Classification of Shunt Type Faults using Wavelet Transform and Convolutional Neural Network

A Dissertation submitted in fulfillment of the requirements for the Degree
of

MASTER OF ENGINEERING
in
Power Systems

Submitted by

Harkamaldeep Kaur
801742010

Under the Guidance of
Ms. Manbir Kaur
Associate Professor, EIED

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

2019

Electrical and Instrumentation Engineering Department
Thapar Institute of Engineering & Technology, Patiala
(Declared as Deemed-to-be-University u/s 3 of the UGC Act., 1956)
Post Bag No. 32, Patiala – 147004
Punjab (India)

DECLARATION

I hereby certify that the work which is presented in dissertation entitled, "**Classification of Shunt Type Faults using Wavelet Transform and Convolutional Neural Network**", in partial fulfillment of the requirements for the award of the degree of **Master of Engineering in Power Systems**, submitted to Electrical & Instrumentation Engineering Department of Thapar Institute of Engineering & Technology (Deemed to be University) is as authentic record of my own work carried under the supervision of **Ms. Manbir Kaur.** It refers others researcher's work which are duly listed in the reference section. The matter contained in this dissertation has not been submitted, neither in part nor in full to any other degree to any other university or institute except as reported in text and references.

Harkamaldeep Kaur

Place: Patiala
Date: 15.07.2019

Harkamaldeep Kaur
Roll No.: 801742010

It is certified that the above statement made by the student is correct to the best of my knowledge and belief.

Manbir Kaur

Date 15.7.2019

Ms. Manbir Kaur
Associate Professor
Electrical & Instrumentation Engineering Department
Thapar Institute of Engineering & Technology, Patiala

ACKNOWLEDGEMENT

With great pleasure and privilege, I wish to express my heartfelt sense of gratitude and Ms. Manbir Kaur Associate Professor, EIED, Thapar University, Patiala for her patient guidance and support throughout this report. I found this guidance valuable and more importantly, her supportive and motivating approach.

I am also thankful to Dr. R.S.Kaler, Professor & Head, EIED as well as Dr. Nitin Narang, Professor & PG Coordinator for the needed support and motivational approach. The paucity of words does not compromise for extending my thanks to all the staff members of EIED, Thapar University, my all family members and friends whose uninterrupted love, inspiration and blessings helped me in completing this research report.

Lastly, I would like to thank my parents for their years of unyielding love and encouragement. They have always wanted the best for me and admire their determination and sacrifice. I am also thankful to the researchers whose work has been consulted, utilized and cited in my dissertation.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

.

ABBREVIATIONS

ANN	Artificial Neural Network
ANNFC	Artificial Neural Network Fault Classifier
ANFIS	Adaptive neuro-fuzzy inference system
BP	Back-propagation
CNN	Convolutional Neural Network
CT	Current Transform
DWT	Discrete wavelet transform
ELM	Extreme learning machine
FT	Fourier Transform
GA	Genetic algorithm
GNN	Graph neural network.
LG	Line to ground
LL	Line to line
LLL	Three phase short-circuit
LLG	Double line to ground
LLLG	Three phase short circuit to ground
MRA	Multi-Resolution analysis
PCA	Principal Component Analysis
PNN	Probabilistic Neural network
PT	Potential Transform
SVM	Support Vector Machine

NOTATIONS

e_l	Error signal at the output of neuron l
d_l	Desired response for neuron l
y_l	Functional signal appearing at the output of neuron l
$\xi(n)$	Instantaneous sum of error squares
N	Time step
w_{li}	Synaptic weight
$\Delta w_{li}(n)$	Correction applied to the weight
$v_l(n)$	Induced local field
$\varphi_l(.)$	Activation function
b_l	Bias applied to neuron l
$x_l(n)$	i_{th} element of input vector
$o_m(n)$	K_{th} element of the output vector
W	Weight of the network
d_{kp}	Desired value for the k^{th} output and the p^{th} pattern
o_{kp}	Actual value of the k^{th} output and the p^{th} pattern
N	Number of entries in training set
W	Total number of parameters (weight and bias) of the network

ABSTRACT

Power system fault detection and its classification play an important role in protection of AC systems. In this study, non-stationary fault patterns those include three phase voltages and current waveforms are obtained for various types of shunt type of faults on a long transmission model using Simulink. Fault location, fault resistance, and distance are considered as the key parameters to study their influence on fault patterns. There are three methods of developing mathematical models for fault classification problems; one: quantitative method; two: qualitative method and three: data-driven method. As the dimensions of the system increases, it is getting more complex to develop mathematical model that can capture the dynamic behaviour of system. Artificial neural network using Levenberg Marquardt has been explored to solve the problem of recognition of faults. The pattern recognition by Levenberg Marquardt somewhat lacks generality and the selection of topology is quite tedious task. On the other hand, it is important to differentiate the fault signal from the disturbances in voltage and current waveforms due to transient disturbances. The other data driven method use feature extraction techniques. Owing to the advantage of the wavelet transform (WT) technique selection of variable size window proportion to frequency of signal processing is proposed to differentiate the fault signal from transient disturbance signal. The noise in the experimental result gives rise to non-zero wavelet coefficient during the steady-state. This has been improved by removing the unwanted noise by selecting proper filter such that fault-induced transient remain retained. The fault signal data has been transformed into informative data using WT. Convolutional neural network (CNN) has been explored with the training set of informative data to solve the problem of fault classification. The results obtained from ANN and CNN are compared to illustrate the capability of CNN

Index Terms: Artificial Neural Network, Convolutional Neural Network, Deep learning, Fault Classification, Feature Extraction, Shunt type Faults, Wavelet Transform

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND

Electric faults occur randomly in a transmission or distribution system and their intensity is different for different types of faults. These faults are the greatest pitfall to the continuity of electricity supply. So diagnoses of fault in a power system is necessary for clearing faults that are mainly occur in a power transmission or distribution system. Thus, to identify and isolate faults rapidly, a well-coordinated system with protection must be provided so that the disruption and damage caused to the power system can be minimized. As number of protection devices are installed in the electrical power system. For e.g. lines are protected by the protection relays which are installed at the end & beginning of the electrical relays to detect electrical faults & switched some off selectivity at the fastest possible. So protection relays for the protection of line generates the message to central alarms (which may be central control centre) where the operator of electrical power system has to draw the conclusion where the location of fault can be.But the task of switching off faults or disconnecting the faults from the electrical power network is important one. In the control centre of electrical power system the task of operator to analyze the alarms received & this task might be difficult because of different reasons. No direct information is obtained on what type of fault has occurred and where, but rather information is obtained in the control centre is about the switches and automatic protection devices to be operated in response to the fault. It is the job of a control engineer, based in the control centre, to infer the cause of a fault, from the string of messages received. Normally for a one off fault, a control engineer can make a relatively swift and accurate diagnosis; however there are times, such as stormy weather conditions, when alarm activity is very high and this stretches and sometimes even exceeds the human ability to cope with the sheer volume of information. Multiple and/or interacting faults can occur at these times along with unforeseen problems in the protection mechanisms. These can severely reduce the speed of diagnosis and hence the overall efficiency. In these circumstances there is an apparent need for an automatic, computer based system which can be used to assist the control engineer[1]. Such a system needs to process switching messages as they arrive with an aim to indicate the component(s) involved and the type of fault which has occurred. It has to operate in real time, carrying out a diagnosis in only a few seconds, and must be able to deal with all conditions including the occurrence of multiple faults or protection problems.

There aremainly two types of faults those occur on power system;symmetrical and unsymmetrical faults. Majority of faults are not balanced three phase faults in electrical system, but they are unbalanced in nature. In addition to this, faults can also be classified as the shunt faults and series faults. In the power system analysis under faulty situations, it is essential to make a distinction between the fault types to ensure the best solution for the power system analysis. The severity of the various types fault varieswith the location of short circuit[2], the path followed by fault current, the impedance of the system and its voltage level.

1.2 METHODS TO IDENTIFY FAULTS:

The extended power system and its application requires some modifications in suitable techniques for the classification of faults in a transmission system to avoid damages to the system and also to increase the efficiency of the power system. Therefore fault classification is done with three techniques which are given below in the form of figure 1.1

- Conventional techniques
- Hybrid techniques
- Modern techniques

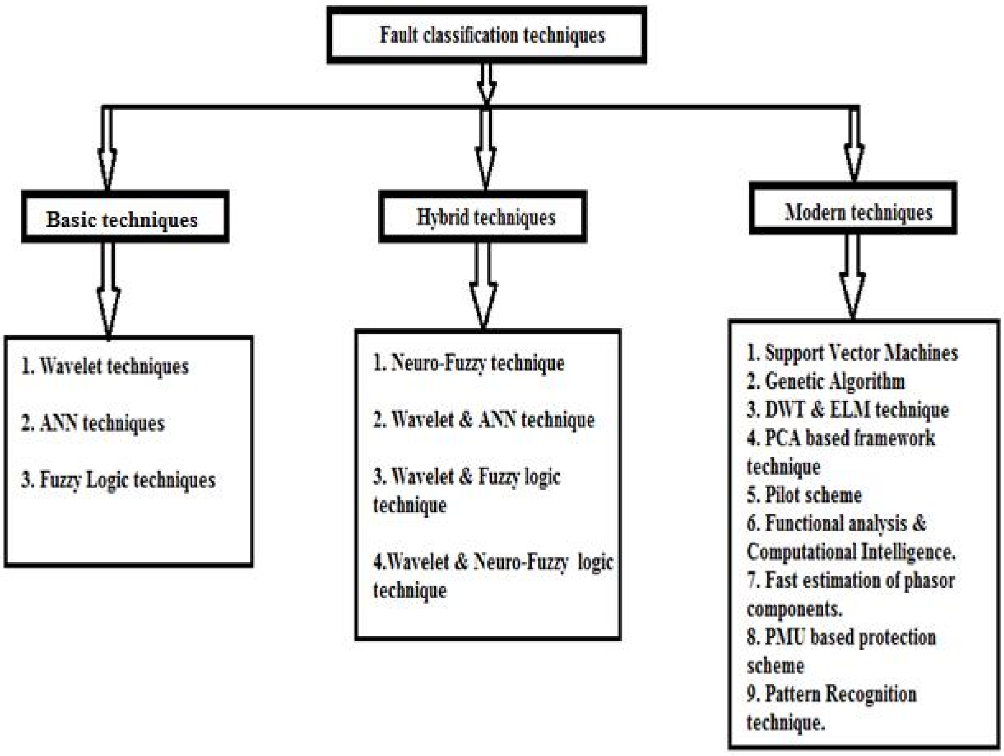


Fig.1.1 Tree diagram of fault classification techniques

1.3 LITERATURE SURVEY

The literature survey is widely classified on the basis of three techniques for the fault classification which are Basic techniques, Hybrid techniques and the modern techniques. In basic techniques in section 1.3.1, first artificial techniques, then fuzzy logic and then the wavelet techniques are discussed. From these techniques wavelet technique is the most efficient discrimination method which is used to extract faulty features for the classification of faults. Then hybrid methods in section 1.3.2, two techniques (ANN and fuzzy, ANN and wavelet transform,and Fuzzy and wavelet transform) are integrated to compensate the limitation of one technique with more strength in other. But out of all, the technique integrated with wavelet is more effective. And the last modern techniques in section 1.3.3, (SVM, GA, PCA, PNN, ELM, etc.) are used now-a-days which have less computational time with higher accuracy to achieve better classification of faults.

1.3.1 Basic techniques: Artificial Neural Network has high robustness and ability to learn from the incomplete & unforeseen data of inputs.The concept of multilayer neural networks is to solve various problems by using supervised learning techniques which has been presented[3].By using neural network backpropagation (BP) algorithm, the fault detection, location and classification in a transmission line system have been explored to implement the proposed network and algorithms , nominal-πmodel of overheadAC lines was considered which model 110 kV line connecting two cities (GEDAREF and EL FAU) with length of line 145 km. Data which is generated was used for single line to ground faults, two phase faults and double line to ground faults. Then neural networks with multi-layers istrained with the trainingdata to classify the fault type and configuration of the network can be seen in figure 1.2.

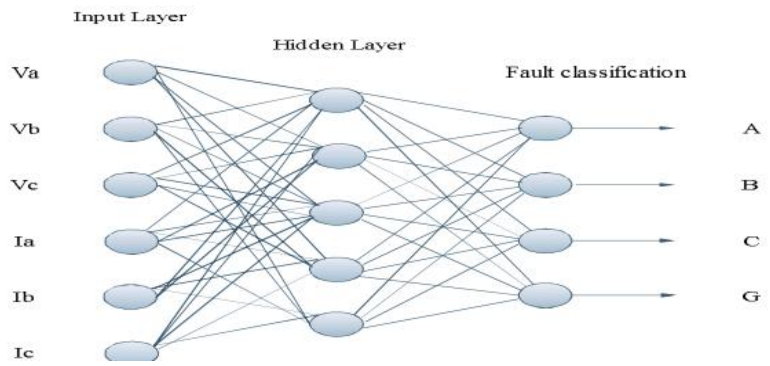


Fig. 1.2 Back Propagation neural network for fault classification

The concept of fault classification based onANN in underground cable and overhead line has been discussed using the Levenberg Marquardt algorithm as supervised algorithm[4]. Fundamental component of voltages and currents measured at one end of the transmission line and the effects of zero sequence currents had been used in fault classification & results obtained show that classification was done with higher rate of accuracy.

Generally, useful information of fault signals is obtained during transients which are naturally dynamic signals due to sudden changes of faults. Therefore, the transient should be considered while identifying faults. To analyse these signals, the wavelet transform is the most efficient approach. A fault classification technique in a transmission line based on Wavelet Transforms for high speed protective relaying has been presented[5] and three samples of line currents have been taken to carry out this classification. The required calculations were very simple as single-level signal decomposition was used to analyze the three line currents. The algorithm used for this approach was fast, secure and provided excellent results in classifying various types of faults.Then a new algorithm to phase selection based on wavelet transforms has been presentedandthe technique was based on using sharp transitions generated on the faulted phase [6].The Four-level signal decomposition was used to analyze the three phase voltages. There was a significant improvement that had been achieved compared to the previous phase-selection schemes regarding speed of operation, computational burden, and performance. Then wavelet multiresolution analysis (MRA) using daubechies eight (D-8), a new method for classification of faults based has been come intopicture.The summation of third-level output of MRA detail signals of current in each phase extracted from the original signals were used as the criterion for the analysis [7]. The absolute value of any two summations (S_a,S_b, S_c) are not equal and are always much higher than the absolute value of the 3rd summation, then it is a L-L-G fault. If S_{min}= $|S_C|$ &$\ll$ $|S_a|$ or $|S_b|$, then it is a L-L-G fault involving phases a, b and ground. The various faults considered in the analysis were L-G, L-L-G, L-L, L-L-L faults.

Discrete Wavelet Transform was used to extract the hidden factors from the fault signals by performing decomposition at different levels for detection, classification and location of transmission line faults [8]. It consist of three phase current signals feed to discrete wavelet transformation filter to decompose them to series of wavelet components which implies a particular octave frequency band containing more detailed information and these are applied to a high pass and a low pass filter to get a level 1 high frequency coefficients for fault detection.Then a new algorithm for fault location and classification using wavelet based on

Clarke's transformation has been developed and mother wavelet db4 was considered most accurate as compared to other mother wavelets with fastest time and smallest error detection [9]. The proposed algorithm could distinguish internal and external faults.

The concept of fuzzy-logic based approach for digital distance protection has been presented to classify faults which are included single phase to ground, two-phase, two-phase to ground[10].In this concept, algorithm was based on the angular differences among the sequence components of the fundamental current as well as on their relative magnitudes during fault.Another concept based on fuzzy logic for fault classification in a transmission line has been developed [11] in which three line currents measurement have been taken to implement the approach.

1.3.2 Hybrid Techniques: Combination of two or more techniques are used to make the system more efficient and to improve the classification tasks. These techniques have advantage of adaption to change, capability to deal with incomplete information, flexibility etc. Here the aim of hybrid technique is to take a technique that has weakness in a particular property & combine it with a technique that has strength in that same property.

The concept of identifying various faults in unbalanced distribution power system using discrete wavelet transform and fuzzy logic has been developed [12]. The proposed approach was fully effective in classifying all ten types of faults and for any possible combination of different power system parameters. Another concept of fault location estimation using hybrid technique combining generalized neural network and wavelet transform has been developed[13]. To extract the features of faulty current signals, wavelet transform was used in terms of standard deviation. Obtained features were used as an input to the GNN (Graph Neural network) model for estimating the location of fault in a given transmission systems. Then results were compared with ANN and found more accurate.Then a new concept of Artificial Neural Network based Fault Classifier (ANNFC) using Discrete Wavelet Transform (DWT) for classification of distinctive faults on three phase transmission line has been developed [14] and using discrete sets of data obtained from wavelet analysis taking Db8 as mother wavelet & addition of fifth level detail coefficients of fault transients both for relay terminal and far end terminal to train ANNFC for fault classification. Another concept based on ANFIS with wavelet transforms in neutral non-effectively grounded distribution system to classify the faults has been discussed [15]. The wavelet transform technique was used to obtain transient currents after faults occur. The fault identifiers were defined on the basis of statistic characteristic of transients in different shunt types of faults. The fault identifiers characterized the traits of fault type and show different disciplinarian in different fault types. These were fed as input into the three ANFISs to identify the type of fault.

1.3.3. Modern Techniques:

In modern techniques, a concept based on SVM for a seriescompensated line has been developed [16].The samples of three line currents and zero sequence current were taken as inputs to the SVM to classify faults.Then the fault classification in power system using hybrid approach consisting of usingwavelettransformationandSupportVectorMachine (SVM) has been developed [17]. The voltage transient information was extracted using DWT. The normalized wavelet energy of post-fault voltage and normalized energy of the post-fault currents were used as the input to the classifier.Another method using SVM with GA for fault classification in TCSC compensated transmission line was compared with existing SVM based methods [18] and higher classification accuracy had been achieved.The improved accuracy was achieved by changing the architecture and input of the classifier. Genetic Algorithm (GA) was used to search globally optimum value of SVM parameters.

Then a new approach based on principalcomponent analysis (PCA) and probabilistic neural network (PNN)to classify faults came into picture [19] and the approach was compared with pattern recognition approaches & found this method overcome previous problems related tofault classification.Another concept of classification of power system faults using Voltage Concordia Pattern Feature Aided PNN [20] which described a method in which the three phase sending end voltage was used to obtain the Concordia pattern followed by feature extraction and classification of fault using probabilistic neural network.Compared to other methods, evaluating Concordia pattern and extraction of features from this technique involved less computation.

