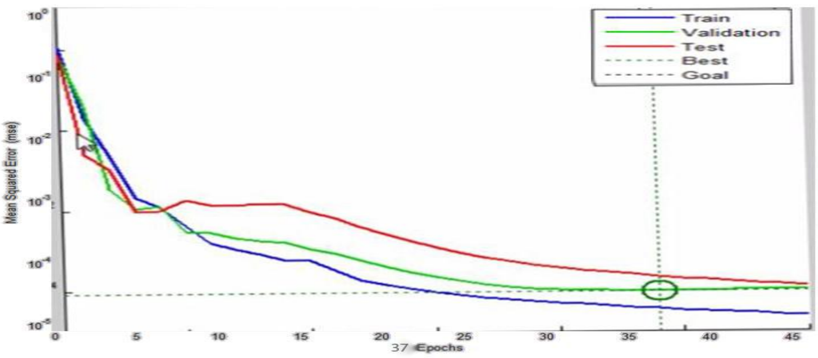

Figure 4.3 Mean-square error performance (6-10-5-3-1).

From the above training performance plots, it is to be noted that very satisfactory training performance has been achieved by the neural network with the 6-10-5-3-1 configuration (6 neurons in the input layer, 3 hidden layers with 10, 5 and 3 neurons in them respectively and one neuron in the output layer). The overall MSE of the trained neural network is way below the goal, and is 6.9776 e-5 at Epoch 37 at the end of the training process. Hence this has been chosen as the ideal ANN for the purpose of fault detection.

Testing the Fault Detection Neural Network

Once the neural network has been trained, its performance has been tested by plotting the best linear regression that relates the targets to the outputs as shown in Figure 4.4

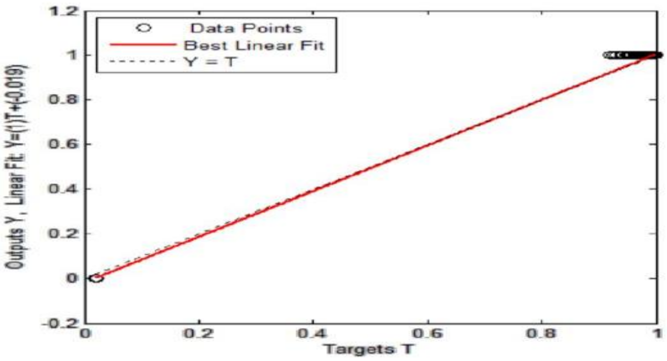

Figure 4.4 Regression fit of the outputs vs. targets for the network (6-10-5-3-1).

The correlation coefficient (r) is a measure of how well the neural network's targets can track the variations in the outputs (0 being no correlation at all and 1 being complete correlation). The correlation coefficient in this case has been found to be 0.99967 in this case which indicates excellent correlation.
The structure of the chosen neural network for fault detection is shown in Fig 4.5 with the input layer, hidden layers and the output layer labelled. It is to be noted that there are 6 neurons in the input layer, 3 hidden layers with 10, 5 and 3 neurons in them respectively and one neuron in the output layer.

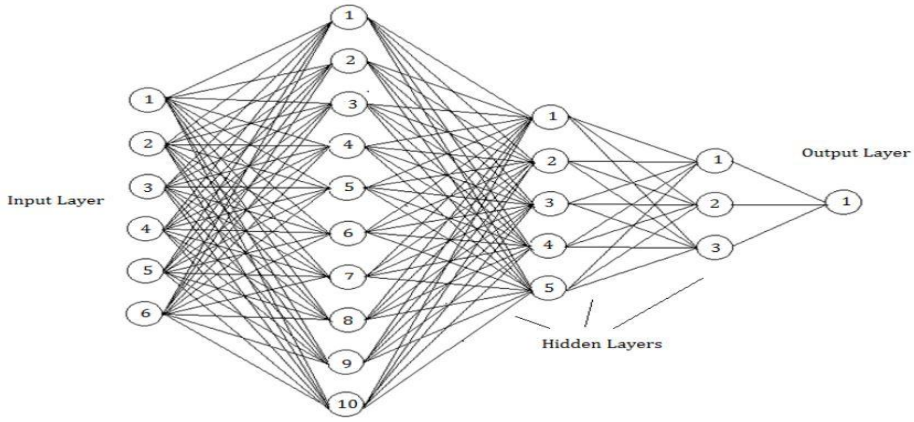

Figure 4.5 the Chosen Fault Detection ANN with (6 – 10 – 5 – 3 – 1)