Based on concept of ELM, classification of faults in series compensated overhead transmission line has been developed [21]. To extract features from the current signal, DWT was used& decompose the signal. Selected features using best selection method were then fed as an input to the extreme learning machine for fault classification. The results show that the proposed technique was robust, fast in learning and classifies the fault very accurately. Another advanced technique for fault detection & classification in transmission lines, using unsupervised feature learning and convolutional sparse auto encoder[22] which automatically learnt information from the signals of voltage and current in the form of dataset. The feature vectors are generated by Convolutional feature mapping and mean pooling.

Table 1.1Metaphor of different fault classification techniques.

S.No.	NAME OF THE APPROACH	TECHNNIQUE USED	SIMULATION TOOL USED	COMPLEXITY
1.	Wavelet approach	Wavelet transform, DWT, Clarks transformation	PSCAD MATLAB/Simulink, ATP, MATLAB wavelet toolbox	Medium
2.	Artificial neural network	Backpropagation (BP) algorithm,Levenberg Marquardt algorithm	MATLAB / SIMULINK	Complex
3.	Fuzzy logic based	Fuzzy logic	PSCAD	Simple
4.	Neuro-Fuzzy	Adaptive Network-based Fuzzy Inference ,wavelet transform ,Genetic Neuro Fuzzy System	PSCAD, ETAPS	Complex
5.	Wavelet & ANN approach	Wavelet transform, Neural network	SIMULINK,Wavelet toolboxes.	Medium
6.	Wavelet & Fuzzy approach	Wavelet transform , fuzzy logic	PSCAD/EMTDC, MATLAB wavelet toolbox	Simple
7.	Support Vector Machine approach	SVM, wavelet transform	PSCAD, SIMULINK , Wavelet and SVM toolboxes	Complex
8.	Genetic Algorithm approach	SVM, Genetic Algorithm	SIMULINK , SVM toolboxes.	Complex
9.	principal component analysis approach	Principal component analysis, probabilistic neural network (PNN)	PSCAD	Medium
10.	Pattern Recoginition Approach	Voltage Concordia Pattern Feature Aided probabilistic neural network (PNN)	PSCAD	Complex
11.	Deep Learning Machine approach	Convolutional sparse auto-encoder	MATLAB / SIMULINK ,PSCAD	Complex

Complexity viz; simple, medium and complex, is based on factors; simulation time, number of inputs and rules involved in the corresponding computational approach.

- If the simulation time of an approach is less than 0.02s then the approach is very simple with a less accurate results and had an accuracy of around 85.34% to 89.56%.

- If the simulation time of an approach is between 0.02s to 0.09s than the approach is little bit complex, with fairly accurate results within the accuracy range of 90.03% to 93.87%
- If the simulation time of an approach is more than 0.09s than the approach is highly complex with the accuracy margin of 95.65% to 97.15%.

Complexity of any approach can also be defined on the basis of number of inputs required for a particular approach which is as follows

- If the number of the inputs used in a particular approach are less than 50 then it is very simple approach in which the memory requirement is vey less.
- If the number of inputs used in a particular approach are in between 50 and 200 then approach is little bit complex and required more memory locations as compared to simple approach
- If the number of inputs used is more than 200 in any particular approach than the approach is highly complex and requires more number of memory locations for data processing.

1.4 OUTCOME OF LITERATURE REVIEW

- ANN based techniques have been used to classify the faults in the last 20 years. These algorithms depend on identifying the various patterns associated with the impedance information & learn the patterns from previous information during training time but it suffers from computational burden with huge number of training cycles.
- The drawback of ANN is that it represents implicit knowledge but it provides key benefit for fuzzy logic which has explicit knowledge representation using if then rules. Generally, Fuzzy logic approaches are simple than neural networks but one of the drawback of fuzzy logic is that it involves some rules which are linguistic.
- But the wavelet transform approaches provides fast and effective analysis for the detection and classification of faults during transient conditions.It offers advantages over the Fourier transform to analyse the frequency components as well. The multi-resolution in DWT provides the proper window selection to many signal analysis but these are dependent on some threshold values.
- But now-days many modern techniques are used to detect and classify the faults. Undoubtedly, all the techniques provide effective results with high accuracy but these techniques are highly complex.So there is a need of new algorithms which should be highly efficient with less complexity and can be used in real-time analysis.

1.5 RESEARCH GAP

Now-days complexity of modern power gridshas been increased and researchers are still doing effortswith efficiently using advanced machine learning approaches for classification of faults as quick as possible.The conventional neural network has gradient diffusion difficultiesas when neural network is updating the weights of each layer, then weight of these layers are restricted to the errors of output layers. And the gradients decreases very sharply as the number of ANN layers increases and which are back-propagated through output layer to input layer. Therefore the weights updated process is very slow when the gradient descent method is used and not learned effectively by earlier layers. In the published literature about techniques of fault classification, no reference has been found that uses CNNs for analysis of faults in the power system. As CNNs have number of hidden layers which can be trained better than the shallow neural networks and can achieve better accuracy.

1.6 OBJECTIVES OF THE DISSERTATION:

The main objective of the work is to develop a simulation model of convolutional neural network (CNN) based fault classifier in a transmission line of the power system which should classify all shunt types of faults considering different fault locations and fault impedances.

1.7 OUTLINE OF THE DISSERTATION:

- Chapter 1 gives the brief introduction about the whole idea behind the need of classification of faults and the literature review on fault classification techniques by various feature extraction techniques using ANN, Fuzzy Logic, Neuro-fuzzy techniques, Wavelet Transform and new advanced techniques like SVM, machine learning etc**.**
- Chapter 2 presents the brief introduction about power system faults, types of faults, causes of faults, effects of faults and the brief analysis of faults.
- Chapter 3 presents the brief introduction of conventional artificial neural network (ANN), its learning techniques and the algorithms used in ANN.
- Chapter 4 provides the importance of wavelet transform & its advantages over the Fourier transform.The brief introduction about discrete wavelet transforms has been discussed.
- Chapter 5 discuss the convolutional neural network and its advantages over ANN. The architecture of CNN has been presented in this chapter. .

- Chapter 6 discuss the solution methodology for ANN and CNN fault classifier in which first the IEEE-14 bus system is simulated with various fault impedances and various fault locations and then the data is collected with all types of shunt faults. The collected data is sampled to train ANN and CNN classifiers.
- Chapter 7 presents the ANN classifier results & the results of discrete wavelet transform and the CNN classifier for the classification of faults.
- And the last chapter 8 conclude the whole dissertation and the future scope of this work.

CHAPTER 2
POWER SYSTEM FAULTS

2.1 INTRODUCTION

In the balanced condition, the current flowing and the operating voltage are within prescribed limits of a power system. Due to a fault in the system this condition can be disrupted. An interference with the normal current flow because of what fault in a circuit is a failure andfault is defined as any abnormal condition in a power system [23].There should be protective devices to prevent from any damage due to high currents. Also the transients occur during faulty conditions can damage the equipments of a power system and the operating conditions of these equipments may be mal-operated for a particular application [24].The possibility of voltage collapse increasesif the occurrence of fault on transmission lines [25].

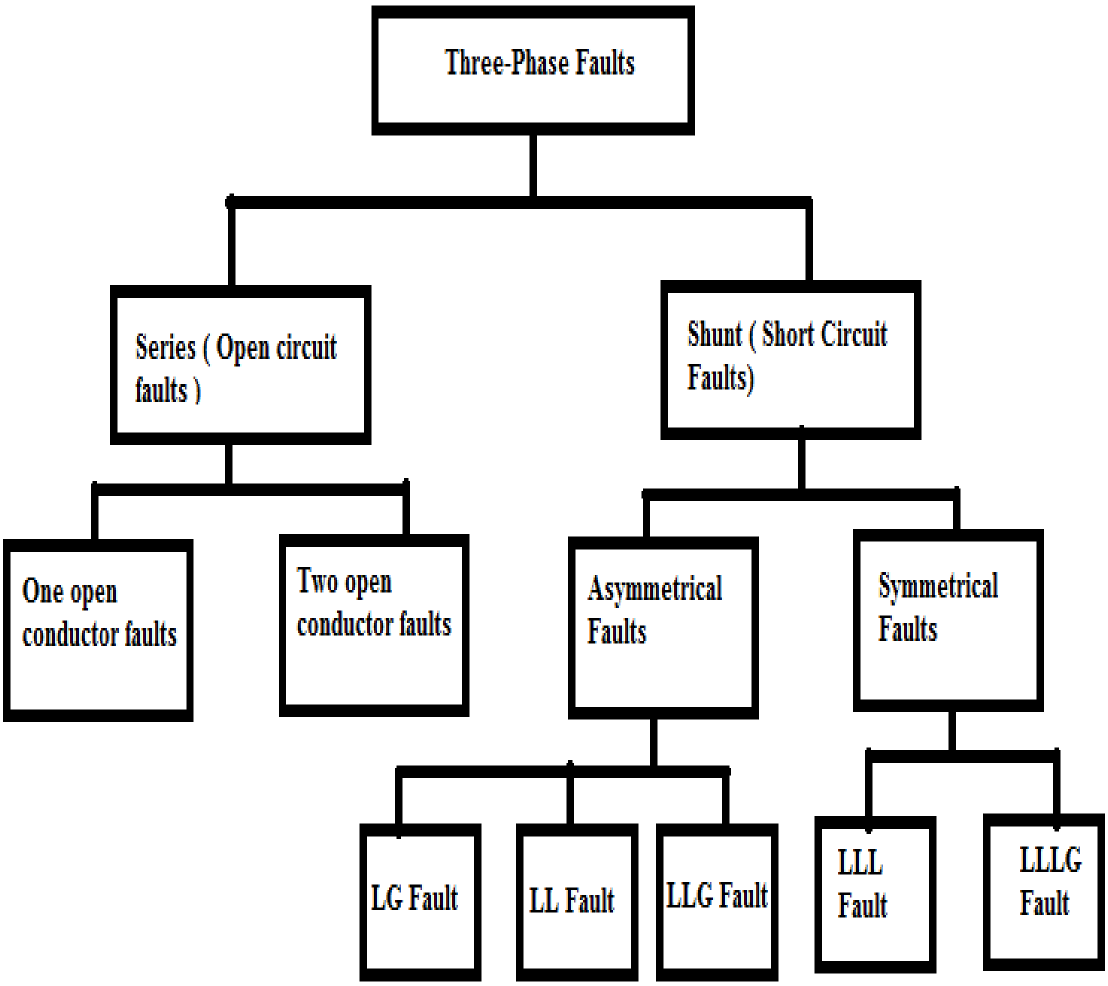


Fig. 2.1 Classification of faults

2.2 TYPES OF FAULTS

The shunt type of faults can be classified as:

- Symmetrical faults
- Unsymmetrical faults

2.2.1 Symmetrical faults: When the disturbance on a power system causes equal fault impedances to the three phases of the system as shown in fig.2.2 this type of fault is called symmetrical fault.This type offault can be L-L-L-G fault or L-L-L fault and also the three phases are equally affected and the system remains balanced so this fault is also called balanced fault and only about 5% of the system faults are three phase faults.

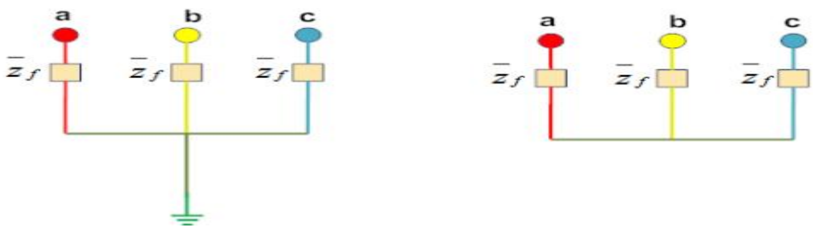


Fig. 2.2 Symmetrical faults representing L-L-L-G and L-L-L faults

Unsymmetrical Faults: When the fault occurrence results in unbalanced system thenit is as called unsymmetrical faults.A singlephasetogroundfault(L-Gfault)is the most commonly occurred unsymmetrical fault. Nearly about 70% of faults in a system are L-G faults. Phasetophasefaults(L-Lfaults)anddoublelinetogroundfaults(L-L-Gfaults)are the other kind of unsymmetrical faults, approximately 10% faults are L-L-G faults and 15% are L-L faults. These faults are shown in Figure 2.3

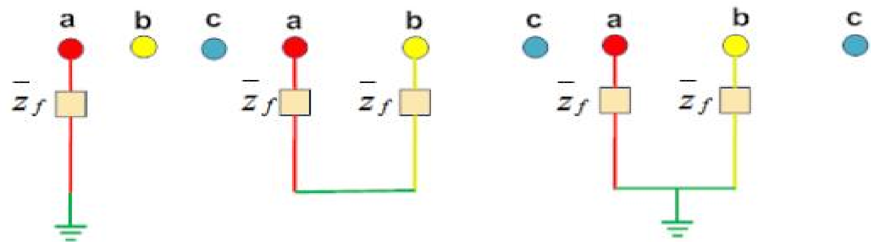


Fig. 2.3 Unsymmetrical faults representing L-G, L-L, L-L-G faults

The external exposure of transmission structures prone to major occurrence of faults. Line insulators to flashoverare caused by lightening strokes.Tower failure may be caused by high velocity winds, ice loading and wind and may lead to mechanically failure of the lines, insulator. It is necessaryto carry out proper fault analysis to select proper rating of circuit breakers with relay settings in proper coordination. Theper phase system is always used for symmetrical faultswhile the unsymmetrical faults using symmetrical components.

Table 2.1 Percentage of occurrence of various short circuit faults

Types of shunt faults	**Representation**	**Percentage frequency of occurrence**
Single line to ground (L-G)		70
Line to line (L-L)		15
Two line to ground (L-L-G)		10
All three phases to ground (L-L-L-G)		2 or 3
All the three lines short circuit (L-L-L)		2 or 3

Table 2.2 Percentage of fault occurrence on different devices of power system

Sr. No.	**Equipment**	**Percentage of total**
1	Overhead line	50
2	Cable	10
3	Transformer	10
4	Switchgear	15
5	Control equipment	3
6	Instrument transformer (CTs and PTs)	2
7	Miscellaneous	10

2.3 EFFECTS OF FAULTS

- During the occurrence of fault, there are different types of effects which can be over flowof current because of low impedance and results in tripping of relays, insulation damage.Also the faults can cause severe shocks to persons who is near during the occurrence of fault at any equipments. The severity of any fault depends on the value of voltage and current.

- Sometimes the heavy currents due to short circuit conditions result in improper working of the component and damaging the insulation windings.
- The faults may sometimes disturbs the near equipments or interconnected circuits to the faulty line which is undesirable for any operating power system.
- Theflashovers produced due to ionization of air during operation of circuit breaker leads to fire which is dangerous to the system as well as for the workers.

2.4 FAULT ANALYSIS

There are number of protection units on a long transmission line to prevent faults. The differential and distance protection schemes are commonly used on the transmission lines of a power system. The distance protection scheme is commonly used as main protection and backup protection for overhead AC lines against all types of shunt faults. The operating time of this protection is mainly depends on the ratio of voltage and current i.e. impedance which further depends on the distance between the relay and the fault. The impedance relay will operate only when the impedance i.e. the ratio of voltage and current falls below a certain value. The CT and PT provides the current and voltage measurement at the fault point during the occurrence of fault. The value of voltage at PT depends on the distance between fault and PT. So distance and impedance are main parameters while analysis of faults.

Let Z_{bus} be the bus impedance matrix of a Generator-Transmission network having N number of buses. Then the bus voltages due to the current injection is given by the equation 2.1

$$V_{i(F)} = Z_{bus} I_{bus(F)} \tag{2.1}$$

Where $I_{bus(F)}$ is the bus current vector having only one non-zero element.Similarly the voltage at i^{th} bus due to fault at p bus is given by

$$V_{i(F)} = V_0 - \frac{Z_{ip}}{Z_{pp}+Z_F} V_0 \ ; \quad i=1,2,\ldots,N \ and \ i \neq p \tag{2.2}$$

Where $Z_{ip}$is the impedance between bus i and p&Z_F is the fault impedance, V_0 is the pre-fault voltage.

For LG fault, the sequence component of fault current at bus p can be expressed as

$$I_p^0\,(F) = I_p^1\,(F) = I_p^2\,(F) = \frac{V_0}{Z_{pp}^1 + Z_{pp}^2 + Z_{pp}^0 + 3Z_f} \qquad 2.3$$

Where Z_{pp}^1, Z_{pp}^2, Z_{pp}^0 are the p^{th} diagonal elements of Z_{bus}^1 , Z_{bus}^2, Z_{bus}^0 matrices

For LL fault , the sequence component of fault current is calculated as:

$$I_p^1\,(F) = \frac{V_0}{Z_{pp}^1 + Z_{pp}^2 + Z_{pp}^0} = -I_p^2\,(F) \qquad 2.4$$

For LLG fault, the sequence components of fault current are calculated as

$$I_p^1(F) = \frac{V_0}{Z_{pp}^1 + \frac{Z_{pp}^2(Z_{pp}^0 + +3Z_f)}{Z_{pp}^2 + Z_{pp}^0 + 3Z_f}} \qquad 2.5$$

$$I_p^2(F) = \frac{V_0 - Z_{pp}^1 I_p^1(F)}{Z_{pp}^2} \qquad 2.6$$

$$I_p^0(F) = \frac{V_0 - Z_{pp}^1 I_p^1(F)}{Z_{pp}^2 + 3Z_f} \qquad 2.7$$

From the above equations, the fault current is dependent on fault impedances. Thus the fault impedance can be considered as main parameterfor the classification of shunt types of faults.

CHAPTER 3
ARTIFICIAL NEURAL NETWORK

3.1INTRODUCTION

An Artificial Neural Network (ANN) is biologically inspired configuration having several layers organized which has set of elementary units of neurons[26]. The basic perceptron model of ANN structure which is also called feed-forward modelis shown in Fig.3.1.From the figure,In each i^{th} layer, there are N_i no. of neurons which has inputs from the previous layer. The signals with excitations are fed as input to the input layer. The neuron performs a non-linear operation and it acts like a processor which produces an output [27]. Every elementary neuron is attached with weights and during the training these weights are adjusted according to the data set.Therefore, ANN produces output by updating the weights based on the given training data set.

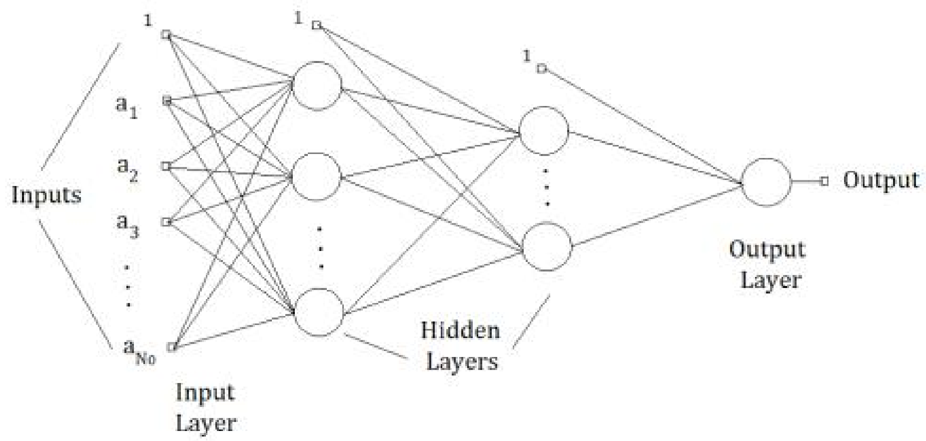


Figure 3.1 A basic three-layer architecture of a feed-forward ANN

The inputs in the form of sets i.e. a_1, a_2 …$a_{n0}$are given to the feedforward ANN as shown in fig.3.1. The training patterns which is considered as most important affects the performance of ANN in terms of functionality of feedforward ANN. The biggest drawback of ANN is that no well-defined expert is there to help in choosing the correct number neurons as well as number of hidden layers. But it has advantage of generalize the data, parallel processing during computing. Therefore it can produce output even if the proper information of input is not applied into the ANN during training time.The only challenge in the application of ANN is to select a proper learning algorithm[28]. But due to their advantages, the ANN has several applications in the field of signal processing and decision-making.

3.2 SUPERVISED LEARNING &UNSUPERVISED LEARNING

Artificial Intelligence models can be categorized as supervised and unsupervised models based on the characteristics offeedback.

3.2.1 SUPERVISED LEARNING ANN MODELS: Supervised learning models learn functions that maps input to output data by using external feedback. Generally, there are two types of Supervised Learning techniques which areclassification and regression. By classification technique, data is separated and by regression technique, data is fitted. The supervised learning includes logistic regression based algorithms, SVM, ANN, and random forestsalgorithms which are the most common.

3.2.2 UNSUPERVISED LEARNING ANN MODELS: Unsupervised models map input data without using any external feedback. The unsupervised learning methods have clustering & dimensionality reduction. The similarities, differences can be foundbyclustering. The similar things are grouped together. The better representation of the data is found by dimensionality reduction.

3.3 ALGORITHMS IN ANN

There are number of optimization algorithms which are used to train the learning procedures in any neural network. All these algorithms have various properties in terms of performance, computational speed and memory requirements. The most important training algorithms for neural networks are Gradient descent, Newton's Method, Conjugate gradient, Quasi Newton and the Levenberg Marquardt (LM) algorithms.

The computational speed v/s memory requirements of different algorithms is depicted by the graph as shown in fig.3.2.

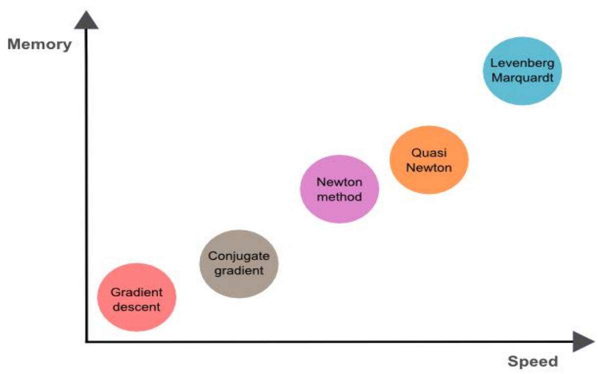


Fig.3.2 Graph between memory and computational speed of different algorithms

As seen from the graph the gradient descent training algorithm is the slowest with one less memory requirement. On the contrary the LM algorithm is the fastest but requires a large memory. To conclude, the gradient descent or conjugate gradient can be used when there is memory saving requirement. The LM algorithm might be best choice when there are few hundreds of parameters to train neural networks.In this work, fault classification mean square error based on LevenbergMarquardt algorithm is calculatedin which the neuron weights are updated to limit the error between the actual output & the desired output[29].

Back Propagation Algorithm

The back propagation which is a learning algorithm based on gradient descent technique limits the error for the non-linear functions of high complexity by selecting right number of hidden layers.It is a supervised learning algorithm. Back Propagation is a systematic approach to train multilayer artificial neural network. The inputs are feed-forward layer wise through the network. The multi-layer model of neural network is shown in figure 3.3.

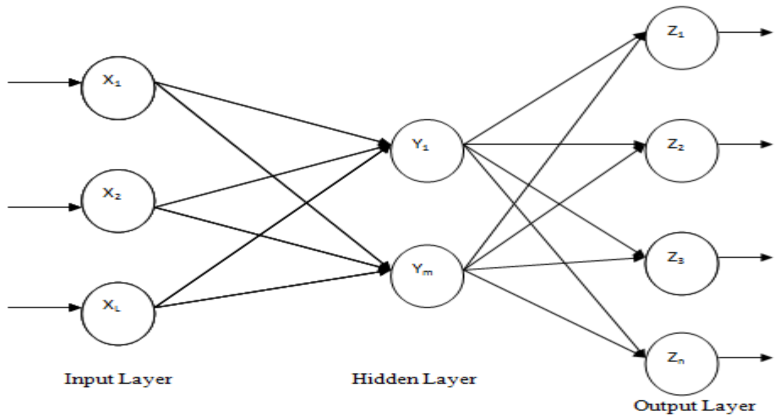


Fig.3.3The multi-layer neuron model.

Where x_1, x_2, x_3…… $x_L$are inputs to the input layer, y_1, y_2, y_3……y_m are the inputs fed to the hidden layer and z_1, z_2, z_3……z_n are the outputs.

Error Back Propagation

The learning process of error back propagation is based on errors calculated at output layers. These errors are then back propagated through hidden layers to input layers and then forward to output layers through hidden layers. It estimates the weights of neurons. After every error calculation, the weights are updated to achieve desired value at the output node. In general, it is a rule formed by the feed-forward neural network to train the parameters of network.The Back Propagation algorithm is briefly described below:

The error signal at the output is expressed in equation 3.1 for l^{th}neuron at n^{th}iteration as:

$$\frac{\partial \xi(n)}{\partial y_l(n)} = \sum_m e_m \frac{\partial e_m(n)}{\partial v_m(n)} \frac{\partial v_m(n)}{\partial y_l(n)} \tag{3.1}$$

The total error energy is expressed in equation 3.2 as:

$$\xi(n) = \frac{1}{2}\sum_{j\in C} e_j^2(n) \tag{3.2}$$

Then average square error energy is expressed in equation 3.3 as

$$\xi_{av} = \frac{1}{N}\sum_{n=1}^{N} \xi(n) \tag{3.3}$$

The input of the activation function with l^{th} neuron is expressed in equation 3.4 as

$$v_l(n) = \sum_{i=o}^{p} w_{li}(n)\, y_i(n) \tag{3.4}$$

The synaptic weight w_{l0} is applied to neuron l and this weight is equal tobias b_l. The functional signal $y_l(n)$ appearing at the output of l^{th} neuron at n^{th}iterationis given by equation 3.5 as

$$y_l(n) = \varphi(v_l(n)) \tag{3.5}$$

By the chain rule ,

$$\frac{\partial \xi(n)}{\partial w_{li}(n)} = \frac{\partial \xi(n)}{\partial e_l(n)} \frac{\partial e_l(n)}{\partial y_l(n)} \frac{\partial y_l(n)}{\partial v_l(n)} \frac{\partial v_l(n)}{\partial w_{li}(n)} \tag{3.6}$$

Differentiating both sides of Eq. (3.2) with respect to $e_j(n)$,we will get the equation 3.7 as

$$\frac{\partial \xi(n)}{\partial e_l(n)} = e_l(n) \tag{3.7}$$

Differentiating both sides of Eq. (3.1) with respect to $y_j(n)$, we get equation 3.8 as

$$\frac{\partial e_l(n)}{\partial y_l(n)} = -1 \tag{3.8}$$

Differentiating Eq. (3.5) with respect to $v_j(n)$, we get equation 3.9 as

$$\frac{\partial y_l(n)}{\partial v_l(n)} = \varphi_l'(v_l(n)) \tag{3.9}$$

Differentiating Eq. (3.4) with respect to $w_{ji}(n)$, we get equation 3.10 as

$$\frac{\partial v_l(n)}{\partial w_{li}(n)} = y_l(n) \tag{3.10}$$

Using Eqs. (3.7) to (3.10) in (3.6), we will get the equation 3.11 as

$$\frac{\partial \xi(n)}{\partial w_{li}(n)} = -e_l(n)\varphi_l'(v_l(n))\, y_i(n) \tag{3.11}$$

By the delta rule

$$\Delta w_{li}(n) = -\eta \frac{\partial \xi(n)}{\partial w_{li}(n)} \tag{3.12}$$

$\eta$is the learning parameter of the BP algorithm. Use Eq. (3.11) in (3.12),

$$\Delta w_{li}(n) = \eta \delta_l(n) y_i(n) \tag{3.13}$$

The local gradient $\delta_j(n)$ is expressed by equation 3.14 as

$$\delta_l(n) = \frac{\partial \xi(n)}{\partial v_l(n)} = \frac{\partial \xi(n)}{\partial e_l(n)} \frac{\partial e_l(n)}{\partial y_l(n)} \frac{\partial y_l(n)}{\partial v_l(n)} \tag{3.14}$$

$$= e_l(n) \varphi_l'(v_l(n))$$

Case 1 When neuron l is taken as an output node

When neuron *l* is located in the output layer of the network, it is supplied with the desired response. The error is computed using Eq. (3.1) as shown in Fig. 3.4. The local gradientis calculated using Eq. (3.14) and it is the best approach for calculating gradient.

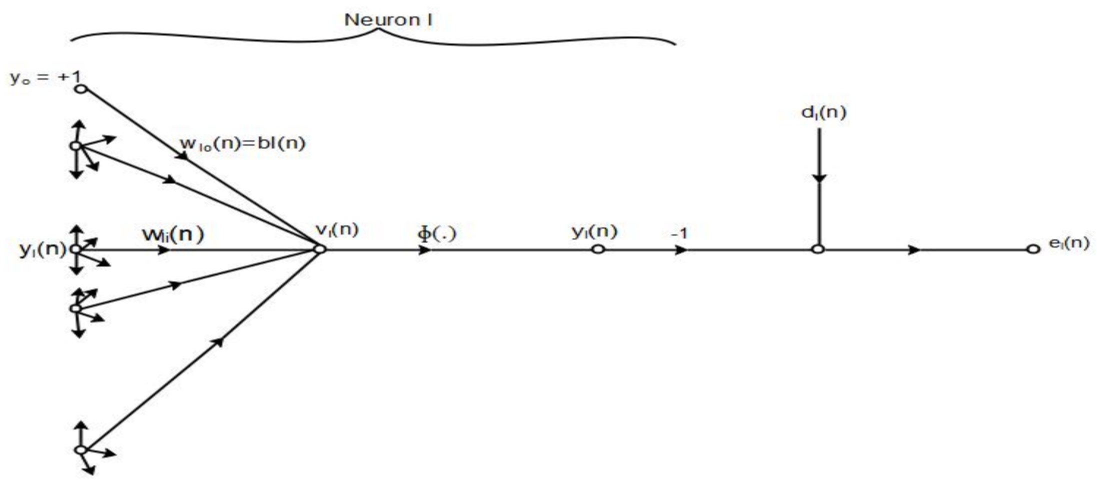


Fig.3.4 Output neuron j signal flow graph

Case 2 When neuron l is taken as a hidden node

When neuron *l* is located in a hidden layer of the network, there is no specified desired response of that neuron. Accordingly, the error signal for a hidden neuron would have to be determined in terms of the error signal of all the neurons as shown in Figure 3.5. According to Eq. (3.14) redefines the local gradient for hidden neuron *l* as

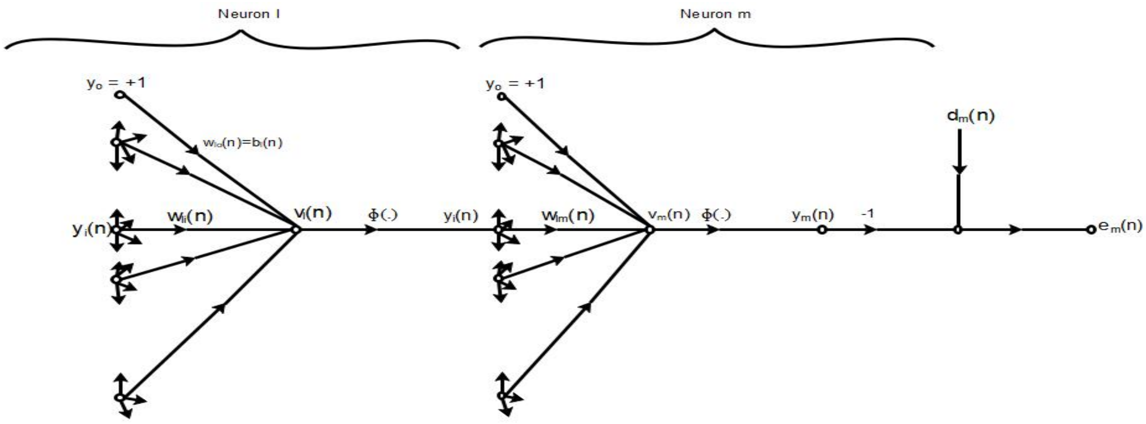


Fig. 3.5 Output neuron k connected to hidden neuron j signal flow graph.

$$\delta_l(n) = -\frac{\partial \xi(n)}{\partial y_l(n)} \frac{\partial y_l(n)}{\partial v_l(n)} \tag{3.15}$$

$$\delta_l(n) = -\frac{\partial \xi(n)}{\partial y_l(n)} \frac{\partial y_l(n)}{\partial v_l(n)}, \text{ neuron } l \text{is hidden}$$

$$\xi(n) = \frac{1}{2} \sum_{l \in C} e_m^2(n), \text{ neuron } m \text{ is an output node} \tag{3.16}$$

Differentiating Eq. (3.16) with respect to functional signal,

$$\frac{\partial \xi(n)}{\partial y_l(n)} = \sum_m e_m \frac{\partial e_m(n)}{\partial y_l(n)} \tag{3.17}$$

Using the chain rule rewriting the Eq. (3.17),

$$\frac{\partial \xi(n)}{\partial y_l(n)} = \sum_m e_m \frac{\partial e_m(n)}{\partial v_m(n)} \frac{\partial v_m(n)}{\partial y_l(n)} \tag{3.18}$$

From Fig. 3.4,

$$e_m(n) = d_m(n) - y_m(n) \tag{3.19}$$

$$= d_m(n) - \varphi_m\left(v_m(n)\right), \quad \text{neuron } m \text{ is an output node}$$

$$\frac{\partial e_m(n)}{\partial v_m(n)} = -\varphi'_m\left(v_m(n)\right) \tag{3.20}$$

The induced local field for neuron *m*,

$$v_m(n) = \sum_{l=0}^{p} w_{ml}(n) y_l(n) \tag{3.21}$$

Differentiating Eq. (3.21) with respect to $y_l(n)$,

$$\frac{\partial v_m(n)}{\partial y_l(n)} = w_{ml}(n) \tag{3.22}$$

Using Eq. (3.20) and (3.21) in (3.18),

$$\frac{\partial \xi(n)}{\partial y_l(n)} = -\sum_m e_m(n)\varphi_m'\left(v_m(n)\right)w_{ml}(n) \tag{3.23}$$

$$= -\sum_m \delta_m(n)\, w_{ml}(n)$$

Getting back propagation formula using Eq. (3.23) in (3.15),

$$\delta_l(n) = \varphi_l'\left(v_l(n)\right)\sum_m \delta_m(n)\, w_{ml}(n), \text{ neuron j is hidden} \tag{3.24}$$

If neuron l is an output node, $\delta_l(n)$ equals the product of the derivative $\varphi_l'\left(v_l(n)\right)$ and the error signal $e_l(n)$, with the neuron j as in Eq. (3.14).

If the neuronis a hidden node $\delta_l(n)$ equals the product of the associated derivative $\varphi_l'\left(v_l(n)\right)$ and the weighted sum computed for the neuron in the output layer that is connected to neuron l, as in Eq. (3.24).The flowchart for the training of ANN for a two-layer neural network by BP is shown in fig.3.6.

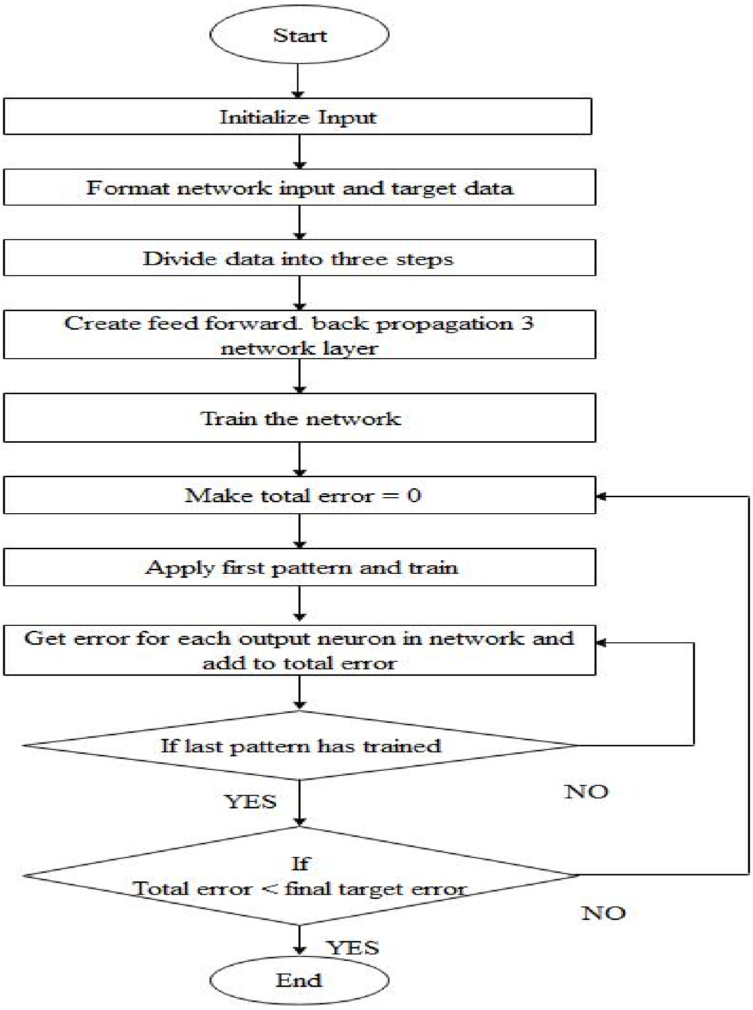


Fig. 3.6 Flowchart of Backpropagation Algorithm

3.4 MERITS AND DEMERITS OF ANN IN FAULT CLASSIFICATION TECHNIQUES

- One of the main advantage of ANN is that it can produce output even with incomplete input data. Another advantage is that it has ability to tolerate faults when one or more units of ANN does not work and still it can produce outputs. Also it has parallel processing capability.
- But it has gradient diffusion problems as when neural network is updating the weights of each layer, then weight of these layers are restricted to the errors of output layers. And the gradients decreases very sharply as the number of ANN layers increases and which are back-propagated through output layer to input layer. Therefore the weights updated process is very slow when the gradient descent method is used and not learned effectively by earlier layers.
- Also,there are number of unlabeleddata in which ANN finds difficulty in processing and training during the operational period of the power system.