Fault Location Neural Network

This section talks about the design, development and the implementation of the neural network based fault locators for the various types of faults. This forms the third step in the entire process of fault location after the inception of the fault. The following subsections deal with the faults and their error performance.
For each of the three phases, faults have been simulated at every 2 Km on a 10 Km long transmission line. Along with the fault distance, the fault resistance has been varied as mentioned earlier for each of the three phases with each of the different fault resistances as 0.25, 0.5, 0.75, 1, 5, 10, 25 and 50 ohms respectively). In each of these cases, the voltage and current samples for all three phases (scaled with respect to their pre-fault values) are given as inputs to the neural network. The output of the neural network is the distance to the fault from terminal A as shown in the table below;

Table 4.1 Percentage errors as a function of fault distance and fault resistance for Fault Location ANN.

Case	Fault Type	% Error Vs. Fault Distance Fault Resistance = 20 Ω			% Error Vs. Fault Distance Fault Resistance = 60 Ω		
		Fault Resistance (Ω)	Fault Location (KM)	% Error	Fault Resistance (Ω)	Fault Location (KM)	% Error
1.	A to Ground	5	5.49	0.163	5	6.56	0.52
2.	B to Ground	7	5.58	0.287	7	5.68	0.34
3.	C to Ground	10	7.12	0.121	10	4.50	1.70
4.	A to B	11	7.30	0.230	11	5.15	1.65
5.	B to C	12	7.50	0.311	12	5.40	1.60
6.	C to A	14	8.20	0.342	14	6.20	1.56
7.	A, B to Ground	16	9.10	0.355	16	8.50	1.40
8.	B, C to Ground	18	9.50	0.401	18	9.20	1.25
9.	A, C to Ground	20	10.00	0.482	20	9.50	1.15

Distance and Fault Resistance as two different cases have been considered (shown in adjacent columns), one with a fault resistance of 20 ohms and another with a fault resistance of 60 ohms. It is to be noted that the resistance of 20 ohms was used as a part of training data set and hence the average percentage error in fault location in this case is just 0.1646 %. The second case illustrates the same with a different fault resistance of 60 ohms which is relatively very high and is not a part of the training set. Hence, the performance of the neural network in this case illustrates its ability to generalize and react upon new data. It is to be noted that the average error in this case is just 0.878 % which is very satisfactory.

The neural network that performed reasonably well was presented along with its respective error performance. The chosen neural network is shown with all its characteristics depicted in detail. The test performance plots are obtained by simulating various faults on different phases at varying locations and calculating the error in the output produced by the Neural Network

Fig 4.6 plots the best linear regression fit between the outputs and the targets of the neural network with 6 neurons in the input layer, 7 neurons in the hidden layer and 1 neuron in the output layer (6-7-1). The value of the correlation coefficient r in this case is found to be 0.99924 which is by far the best and the closest to one.

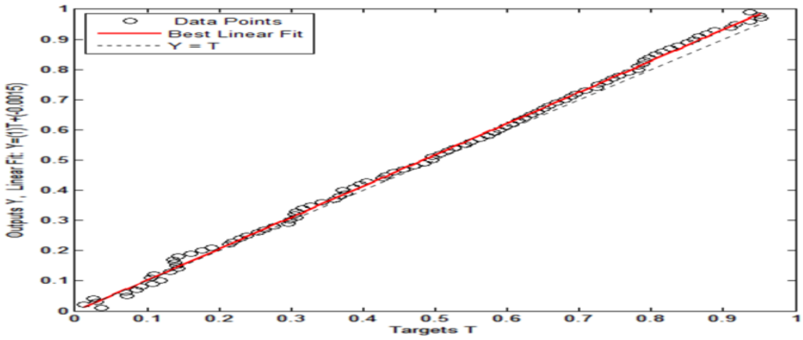

Figure4. 6 Regression fit of the outputs versus targets with (6-7-1).