But to solve the problems given above, the effective approach now-a-days is deep learning which has training procedures which iterates and updates the weights in each layer of network effectively. It vanishes the gradient problems of traditional neural networks.

CHAPTER 4
WAVELET TRANSFORM

4.1 INTRODUCTION

'Wavelet', a concept introduced by Jean Morlet a French geophysicist in 1982 means a small oscillating wave of effectively finite duration that has an average value of zero and has derived the concept of wavelet transform as a new tool to analyse the seismic signal. Wavelet analysis is an exciting method that can detect and analyse abrupt changes in signals. The wavelet theory is applicable to solve complex tasks in mathematics, physics, and engineering, with modern applications as diverse as wave propagation, data compression, signal processing, image processing, pattern recognition, computer graphics etc. [30], Wavelet can be seen as a complement and has many advantages over classical Fourier decomposition method.

By definition, a wavelet is an oscillation that decays quickly as shown in figure 4.1.

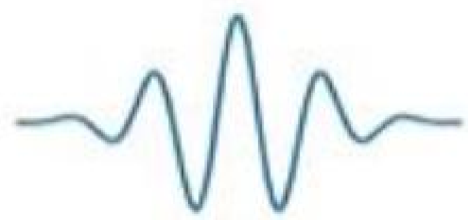


Fig. 4.1 The wavelet for large scale factor with low frequency

Mathematically, the conditions to be wavelet are expressed as in equation 4.1and 4.2

$$\int_{-\propto}^{+\propto} |\varphi(t)|^2 dt < \propto \quad (4.1)$$

$$\text{And} \int_{-\propto}^{+\propto} |\varphi(t)| dt = 0 \quad (4.2)$$

Originally wavelet is considered as family of functions constructed through translations and dilations of a single function called 'mother wavelet' and is expressed by equation 4.3

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \varphi\left(\frac{t-b}{a}\right), such\ that\ a, b \in R\ \ and\ a \neq 0 \quad (4.3)$$

Parameter a is called scaling parameter that represents the degree of compression.

Parameter $b$is called translation parameter that determines the time position of wavelet.

In case $|a| < 1,$ then wavelet is compressed one of short time domain and mainly corresponds to higher frequencies as shown in figure 4.1.

On the other hand, in case $|a| > 1$, then wavelet is the stretched version of mother wavelet. It has large time domain and corresponds to low feequencies as shown in figure 4.2.

Compressed wavelet helps in capturing the abrupt changes in signals and stretched wavelet helps in capturing the slow varying signals. Higher frequencies can be better analyzed intime, and lower frequencies can be better analyzed in frequency. The components which has high frequency can be located better in than a component which has low frequency and vice versa [31].

Fig. 4.2 The wavelet for small scale factor with high frequency

This time-frequency representation provided by Wavelet transform stands it as one of the powerful and important tool of signal representation and time-frequency signals. Now-days, it is being used in image processing, data compression, and signal processing [32]. There are two types of signals

- Continuous Signals
- Discrete Signals

Continuous signals are described as continuous function of time. Speech signal or sound signal is one of the best example of continuous signal. Wavelet is used to filter noise in the sound or speech signal.

Discrete signals are represented by numbers or pair of numbers. Black and white photo si one of the best example of discrete signal. Wavelet supports the identification of grey areas (partially black or white) as black or white areas using decomposition method.

For representation of a signal , wavelet is represented by a function having a finite sequence of wavelet coefficients. In simple form, wavelt function is represented by equation 4.4

$$f(x) = \sum_{a \in R} \sum_{b \in R} d_{a,b} \varphi_{a,b}(x) \tag{4.4}$$

Where $d_{a,b}$ are wavelet coefficients.

4.2 ADVANTAGES OF WAVLET TRANSFORM OVER FOURIER TRANSFORM

Wavelet transform is preferred over Fourier transform on account of following reasons:

1. In FT, the signals are represented in sinusoidal form by single peak. It provides information of signals in terms of frequency only, not in the time domain. But the basis of WT are located in time-domain. In general, it represents small waves. The analysis of WT is done by scaling and wavelet function. Basically scaling and translation of signals are done by these functions. Thus, the Wavelet transform is represented in both time and

frequency simultaneously. Also, it provides a multi-resolution system such as image processing.

2. Fourier transform is based on single parameter function i.e. scaling only. Where as wavelet transform is a two parameter function both scaling and translation is possible
3. For discontinuos signal, the significant coefficients of large magnitude is produced by FT, but only few such coefficients are generated by WT.
4. The method of non-linear approximation is used to check the approximation power of transform. In this method, only few coefficients are taken and rest of the coefficients set to zero. And the signal is reconstructed using these coefficients. The fast and better results are obtained by wavelet transform of non-linear approximation method as compared to FT. The most signals are natural which has few discontinuities. For instance speech, natural images etc. Also WT has applications of compression and de-noising of signals with good efficiency.
5. The transform of windowed signals is not taken by FT i.e. single peak of sinusoidal signals is taken and no negative frequencies are computed. But the window width can be changed by WT.
6. One of the biggest advantage of WT is to separate fine information of the signals. The very fine information of signals can be analysed by small wavelets the course details by large wavelets.
7. They compress and scale the signals without so much degradation.

4.3 TYPES OF WAVELET TRANSFORMS

There are basically two types of wavelet transform i.e. time-scale and time-frequency wavelet transform. Out of these two transforms, time-scale is commonly used. The signals can be decomposed into a wide range of scales by this transform and this is called multi-resolution analysis. Some wavelet families are Haar, Coiflets,Biorthogonal,Morlet,Daubechies, Mexican Hat,Meyer,Symlets, real and Complex Wavelets. The discrete wavelet transform gives sufficient information about the original signal for analysis & synthesis, with less computation time. The Daubechies family of wavelet are orthogonal discrete wavelet transform which is characterized by number of vanishing moments with scaling function which results in multiresolution analysis. Some signals has self-similarity properties, fractal problems and some signals has discontinuities, to solve such problems this family of wavelets is broadly used.The db1-db10 are commonly used daubechies wavelets. The index number denotes the number of vanishing moments. The number of vanishing moments are equal to half the number of coefficients.The number of vanishing moments represents approximation order and also smoothness of the wavelet. For instance, db2 has two and db4 has four vanishing moments. The db2 wavelet can easily encode the polynomial with one coefficient and db4 can encode the polynomial with two coefficients and so on.

There are different kinds of ‘Daubechies’ families of wavelets as shown in fig.4.3.

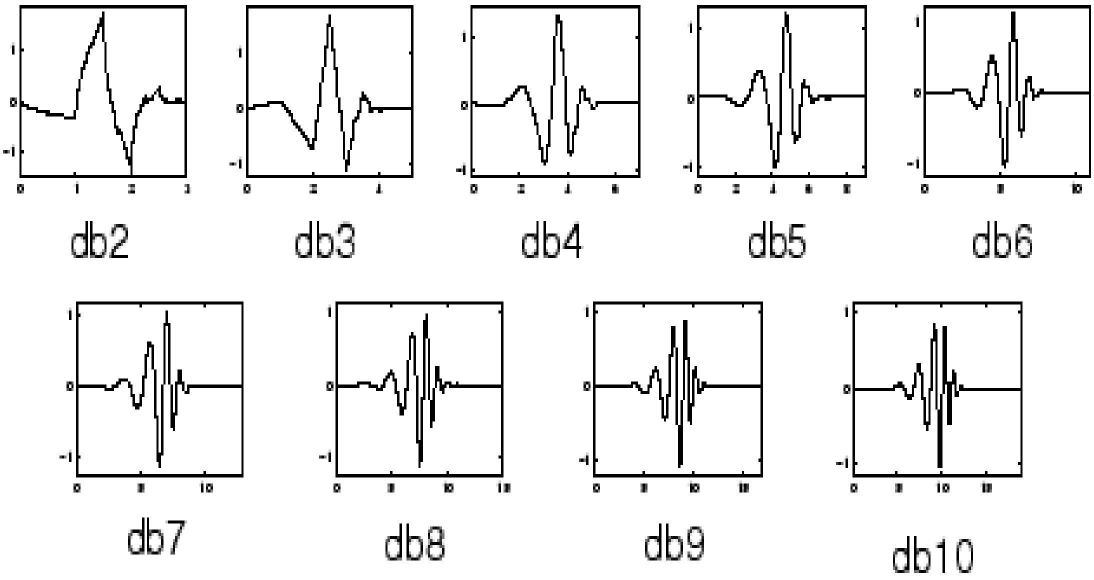


Fig.4.3 Differentdaubechies families of wavelets

4.4 DISCRETE WAVELET TRANSFORMS

The signals analyzed by DWT are decomposed at various bands of frequency with changing resolutions. The decomposed signal has information in terms of approximation and detailed coefficients [33]. The decomposition of the signal is done with high-pass filters and low-pass filters which provides filtering of the signal which is in time domain at different frequency bands. The different filtering procedures with different cut-off frequencies are done to analyze the signals of different scales.

The procedure of discrete wavelet transformation is as follows:

Step 1: Present the two parameter wavelet function signal in arrangedby digital half band which has low-pass filter with impulse response $h[n]$.

Step 2: Filtered signal is corresponded to mathematical operation that has impulse response of the filter by convoluting the signal. It defines the convolution operation in discrete time as in equation 4.5.

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] * h[n-k] \tag{4.5}$$

Step 3: When signal passes through a half band low-pass filter, it gets vanished half of samples by following Nyquist's rule. Now highest signal frequency can be seen in Л/2 radians despite of Л radians.

Step 4: Double the scale of the signal.

High frequency data is removed by filtering but, it remains same. Scale is altered by process of subsampling on the other side, resolution is affiliated to the information of amount which is in signal. Thus, filter operates affects to it. Half frequencies are removed by half band low-pass filtering that may interrupt the half of the information. Therefore, after filtering operations, resolution is halved. It does not influence the resolution to scale sampling operation removed half component from signal makes half the number of samples. Without any loss of data, half samples have ability to discard. All in all, the low-pass filtering halves the resolution, but scale remains unchanged. Then signal is sub-sampled by 2.

This procedure can mathematically be expressed as in equation 4.6

$$y[n] = \sum_{k=-\infty}^{\infty} h[k] * x[2n-k] \tag{4.6}$$

Step 5: The original signal x[n] is first passed through a half-band high-pass filter g[n] and a low-pass filter h[n].

Step 6: After the filtering, half of the samples can be eliminated according to the Nyquist's rule, since the signal now has a highest frequency of p/2 radians instead of p. The signal can therefore be sub-sampled by 2, simply by discarding every other sample. This constitutes one level of decomposition and can mathematically be expressed as follows in equation in 4.7 and 4.8 [36].

$$y_{high}[k] = \sum_n x[n] * g[2k-n] \tag{4.7}$$

$$y_{low}[k] = \sum_n x[n] * h[2k-n] \tag{4.8}$$

Where $y_{high}[k]$ and $y_{low}[k]$ are the high-pass filter output and low-pass filter output, after sub-sampling by 2. This decomposition halves the time resolution since only half the number of samples now characterizes the entire signal. However, this operation doubles the frequency resolution,since 61 the frequency band of the signal now spans only half the previous frequency band, effectively reducing the uncertainty in the frequency by half. The above procedure, which is also known as the sub-band coding, can be repeated for further decomposition. At every level, the filtering and sub-sampling will result in half the number of samples (and hence half the time resolution) and half the frequency band spanned and hence double the frequency resolution.

Figure 4.4 illustrates this procedure, where $x[n]$ is the original signal to be decomposed, and $h[n]$ and $g[n]$ are lowpass and highpass filters, respectively. The bandwidth of the signal at every level is marked on the fig.4.4 as "f"

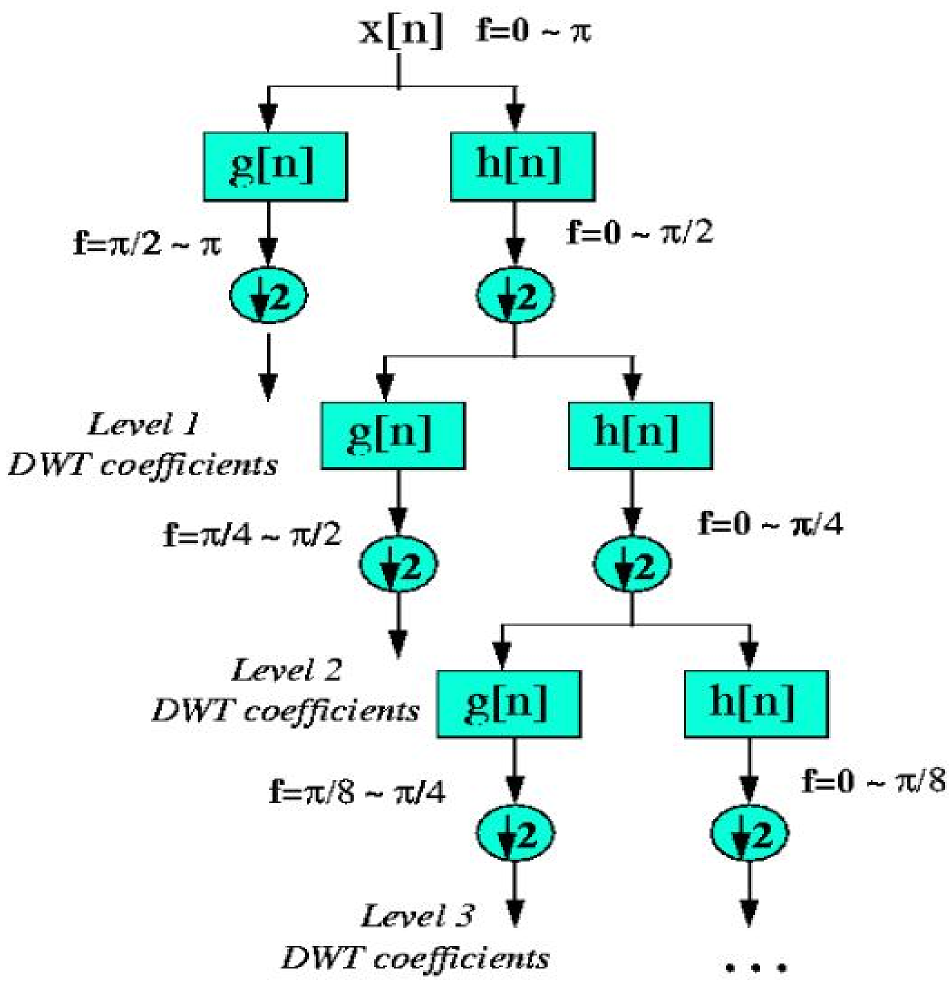


Figure 4.4 Different level decompositions of DWT

4.5 CONCLUSIONS

The time-frequency representation is provided by Wavelet transform. The time-scale wavelet is commonly used to decompose the wide range of signals with scaling function. As compared to FT, WT has many advantages of multiresolution analysis of the signals, of being able to separate the fine details in a signal, better and fast results for nonlinear approximation and so on. The db1-db10 are commonly used daubechies wavelets of DWT. The effectiveness of the wavelet analysis is largely influenced by the choice of the mother wavelet. The choice of the appropriate mother wavelet depends on the nature of the signal and on the type of information to be extracted from the signal.

CHAPTER 5

CONVOLUTIONAL NEURAL NETWORK

5.1 INTRODUCTION

The neural networks have good characteristics to learn complex input-output relationships which are nonlinear, have sequential procedures of training, and adapt themselves to the data. The radial-basis function networks and feed-forward networks of neural network are the most commonly used neural networks for pattern classification and for different visual recognition tasks, image classification which requires extraction of features[34]. However, an expert needs to identify most of the applied features and is coded manually according to the data sets. This is difficult task to write a robust algorithm for detection of handwritten alphabets/digits or objects in images etc., and also requires more computational time.

Deep learning is another class of artificial neural network that too mimics human brain. But the structure has multiple hidden layers in between input and output layers and is capable of modelling and processing nonlinear relationships. The term '*deep*' refers to the number of sequential layers in the network for data processing, understanding of human speech and to recognise the objects visually.

The information passes in deep learning is trained layer wise and the outputs calculated at the previous layer actlike inputs to the next layer. The input layer refers to first layer, many number of hidden layer refers to the middle layers and output which is last layer. All the layers are simple, having activation function with uniform algorithms.

The deep learning has significance in extraction of features by using algorithms which automatically learns the data, training [35].

The earliest efforts in developing deep learning algorithms date to 1965, when Alexey Grigoryevich Ivakhnenko and Valentin Grigorʹevich Lapa used models with polynomial activation functions, which were subsequently analysed statistically.

There are number of architectures of deep learning available but the most effective architecture to classify images isConvolutional Networks (ConvNets). The CNN models are started in 2012 with AlexNet and after that these models have grown rapidly. From the science of biology their multistage architectures are inspired. One of the biggest advantage of CNN is that the invariant features are learned hierarchically and automatically through these models without any human supervision. Firstly, the low level features are learned then these features are recognized and combined to learn the new complex patterns [36].

5.2 ADVANTAGES OF CNN OVER ANN

- As there is large number of unlabelled data, in which ANN finds difficulty in processing and training. Some times this difficulty can be seen during operational period of power system. But deep learning overcomes this difficulty by using the both supervised and unsupervised learning effectively. It uses the unsupervised learning during training of large amount of data to train the dense hidden layers. It is very effective approach to converge the parameters into the local optimum during training of such deep neural networks.
- These neural networks have very strong ability to extract features from the samples and visualization which results in effective classification of features.
- The training process of CNN is layer-wise and the pre-training procedures, the weight parameters of each layer is updated by using the gradient descent approach.

However the deep learning approach is limited by learning through observations only and selection of biases is a great issue. For successful process of deep learning, it suffers from the requirements of large amount of data and hence more computationally expensive. Alss there does not exist any strong theoretical foundation.

5.3ARCHITECTURE OF CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are generally trained as supervised models in which both input images and the target labels are available within the training data.Basically CNN is a mathematical construct trained to learn feature extract. In contrast to three layers of conventional neural network, CNN is composed of variety of layers : input, convolution, MAX pooling, hidden and output layer. The feature extraction part is done by convolution layer and MAX pooling layer.The extracted features obtained from convolution and pooling layer are mapped into final result as output which is used for classification . Height, width, depth are three dimensions to be represented by each layer.Height, width, depth are three dimensions to be represented by each layer. The architecture of CNN is shown in fig.5.1.Images are made up of pixels in the form of digits. The architecture of CNN can easily be understood by knowing the steps involved in building a convolutional neural network.So there are generally five steps involved which are described below:

STEP 1: CONVOLUTION

The word ‘convolution’ in CNN comes from the convolution theorem which is integration of two combined functions and modify the one function with the other function as given by equation 5.1.

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\beta)\, g(t-\beta)\, d\beta \qquad (5.1)$$

But in that context, the convolution formula can be described as a weighted average of the function $f(\beta)$ at the moment t where the weighting is given by $g(-\beta)$ simply shifted by amount t. As t changes, the weighting function emphasizes different parts of the input function.

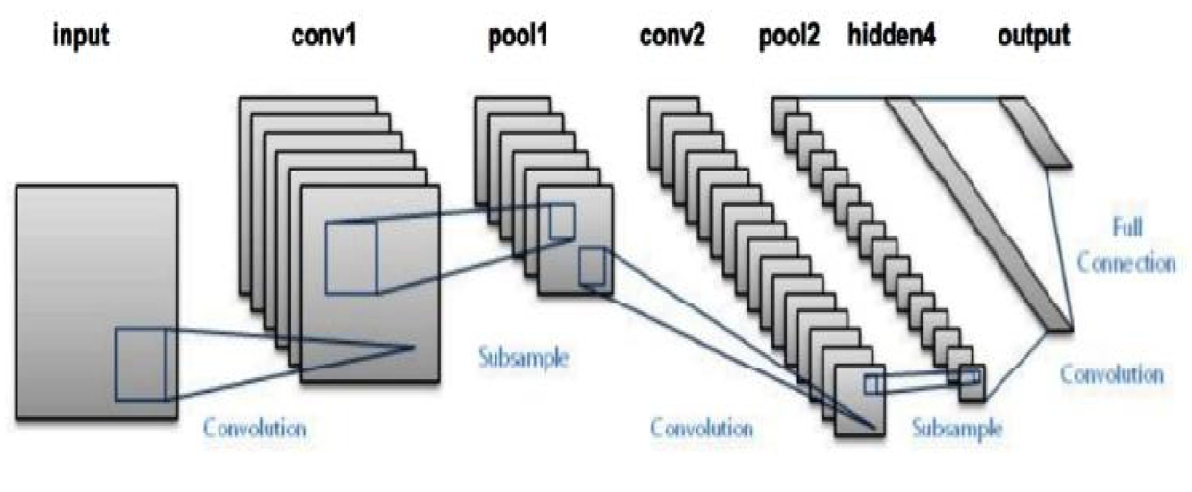


Fig.5.1The Convolutional Neural Networks (CNN) architecture [37]

In Discrete form: For complex-valued functions f, g defined on the set Z of integers, the discrete convolution of f and g is given by equation (5.2)

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m] g[n-m] \qquad (5.2)$$

$$\text{or} \qquad = \sum_{m=-\infty}^{\infty} f[n-m] g[m]$$

In general, to learn feature representation is the main motive of convolution layer. The three terms input image, feature kernals and feature map are involved in convolution process. The image which is being detected is the input image, then next is the feature kernel also called filter is a matrix having some values defined and this kernel is multiplied with input image as element wise to produce another feature map which is called convolved feature map. In general this step is done to reduce the size of the image which results in processing fast and easy, also some features are eliminated and only the unique features are preserved.Equation 5.3 used for the convolution process can be written as

$$S(i,j) = I \times K(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \qquad (5.3)$$

where$I$ is an input image, K is kernel/filter used in convolution process, m is row of image, and n is column of image.

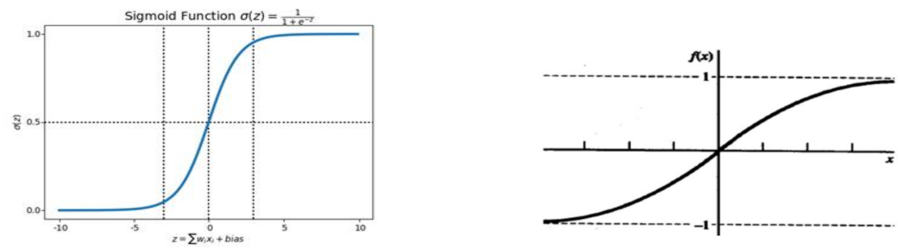
STEP 2: APPLYING ACTIVATION FUNCTION

The commonly used activation functions are sigmoid, Hyperbolic tangent function (tanh) and Rectified Linear unit (ReLU) [38]. These functions are applied to create non-linearties so that neural network becomes more powerful and should ability to learn complex data.

Sigmoid Activation function: It is a activation function as shown in fig.5.2 and expressed in equations (5.4) & (5.5) are

Unipolar Sigmoid: $$f(x) = \frac{1}{1+e^{-x}} \qquad (5.4)$$

Bipolar Sigmoid: $$f(x) = \frac{1-e^{-x}}{1+e^{-x}} \qquad (5.5)$$



(a) Unipolar sigmoid function (b) Bipolar sigmoid function

Fig. 5.2 Sigmoid function [38]

The range of this function is between 0& 1. It can be easily understood and applied but it has major disadvantages i.e. vanishing gradient problem, have slow convergence and it makes optimization difficult because of gradients updation in different directions.

Hyperbolic Tangent function-Tanh: It is a activation function as shown in figure 5.3 and of the form of equation 5.6.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (5.6)$$

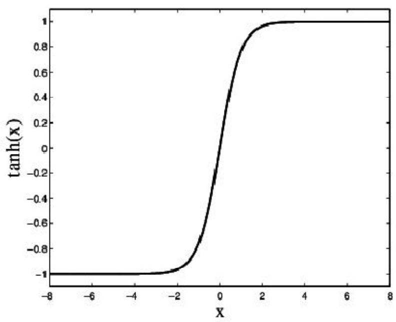


Fig. 5.3Tanh activation function [38]

This function is preferred over sigmoid and has output at zero centred as its range is in between -1 and 1. Due to this range, optimization is easy but still has difficulty in vanishing gradients.

Rectified Linear unit (ReLU): TheReLU function in mathematical form as expressed in equation 5.7 and fig.5.4 represents the ReLU function.

$$R(x) = max(0,x): \text{ if } x< 0 \text{ , } R(x) = 0 \text{ and if } x>= 0 \text{ , } R(x) = x \tag{5.7}$$

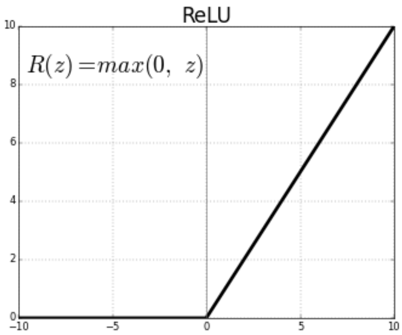


Fig.5.4ReLU activation function [38]

As seen from the above expression that it is very simple and efficient activation function.it is used to create non-linearity in the CNN. The main advantage of ReLU activation function is that it gives large and consistent gradients (does not saturate) when active and also efficient to optimize, converges much faster than sigmoid activation function or tanh activation function. But the only drawback of this function is that it can be applied only within the hidden units of the network.

STEP 3: POOLING

The pooling is very similar to convolutional process i.e. convolving kernals on input sets. It is mainly done to detect the lighting in pictures, angles of the pictures. In general, there are three types of spatial pooling namely; MAX pooling, Average poling and Sum Pooling.

The MAX pooling is basically done to reduce the size of picture by preserving the relavant features of the image. It results in reducing overfitting data when CNN is fed with so much input information. MAX pooling uses the maximum value from each cluster of neurons at the prior layer. MAX pooling generates a new, output matrix where each element is the max of a region in the original input.

STEP 4: FLATTENING

After the pooling, the resultant feature map is called pooled feature map which is flattened now in this step. This is done to transform the entire feature matrix into a single column matrix which is further fed for processing to the neural network.

STEP 5: FULLY CONNECTED LAYER

In the last step, the fully connected neural network is fed by flattened feature map. The fully connected layer is very similar to the simple ANN architecture which has hidden layers present inbetween the input and output layers. The information is fed through this network and then error is calculated and then error is back-propagated to adjust the weights. The total error at the output layeris expressed as in equation 5.8.

$$E = \frac{1}{2N}\sum_{i=1}^{N}(\mathrm{y}-\mathrm{x})^2 \tag{5.8}$$

Where x is the desired value and y is the actualoutput.

The BP algorithm based gradient descent method is used to calculate gradients of all errors for adjusting the weight parameters and limits the output errors [39]. But these errors are figured out in terms of probability for the different classifications using soft-max function [40]. The soft-max function specifies a discrete probability distribution for K classes, denoted by equation 5.9 as

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} \text{for} j = 1, \ldots, k \tag{5.9}$$

Where z is a vector of the inputs to the output layer and j indexes the output units.

To summarise, flow chart of CNN processing is as shown in figure 5.5

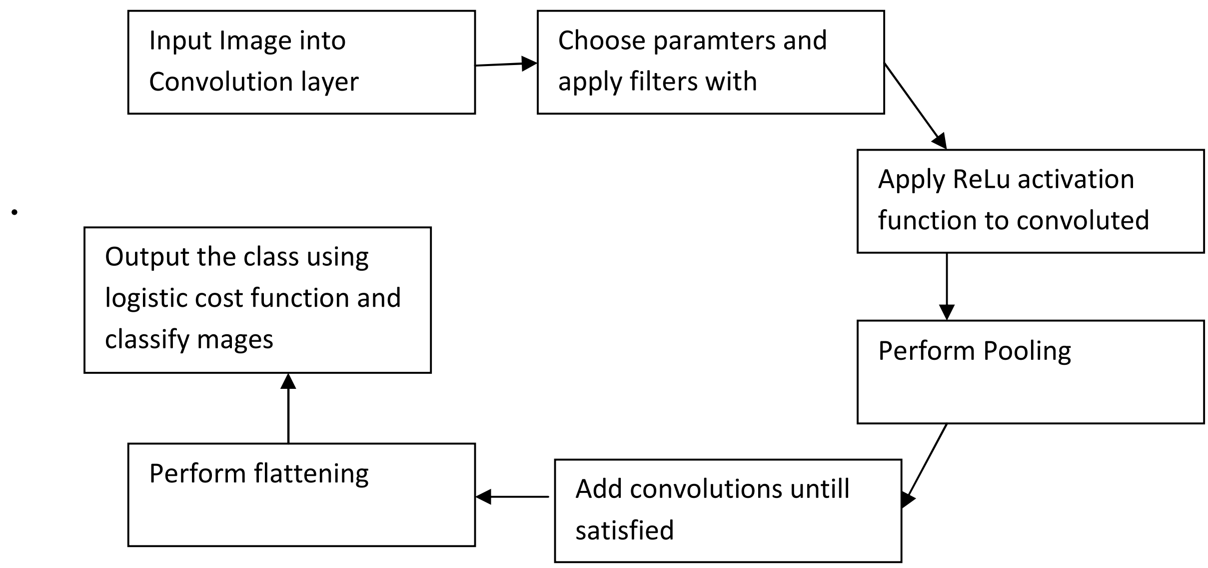


Fig. 5.5: Flow chart of Deep Learning

CHAPTER 6

SOLUTION METHODOLOGY

6.1 SIMULATION

The proposed ANN and CNN framework has been implemented for shunt type fault classification on 200km long transmission line. Figure 6.1 depicts the simulation model of the IEEE 14 bus system. Fault patterns are generated for 11 types of shunt faults considering two casesFault locations and the other is fault impedances.Shunt type faults those are considered are listed in table 6.1

Table 6.1 All 11 types of faults with labels

Label of fault	**Name of the fault**
A-G	Line A to Ground fault
B-G	Line B to ground fault
C-G	Line C to ground fault
A-B	Line A to Line B fault
B-C	Line B to line C fault
A-C	line A to line C fault
A-B-G	Line A to line B to ground fault
B-C-G	Line B to Line C to ground fault
A-C-G	Line A to Line C to ground fault
A-B-C	Three line short circuit fault of three phases A,B and C
A-B-C-G	Three line short circuit fault of three phases A,B and C with ground fault

Generally, The problem of power system fault classification issolved in three stages.

Stage 1: Data is collected for different shunt type of faults on 200km long line IEEE14 bus system using Simulink simulation. The voltages and currents of three phases are considered as fault patterns those are areobtained by considering two cases

- Fault impedance
- Fault location

Stage 2: Data is sampled to extract features from three phase voltages and current patterns.

Stage 3: Extracted features in the form of discrete images are presented to classifier. In this study convolutional neural network is used for classification of fault patterns.

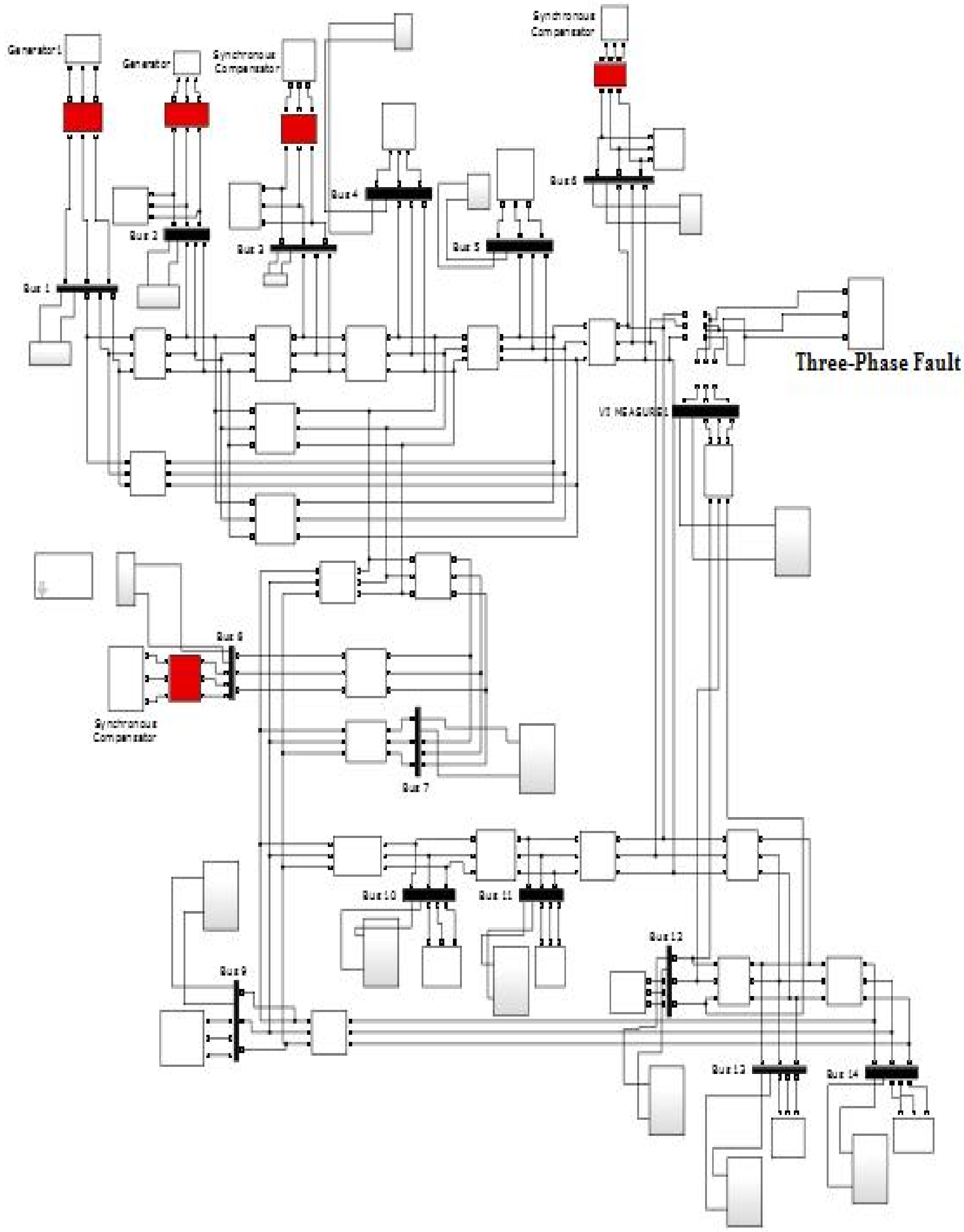


Fig. 6.1 Simulink model of IEEE-14 Bus system

Table 6.2 Different Parameters with variation values

Sr.No.	Parameter	Variation Values
1	Fault Location (Line length=200km) (Resistance = 0.001 ohm)	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 (km).
2	Fault resistances (line length=200km)	0.001, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95(ohms).

The table 4 shows that or fault location, the model was simulated 19 times for each fault at different fault locations; therefore total simulations for all 11 types of fault are 19*11=209. Similarly, for fault resistances, the model was simulated 20 times for each fault at different values of resistances; therefore total simulations for all 11 types of faults are 20*11=220

6.2 DATA SAMPLING

Case 1: FEATURE EXTRACTION IN ANN

During the faults, there are sudden sharp changes which are noticed in the faulty lines of the system as system changes from the healthy state to a abnormal state as shown in fig.6.2 and fig.6.3.

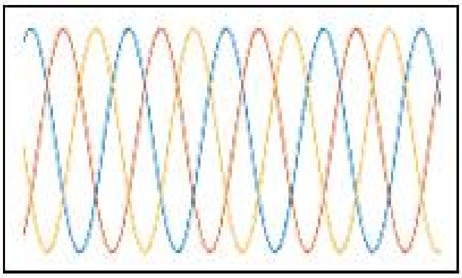


Fig. 6.2 Healthy Waveform of current

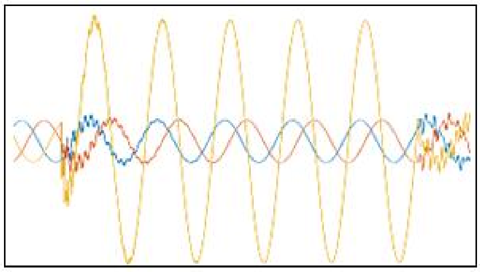


Fig.6.3 Faulty Waveform of current

The fault is assumed to occur at the time instant 0.2 seconds. In total 80,000 data points are sampled for normal and faulty conditions. For 20 cycles of input under observation, 20 points are selected randomly for three-phase voltage and current. The current and voltage values are normalized for the faulty section concerning healthy section.