Fig 4.7 plots the mean-square error as a function of time during the learning process and it can be seen that the learning achieved the best MSE is about 0.0005056.

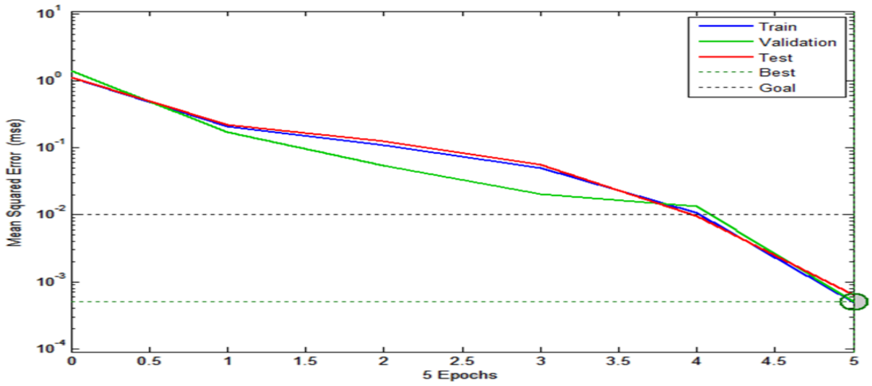

Figure 4.7 Mean-square error performance of the network with (6-7-1).

Figure 4.8 shows the plot of fault location against the percentage error in locating the fault position along the transmission line. The plot indicates that the best fault locations with minimal errors are 2.7KM, 4.7KM and 4.2 KM respectively.

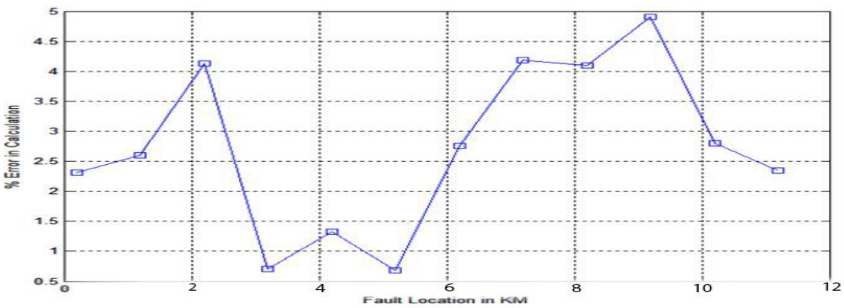

Figure 4.8 Percentage Error and Fault in KM plot.

Fig 4.9 shows the structure of the chosen ANN with 6 neurons in the input layer, 7 neurons in the hidden layer and 1 neuron in the output layer (6-7-1).

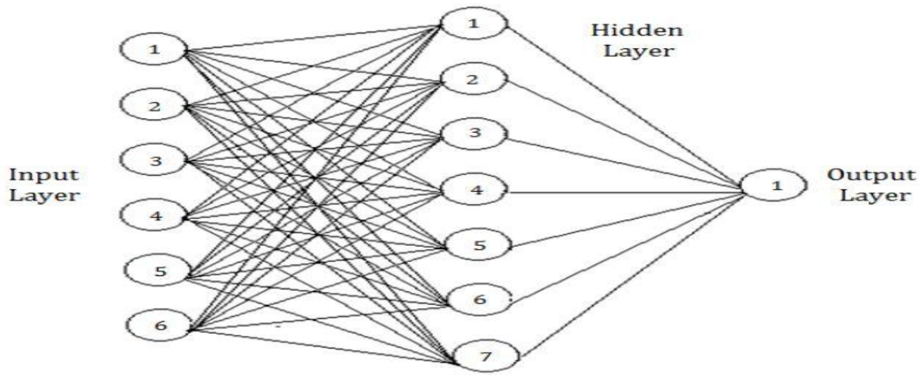

Figure 4.9 Structure of the chosen ANN with configuration (6-7-1).