The table 6.3 and table 6.4 depict the values of voltage and current of each phase after normalization for every cycle of ABG fault at 0.001 ohm resistance. The 10 random values of voltage and current of each phase are taken after every cycle.

Table 6.3 Faulty Voltage values after normalization for every cycle for ABG fault at 0.001-ohm resistance

Sr. No.	V_a	V_b	V_c
1.	0.1487	0.7474	-0.8806
2.	0.1332	0.7482	-0.8952
3.	0.1578	0.7305	-0.8930
4.	0.1327	0.7545	-0.8944
5.	0.1389	0.7441	-0.8911
6.	0.1362	0.7425	-0.8774
7.	0.1489	0.7292	-0.8847
8.	0.1611	0.7402	-0.9019
9.	0.1512	0.7459	-0.8970
10.	0.1441	0.7406	-0.8847

Table 6.4 Faulty Current ($1x10^{-10}$) values after normalization for every cycle for ABG fault at 0.001-ohm resistance.

Sr. No.	I_a	I_b	I_c
1.	-4.4	3.8	-0.11
2.	-3.8	3.4	-0.13
3.	-3.3	2.9	-0.10
4.	-3.0	2.6	-0.86
5.	-2.9	2.4	-0.88
6.	0.19	-0.093	-0.25
7.	0.003822	-0.052	0.0067
8.	0.13	-0.066	-0.048
9.	0.2	-0.024	-0.16
10.	0.1	-0.014	-0.090

Total no. of features for different fault location case= 60*209

Total no. of features for different fault resistance case= 60*220

Case 2: FEATURE EXTRACTION IN CNN

During the faults, there are sudden sharp changes which are noticed in the faulty lines of the system as system changes from the healthy state to a abnormal state as shown in figure

6.2and figure 6.3.Fault pattern is a non stationary signal. DWT helps to know the frequency of fault signal and the time of its occurance as well. In fault classification, discrete wavelet transform can be used as filter for extracting faulty harmonics by using db2 as a mother wavelet. The noise signals are high frequency signals. DWT deconstructs the fault signal into an approximate signal of low frequency and detailed signal of high frcquency components. A suitable threshold of 2.5 times the normal fault signal peak is used to filter high frequency noise signal. The detailed coefficients of voltage and currents are calculated using the equation (6.1)

$$D_j(t) = \sum_{k \epsilon Z} C(j,k)\, \phi_{j,k}(t) \tag{6.1}$$

$\phi_{j,k}(t)$represents wavelet function where $j$denotes scaling, k is translation, t is translation time, $C(j,k)$represents scale function. The different families of daubechiesare used to compare the accuracy of fault classifier for a given number of epochs during training of the CNN classifier but db2 is selected as best with high accuracy

6.3 TRAINING OF FAULT CLASSIFIER

Case 1: For ANN

Feed-forward Neural Network is selected as a classifier which is most commonly used ANN for the classification of faults [41] and the Levenberg-Marquardt Back-propagation is chosen as a training algorithm based on mean square error. The weight parameters of layers are updatedand error is calculated based on mean square error as expressed in equation (6.2).

The total mean square error is

$$E(x,w)=\frac{1}{2} \sum_{p=1}^{p} \sum_{m=1}^{m} e_{p.m.}^2 \tag{6.2}$$

Where x, w represent input vector and weight vector, m is number of outputs and p is number of patterns, $e_{p.m.}$ represents error at output m when applying pattern p& it is defined as in equation (6.3).

$$e_{p.m.} = target\ output – actual\ output \tag{6.3}$$

The fig.6.4 depicts thearchitecture of one neuronj which has connections with the other part of the network and nodes y_{ji} represents network inputs, w_{ji} represents the weights of neurons and w_{j0} weighted bias of neuron j & $F_{mj} y_{ji}$ is the non-linear relationship between the output node $y_{ji}$of neuron and the network output o_m.

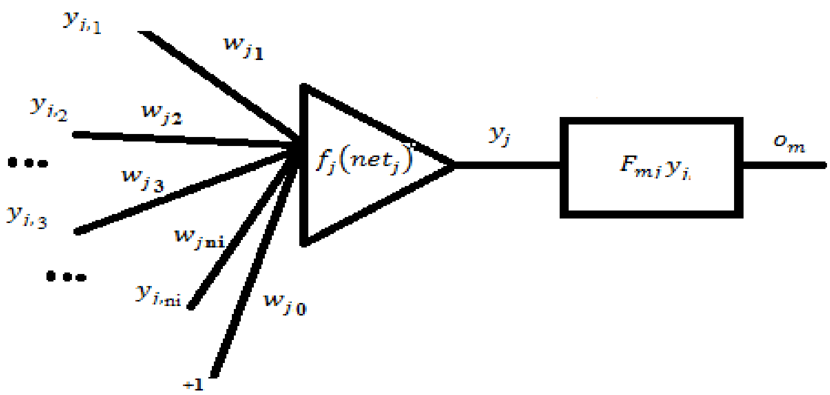


Fig.6.4 The neural network architecture

For forward computation (input to hidden to output layers)**:** Calculate net for all layers asexpressed in equation (6.4).

$$net_j = \sum_{i=1}^{ni} w_{ji}\ y_{ji}\ + \text{bias} \tag{6.4}$$

Calculate output as defined in equation (6.5).

$$y_j = f_j(net_j), \tag{6.5}$$

where f_j represents activation function for neuron j

Bipolar Sigmoid Activation function at the input to the hidden layer is expressed in equation (6.6).

$$f(x) = \frac{1-e^{-s_j net_j}}{1+e^{-s_j net_j}} \tag{6.6}$$

Linear Activation function at hidden to output layer is expressed in equation in (6.7)

$f(net_j) = net_j$(6.7)

Calculate slope as expressed in equation (6.8).

$$s_j = \frac{dy_i}{dnet_j} = \frac{df_i(net_j)}{dnet_j} \tag{6.8}$$

The slope is calculated for all outputs of the layers.

For backward computation (output to hidden to input layers):

Calculate error for all outputs of the layers.

Jacobian matrix calculation:

The Jacobian matrix is expressed in equation (6.9)

$$J = \begin{bmatrix} \frac{de_{i,1}}{dw_1} & \frac{de_{i,1}}{dw_2} & \frac{de_{i,1}}{dw_3} & \cdots & \frac{de_{i,1}}{dw_n} \\ & \vdots & & \ddots & \vdots \\ \frac{de_{i,m}}{dw_1} & \frac{de_{i,m}}{dw_2} & \frac{de_{i,m}}{dw_3} & \cdots & \frac{de_{i,m}}{dw_n} \end{bmatrix} \quad (6.9)$$

The size of matrix is $p*m*n$, p= no. of patterns, m= no. of outputs, n= no. of weights, i& j are represents weights (1 to n).

The each element of the matrix is calculated as expressed in equation (6.10)

$$\frac{\partial e_{p,m}}{\partial w_{j,i}} = \frac{\partial(d_{p.m}-o_{p.m})}{\partial w_{j,i}} = -\frac{\partial o_{p,m}}{\partial w_{j,i}} = -\frac{\partial o_{p,m}}{\partial y_j}\frac{\partial y_j}{\partial net_j}\frac{\partial net_j}{\partial w_{j,i}} = -F_{mj}s_j y_{j,i} \quad (6.10)$$

Where F_{mj} is the derivative of non-linear function between output m and neuron j.

Calculation of updated weights.

The weight parameters are adjustedas expressed in equation (6.11).

$$w_{k+1} = w_k - (J_k^T J + \mu I)^{-1} J_k e_k \quad (6.11)$$

Wherew_k= current weight, w_{k+1}= next weight, e_k = last total error, J = Jacobian matrix, μ is combination coefficient, I = identity matrix.

The steps for training neural network for classification of shunt faults are discussed below in terms of flow chart as shown in fig.6.5.

i. Initialise the weights randomly in the range [-1 to 1].
ii. Then calculate the total mean square according to the initial weights which are randomly generated.
iii. The weights are adjusted and updated as given by Equation (6.11)
iv. Again calculate the total mean square error with the updated weights.
v. Now if the total MSE is greater than the desired value, then retract the step by increasing combination coefficient μ by a factor of 10 and go to step ii and update the weights.
vi. But if the total MSE is decreased then accept the step as it is and decrease the coefficientμ by dividing with 10.
vii. Next go to step ii with updated weights until the total MSE is less than the desired set value.

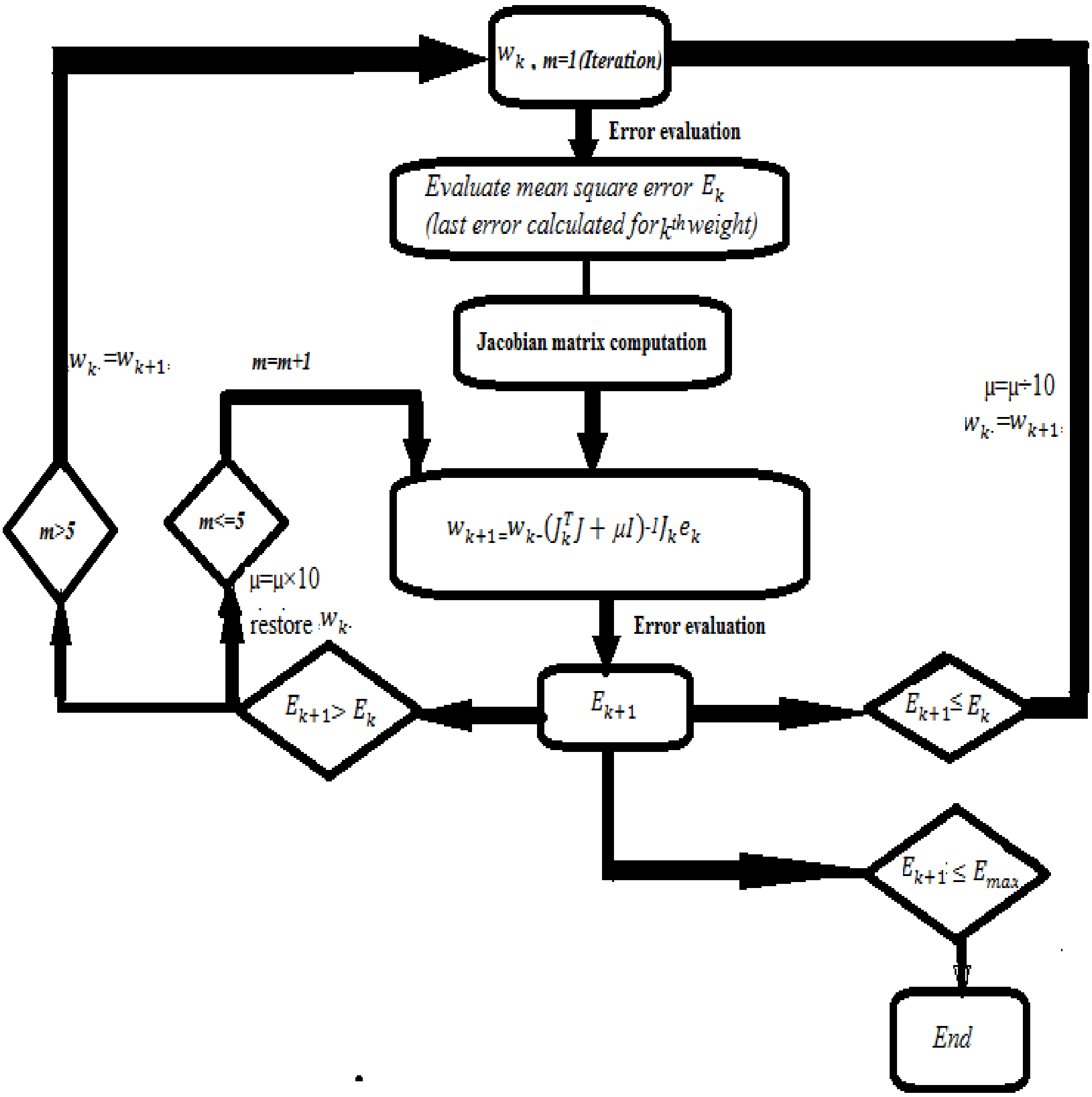


Fig.6.5 The flow chart for training neural network using LM algorithm

The neural network configuration of neural network for the faults classification is shown in fig.6.6

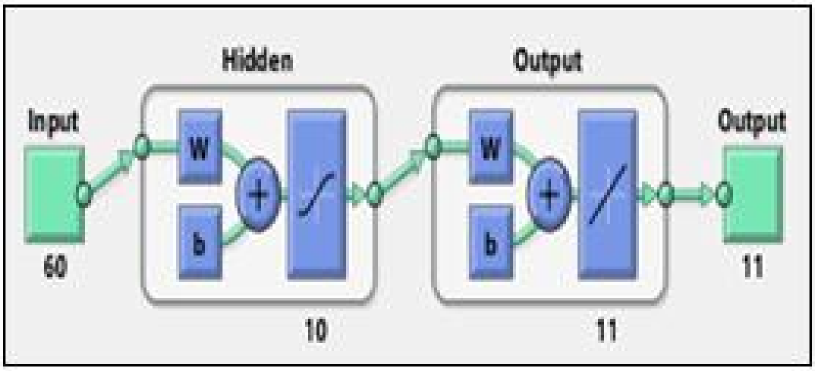


Fig.6.6 Neural network configuration for fault classification training

The table 6.5 shows the number of parameters used in architecture of ANN

Table 6.5 Number of parameters for ANN architecture

INPUT	No. of input layers	1
	No. of input neurons	60
	No. of input patterns for fault location	60*209
	No. of input patterns for fault resistance	60*220
HIDDEN LAYER	No. of hidden layers	2
	No. of hidden neurons	10
OUTPUT	No. of output layers	1
	No. of output neurons	11
	No. of output patterns for fault location	11*209
	No. of output patterns for fault resistance	11*220
TARGET	No. of targets (ABCG, ABC, ABG, AB, ACG, AC, AG, BCG, BC, BG, CG)	11

Case 2: For CNN

Convolutional neural network topology for fault classifier is proposed with eight layers as shown in fig.6.7. The 8 layers are: Input layer, Convolution 2-D layer, Batch Normalization layer, ReLU activation layer, MAX Pool layer, Fully connected layer followed by softmax function as expressed in above equation (5), and classification layer.

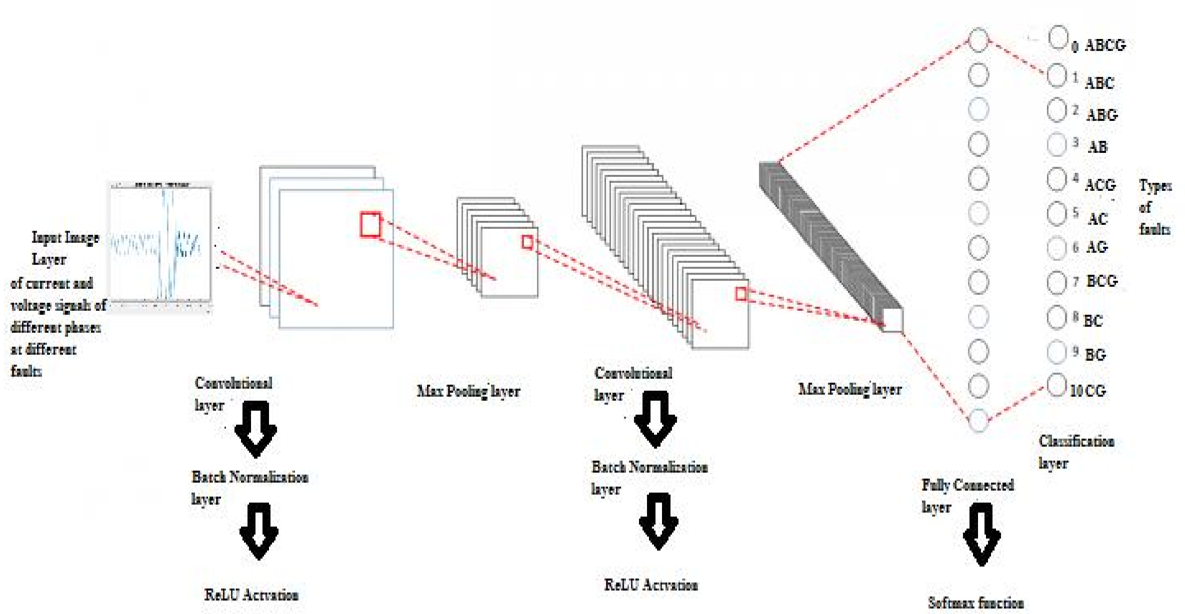


Fig.6.7 Proposed Architecture of CNN as fault classifier

Input Image layer: An image input layer inputs 2-D images of current and voltage signals of each phase at different faults locations and at different fault impedances to a network of CNN and then normalized the data with zero centre method.

Convolutional 2D layer: This layer uses the convolution process as by applying filters to the input image by sliding the filters vertically and horizontally and calculating the dot product of the weights with the input applied and then adding bias to it.

Batch Normalization Layer: This layer helps in normalizing each input across a mini-batch which further helps in speed up of training process of CNN.It is always used between convolution layer and the ReLU layer.

Input: values of x over a mini-batch ;$B = \{x_1, x_2, ...x_n\}$, parameters to be learned β, γ

The mini-batch mean is expressed in equation 6.12.

$$\mu_\beta = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{6.12}$$

Then calculate the mini-batch variance whichis expressed as in equation 6.13

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_\beta)^2 \tag{6.13}$$

Normalize the input values as expressed in equation 6.14.

$$\varkappa = \frac{x_i - \mu_\beta}{\sqrt{\sigma_B^2 - \epsilon}} \tag{6.14}$$

Then the output after scaling and shifting is expressed in equation 6.15

$$y_i = \gamma\varkappa + \beta = BN_{\gamma,\beta}(x_i) \tag{6.15}$$

ReLUlayer:AReLU layer is a activation layer which sets the threshold values to each of the inputs, which gives value less than zero is set to zero. The mathematically form of this activation function is expressed in equation 5.7.

Max Pooling layer: This layer performs down-sampling which divides the input into rectangular pooling regions, and calculating the maximum of each regions.

Fully Connected Layer:A fully connected layer performs similar to the layers which are hidden layers used in ANN which multiplies the input values of inputs with weight matrix and then adds a bias values

SoftmaxLayer: Asoftmax layer is used with softmax function to the input. The output at the softmax layer indicates the probabilities between 0 to 1 for the different classes at output layer. The softmax function is expressed in above equation 5.9.

Classification Layer: The final layer is the classification layer for classify all shunt types of faults. The probabilities calculated in above softmax layer is used in this layer for each input values to assign the input to one of the mutually exclusive classes and calculatethe cross entropy loss for different classification tasks.

For training process, the parameters setting at various layers of convolutional neural network is as given in table 6.6.

Table 6.6 Parameters of CNN classifier

Parameter	**Value**
Input size	[1334, 6, 1]
Filter size	[3, 3]
Number of filters	8
Learning rate (weights)	1
Epsilon	$1.0000*e^{-05}$
Pool size	[2, 2]
Input (Fully Connecetd layer)	16008
Output size	11
Weights matrix	11*16008
Output Class matrix	11*1
Output Label (for 11 shunt type of faults)	0,1,2,3,4,5,6,7,8,9,10
Loss function	Crossentropyex

The proposed topology is depicted in the flow chart shown in figure 6.8.

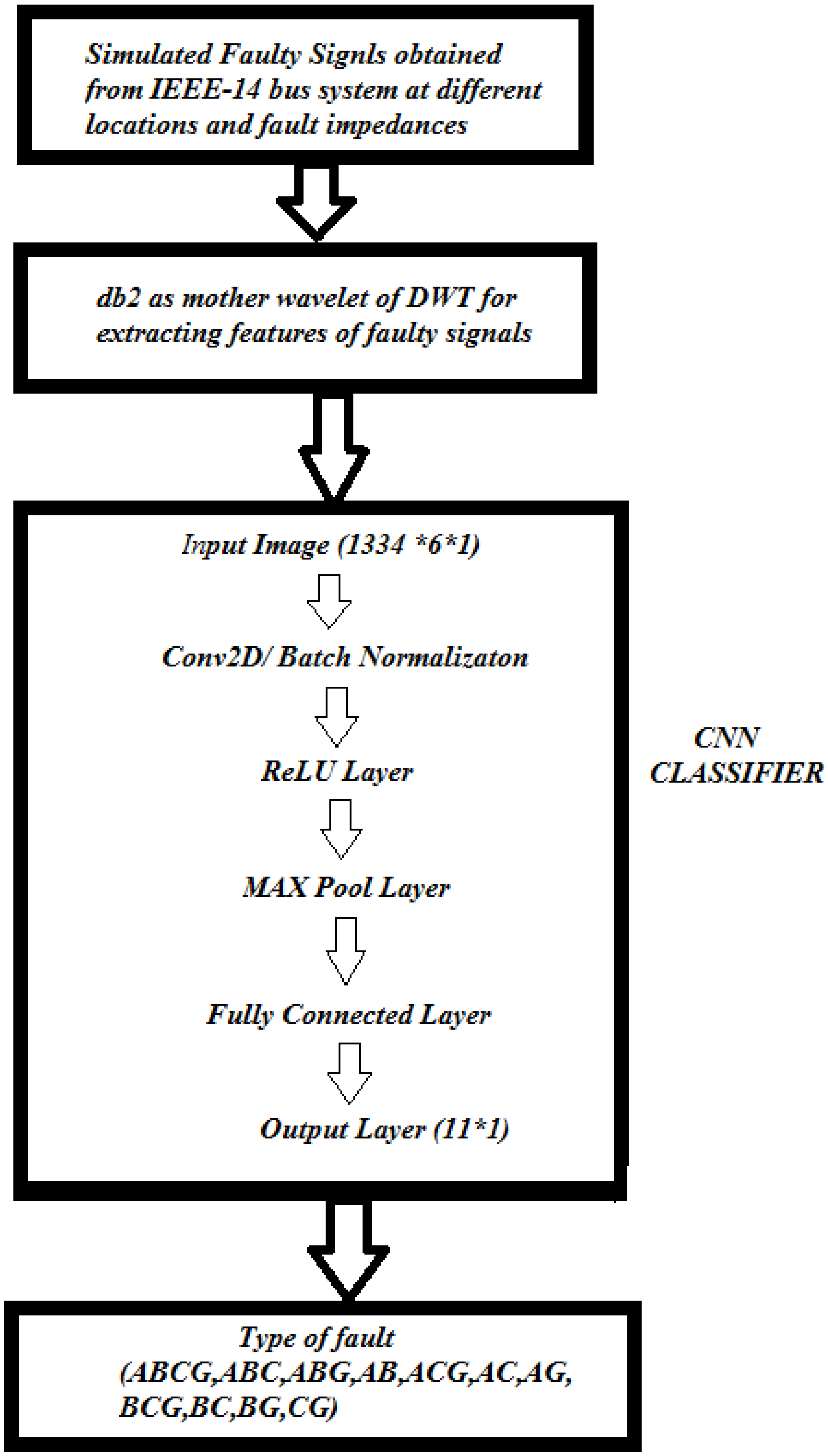


Fig.6.8 Flowchart for Fault Classification using CNN

CHAPTER 7

RESULTS AND DISCUSSIONS

7.1 SIMULATION RESULTS

Case 1: Simulation Results for fault impedance variations

The results obtained during simulation of IEEE-14 bus system as shown in figure 7.1, 7.2,7.3 and 7.4 atvarious resistances.The given current waveform containing 20 cycles for 0.4 seconds. The first 10 cycles are normal and next 10 cycles are abnormal. As seen from the diagram the fault is occurring at 0.2 seconds containing transient and sub-transient components of current. The current of phase A sharply rises with line to ground fault at phase A and then settled to normal after 0.3 seconds for 15 ohm resistance but the current rise for 75 ohm is low as the fault impedance increases. Similarly, this form of data is collected for various resistances for all types of shunt faults.

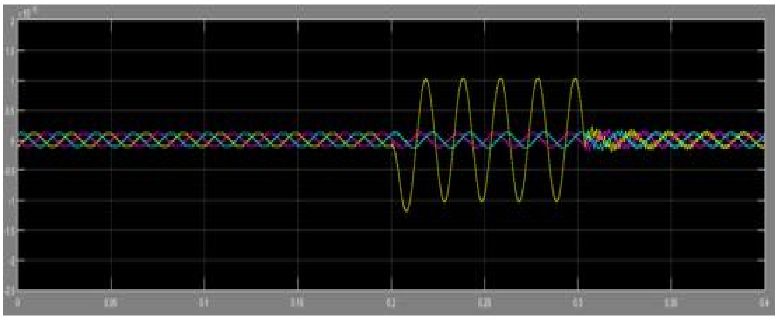


Fig.7.1 Current waveforms during line to ground (AG) fault at 15-ohm resistance

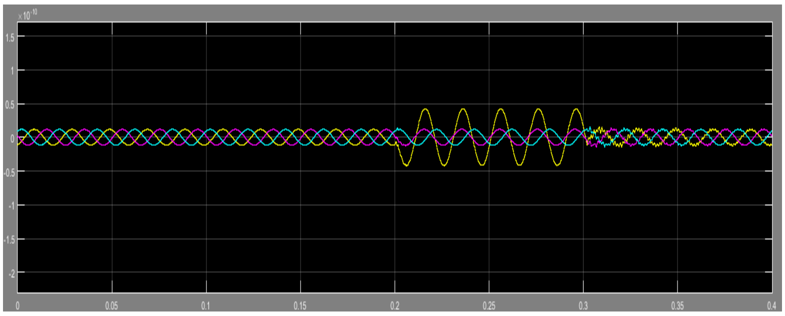


Fig.7.2 Current waveforms during line to ground (AG) fault at 75-ohm resistance.

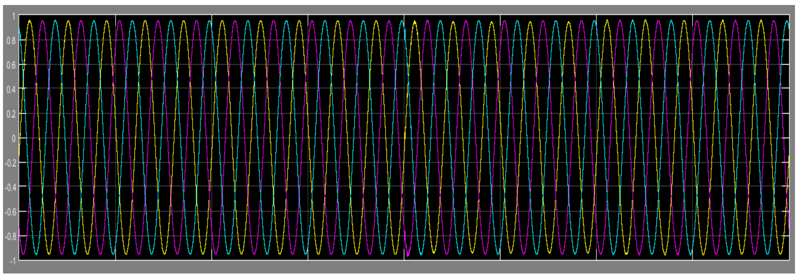


Fig.7.3 Voltage waveform during line to ground (AG) fault at 15-ohm resistance

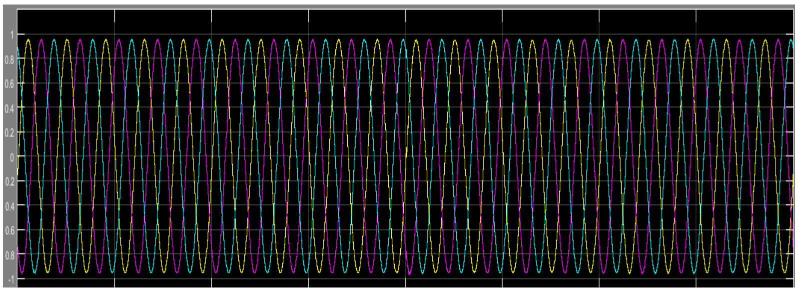


Fig.7.4 Voltage waveform during line to ground (AG) fault at 75-ohm resistance

Case 2: Simulation Results for fault location variations

The results obtained during simulation of IEEE-14 bus system are shown in figure 7.3, 7.4 atvarious locations. The given current waveform contains 20 cycles for 0.4 seconds. The first 10 cycles are normal and next 10 cycles are abnormal. As seen from the diagram the fault occurs at 0.2 seconds containing transient and sub-transient components of current. The current of phase A rises sharply with line to ground fault at phase A for 120 km than the current rise for location at 90 km is low as the location varies. Similarly, this form of data is collected for various locations for all types of shunt faults.

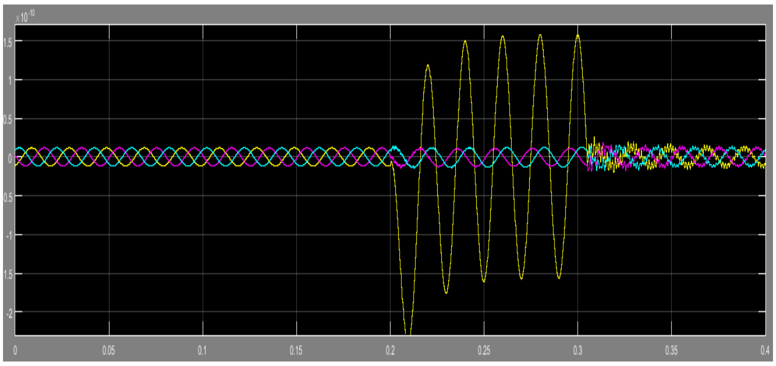


Fig.7.5 Current waveforms during line to ground (AG) fault at 120 km

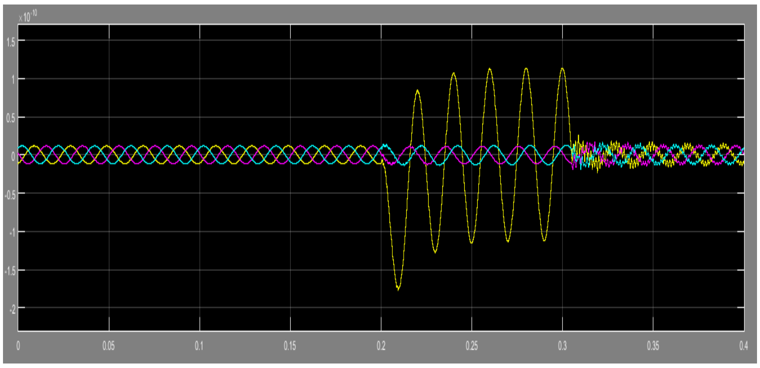


Fig.7.6 Current waveforms during line to ground (AG) fault at 90 km

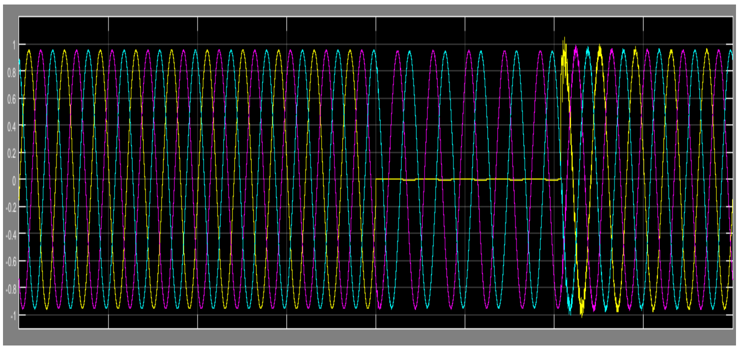


Fig. 7.7 Voltage waveform during line to ground (AG) fault at 120 km

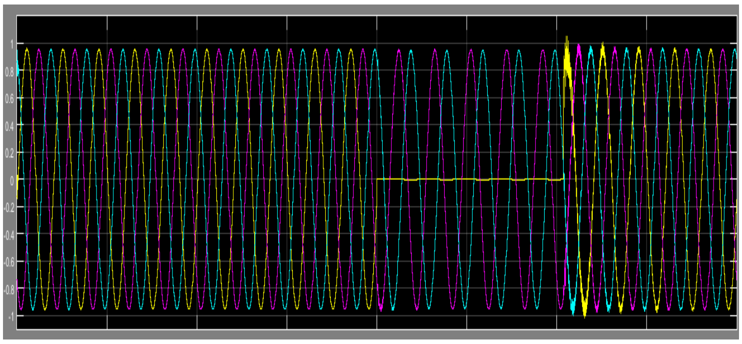


Fig. 7.8 Voltage waveform during line to ground (AG) fault at 90 km

From the simulation results, there are different values of current during fault conditions at time 0.2 seconds which are required as training data to train the ANN classifier. Similarly, different voltage values are taken as shown in figure 7.7 and 7.8.

7.2 FAULT CLASSIFIER RESULTS

7.2.1 ANN AS FAULT CLASSIFIER:

The fig.7.5 and 7.9 shows the confusion matrix for ANN fault classifier for different fault impedances and for different fault locations. Actually, the confusion matrix gives the accuracy of ANN classifier. It will list the correct classifications for all types of shunt faults. The 11*11 matrix depicts the all types of shunt faults in the order as which is shown in table 7.1. The right outer column and bottom outer row of the matrix gives the individual fault accuracy with error percentage.

Table 7.1 Labels of all shunt types of faults for ANN classifier

Label of fault	Type of fault
1	ABCG
2	ABC
3	ABG
4	AB

5	ACG
6	AC
7	AG
8	BCG
9	BC
10	BG
11	CG

Case 1: For fault impedance variations

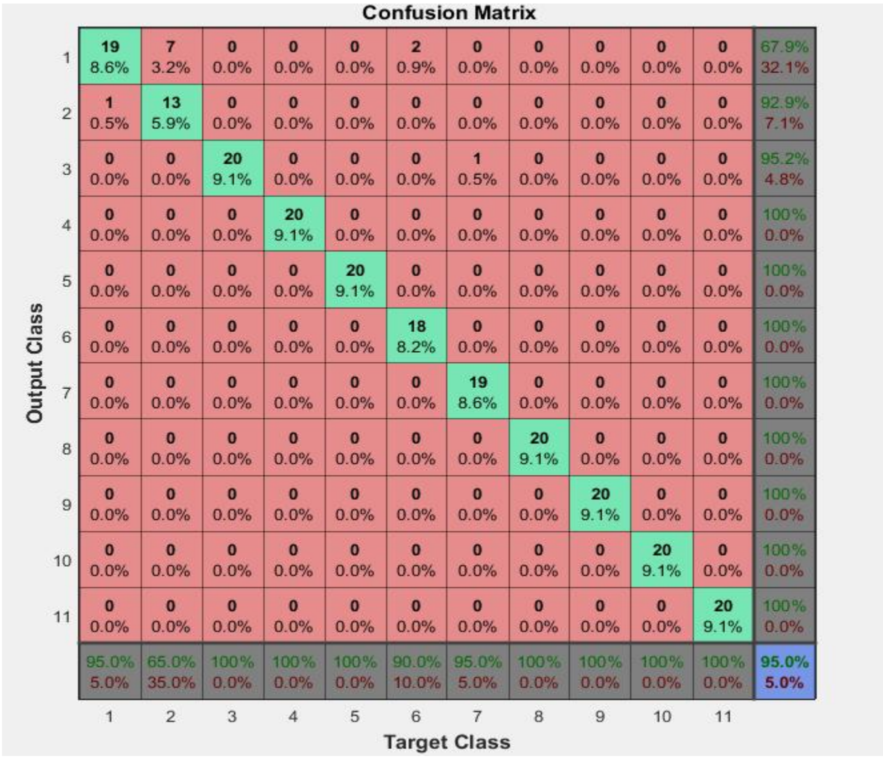


Fig.7.9 Confusion matrix showing an overall accuracy of 95% for different types of faults.

From the above fig.7.9, the overall accuracy for the classification of faults is 95% which means 95% data is correctly classified.