VI. Conclusion

The traditional power grid can no longer keep pace with the ever increasing demand for electrical energy. Hence, multi directional power system flow (Smart) network is required through the incorporation of computerized and automated systems to gather, analyze and respond to information timely. Neural Networks are indeed a reliable and attractive scheme for an ideal transmission line fault location scheme especially in view of the increasing complexity of the modern power transmission systems.Back Propagation neural networks are very efficient when a sufficiently large training data set is available and hence Back Propagation networks can be employed to improve distribution infrastructure in Nigerian grid. The deployment of smart grid technologies in Nigeria remains the 'way forward' to bring Nigeria out of the dark where it has remained for years. There is an imperative integration of smart grid technologies into to the power flow for optimum power performance.

References

[1]. Ahmed M. K**,** AliyudaA andBute M. S. (2017). Multi-Agent Based Monitoring and Control of Power Distribution System. In Proceedings of the 13th International Conference of Nigeria Computer Society.

[2]. Chen, L., Lu, R., Cao, Z., AlHarbi, K. and Lin, X. (2016)."MuDA: Multifunctional data aggregation in privacy-preserving smart grid communications, "Peer-to-Peer Networking and Applications, vol. 8, no. 5, Sep 2015: 777–792

[3]. D. He, N. Kumar, and J.-H. Lee, (2016) "Privacy-preserving data aggregation scheme against internal attackers in smart grids," Wireless Networks, vol. 22, no. 2, pp. 491–502. [Online]. Available: http://link.springer.com/10.1007/s11276-015-0983-3

[4]. Deng, R., Xiao, G., Lu, R. and Chen, J. (2016). "Fast Distributed Demand Response with Spatially and Temporally Coupled Constraints in Smart Grid,"IEEE vol. 18, no. 3,

[5]. Fan, C. I., Huang, S. Y.and Lai, Y. L. :(2012). "Privacy-Enhanced Data Aggregation Scheme Against Internal Attackers in Smart Grid," IEEE Transactions on

[6]. Gong, Y., Cai, Y., Guo, Y., and Fang, Y. (2016) "A Privacy-Preserving data Aggregation Scheme for IBDR in the Smart Grid," IEEE Transactions on Smart Grid, vol. 7, no. 3, pp. 1304–1313, [Online]. Available: http://ieeexplore.ieee.org/document/7069275/

[7]. H. Wang, S. Zhang, and D. He, (2016) "Balanced anonymity and traceability for outsourcing small-scale data linear aggregation in the smart grid," IET Information Security, [Online]. Available: http://digital- library.theiet.org/content/journals/10.1049/iet-ifs.2016.0150

[8]. Liang, X., Zhang, K., Lu, R., Lin, X. and Shen, X. (2016) "EPS: An Efficient and Privacy-Preserving Service Searching Scheme for Smart Community,"IEEE

[9]. Rajagopalan, R. and Varshney, P. K., (2006) "Data aggregation techniques in sensor networks: A survey "Electrical Engineering and Computer Science. Paper 22. P 5-20: [Online]http://surface.syr.edu/eecsSyracuse University.

[10]. Shen, H., Zhang, M. and Shen, J. (2017). Efficient Privacy-Preserving Cube-Data Aggregation Scheme for Smart Grids. In:IEEE Transactions on Information Forensics and Security, 12 (6), pp. 1369 – 1381.

[11]. Shi, Z., Sun, R., Lu, R., Chen, L., Chen, J. and Sherman X. (2015) "Diverse Grouping-Based Aggregation Protocol With Error Detection for Smart Grid Communications,"IEEE Transactions on Smart Grid, vol. 6, no. 6. pp. 2856–2868, [Online]. Available: http://ieeexplore.ieee.org/document/7169604/.