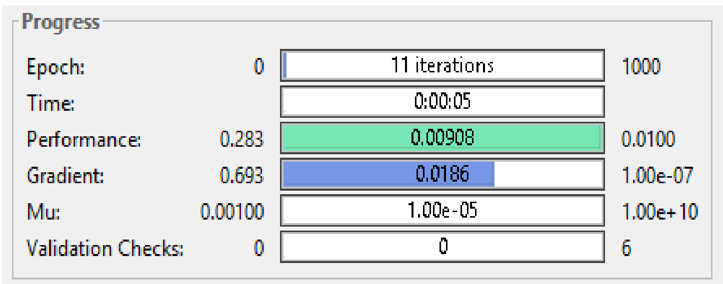


Fig.7.10 Training state configuration

For fault classification, at different resistances, a neural network takes 11 epochs during the training of neural network and mean square error becomes minimumof 0.00908 as shown in fig.7.10. The error obtained during training of ANN classifier in terms of error histogram is shown in fig.7.11 and the training state during classification of faults as shown in fig.7.12

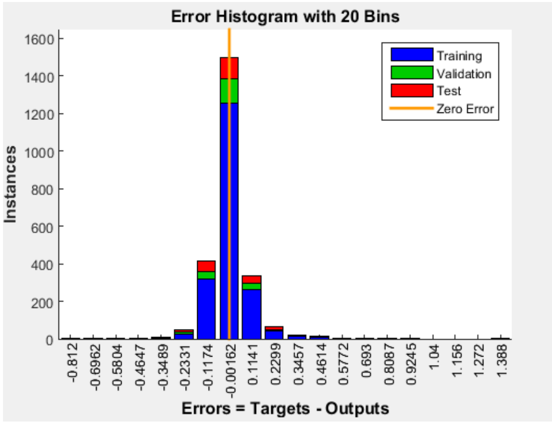


Fig.7.11 Error histogram

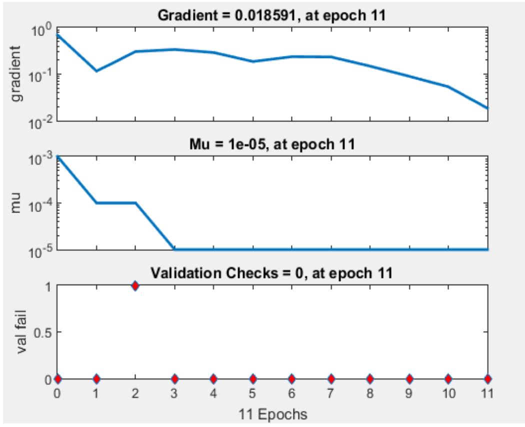


Fig.7.12 Training state plot

Case 2: For fault location variations:

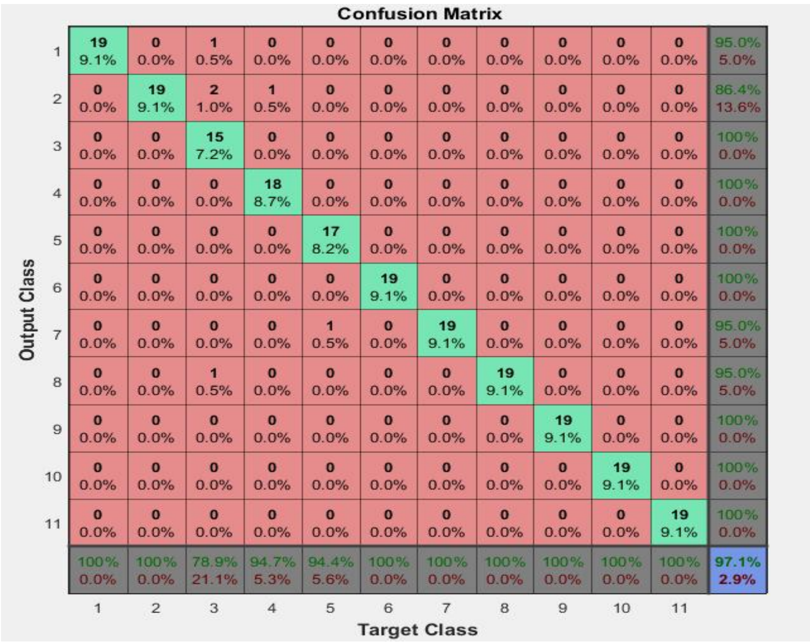


Fig.7.13 Confusion matrix showing overall 97.1% accuracy for all types of faults

From the above fig.7.13, the overall accuracy for the classification of faults is 97.1% which means 97.1% data is correctly classified.

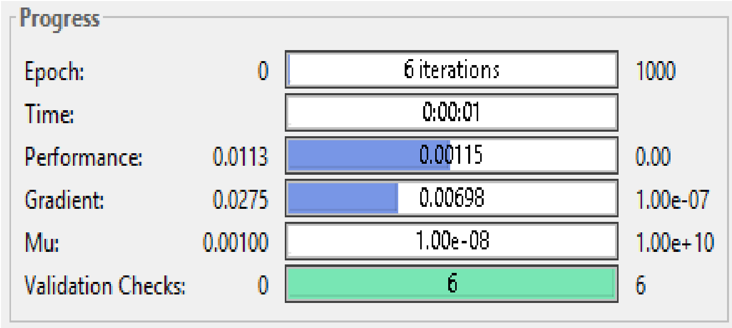


Fig.7.14 Training state configuration

For fault classification, at different fault locations, the neural network takes six epochs with less training time during the training of neural network and mean square error becomes minimumof 0.00698 as shown in fig.7.14. The error obtained during training of ANN classifier in terms of error histogram is shown in fig.7.15 and the training state during classification of faults as shown in fig.7.16.

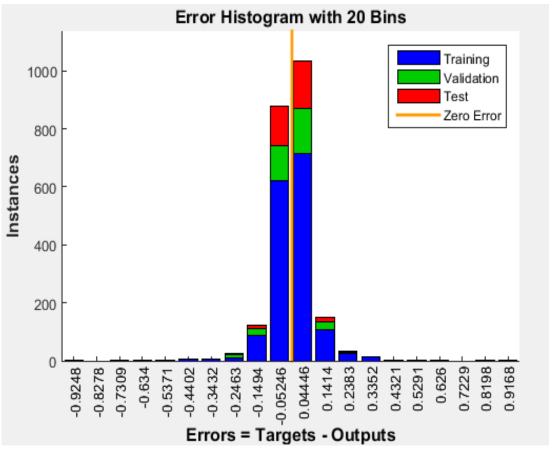


Fig.7.15 Error histogram

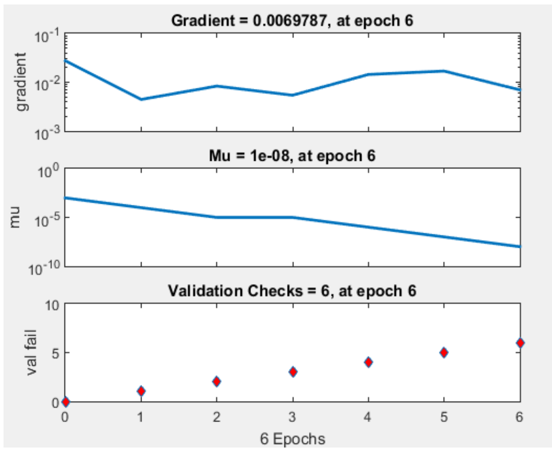


Fig.7.16 Training state plot

7.2.2 CNN AS FAULT CLASSIFIER:

Case 1: For different fault impedances

The db2 is considered as mother wavelet for ABCG fault at fault impedance 55 ohm. Figures 7.17, 7.18 and 7.19 shows the implementation of db2 as mother wavelet on the current signals. Similarly db2 as DWT is used to transform the voltage signals to calculate the detailed coefficients. Then these 6 calculated values of current and voltage are used to train the CNN fault classifier at various fault impedances.

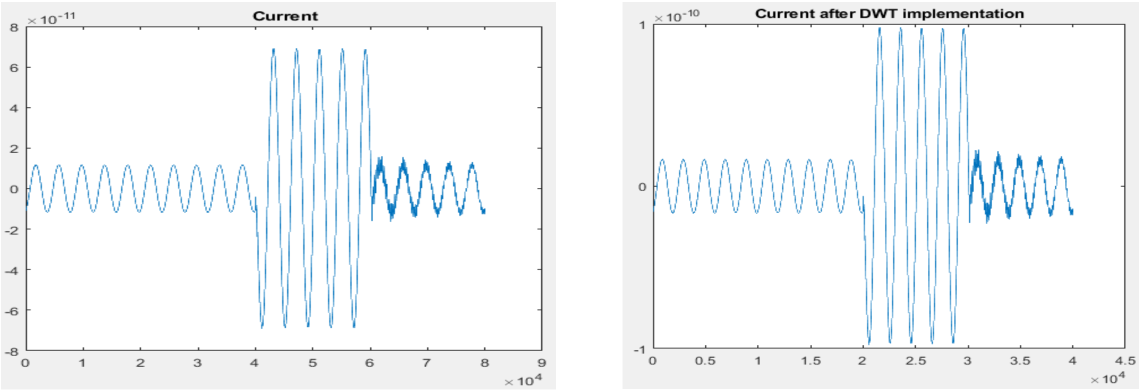


Fig.7.17 Current Waveforms of phase A before and after DWT

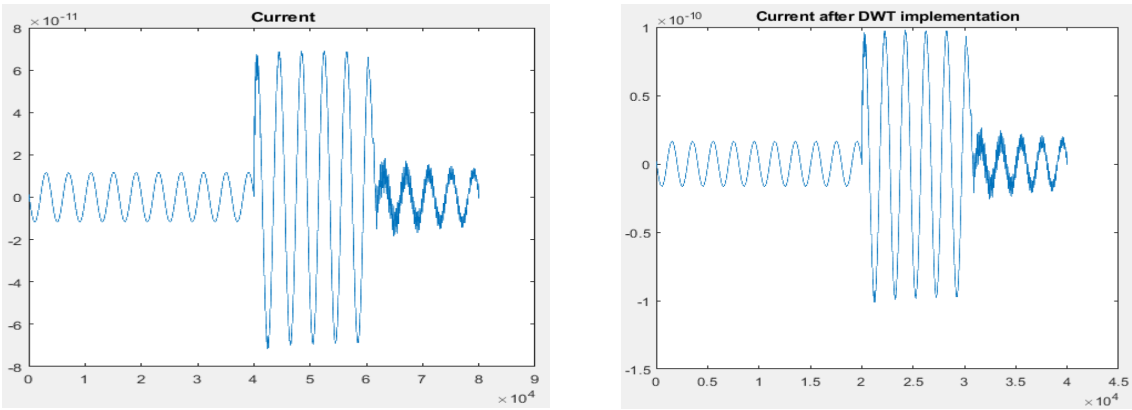


Fig.7.18Current Waveforms of phase B before and after DWT

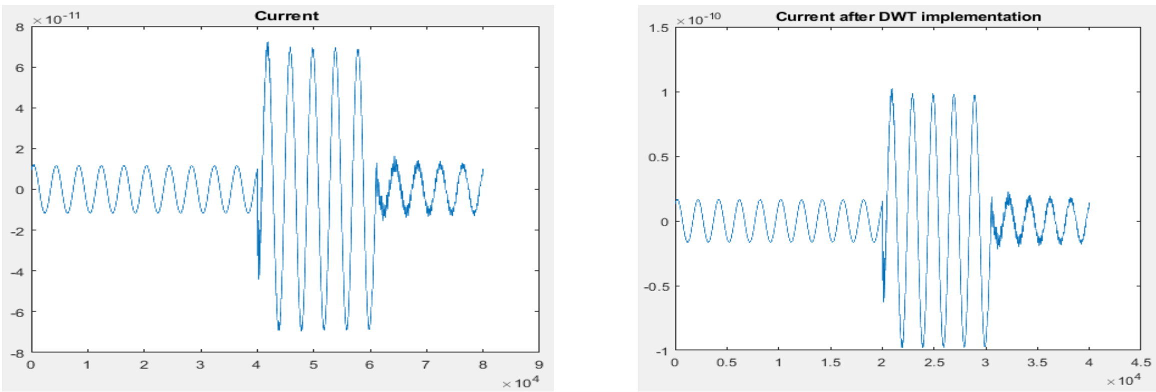


Fig.7.19 Current Waveforms of phase C before and after DWT

The performance of the CNN as fault classifier is in terms of accuracy in detection of individual fault is shown in table 7.2. The plot between average accuracy of classification of eleven different kind of shunt type faults at different fault impedances shows that a reasonable accuracy of 98.18% is achieved in nearly 25 epochs amounting to batch accuracy of 100% in 30 epochs and the batch loss reduces from 2.4 units to 0.0234 units as shown in table 3. That obtained is quite satisfactory.

Table 7.2 Accuracy for shunt type of faults

Type of fault	Target Label	Accuracy(%)
ABCG	0	100%
ABC	1	100%
ABG	2	90%
AB	3	90%
ACG	4	100%
AC	5	100%
AG	6	100%
BCG	7	100%
BC	8	100%
BG	9	100%
CG	10	100%
Overall accuracy		**98.18%**

The overall accuracy of the CNN fault classifier is 98.18% for different fault impedances as shown in table 7.2.

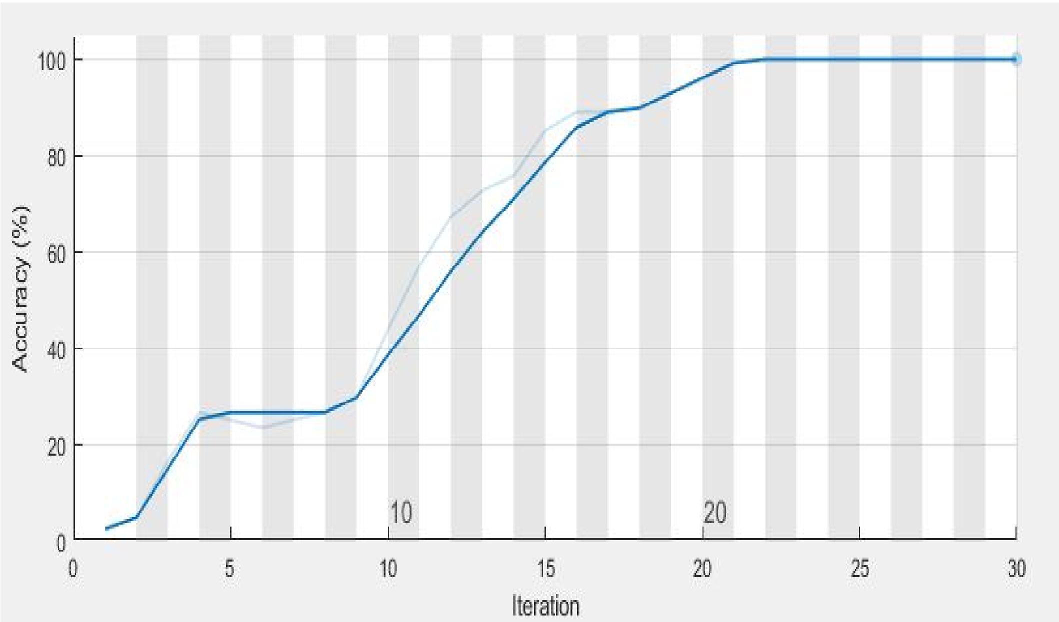


Fig.7.20 Graph showing accuracy with iterations during training of CNN fault classifier at various fault impedances

The above graph shown in figure 7.20 gives the overall accuracy of different types of faults with maximum 30 iterations.

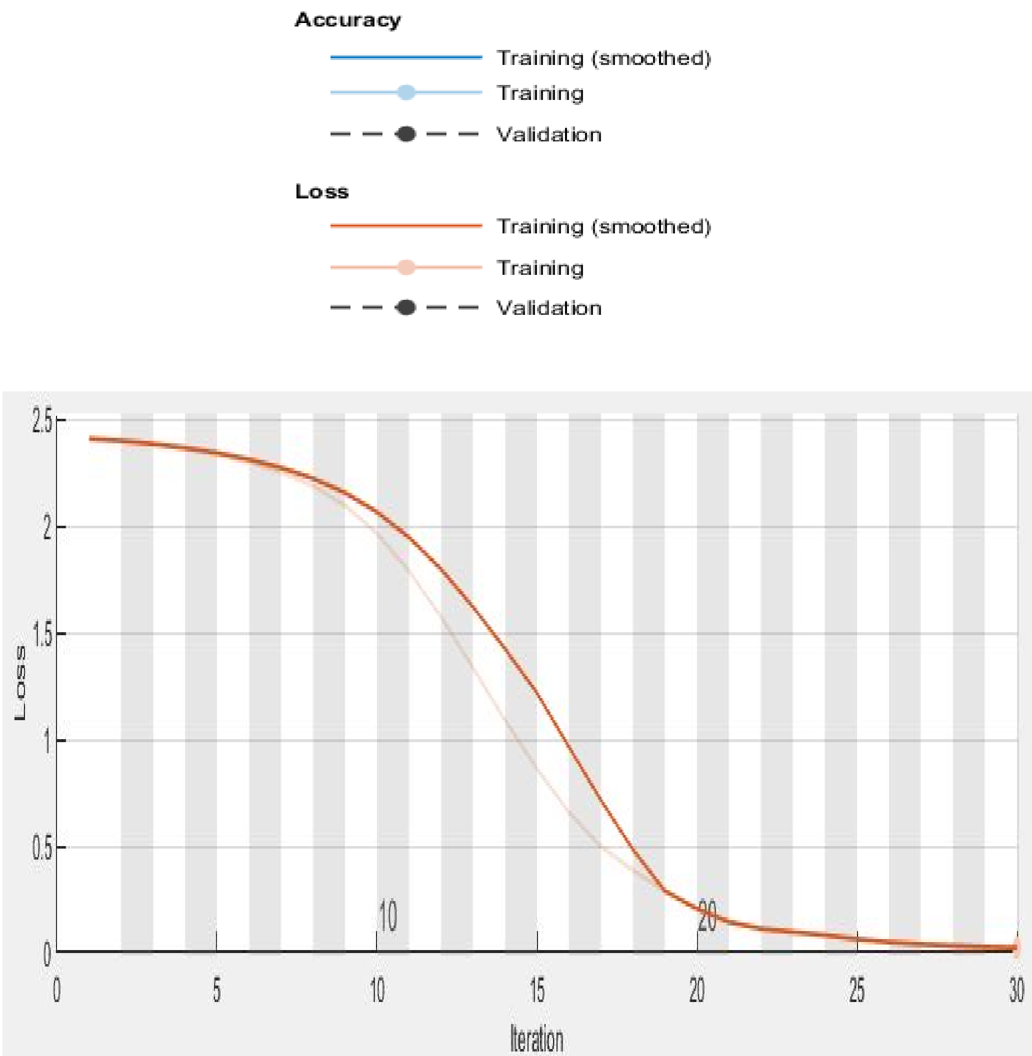


Fig.7.21 Mini-Batch loss during training of CNN fault classifier at various fault impedances

The above graph shown in fig.7.21 shows the loss during the training of CNN fault classifier at constant base learning rate 0.0100.

Table 7.3 Performance of the CNN fault classifier

Epoch	Time elapsed (seconds)	Mini Batch accuracy (%)	Mini Batch loss
1	6	2.34%	2.4096
30	32	100.00%	0.0234

The results in table 7.3 give the assessment of the time elapsed during training of CNN as fault classifier at various fault impedances with mini-batch loss at base learning rate 0.01.

Case 2: For different fault locations

The db2 as mother wavelet for AB fault at fault location 10 km. The three figures 7.18, 7.19 and 7.20 shows the implementation of db2 as mother wavelet on the current signals. Similarly db2 as DWT is used to transform the voltage signals to calculate the detailed coefficients. Then these 6 calculated values of current and voltage are used to train the CNN fault classifier at various fault locations.

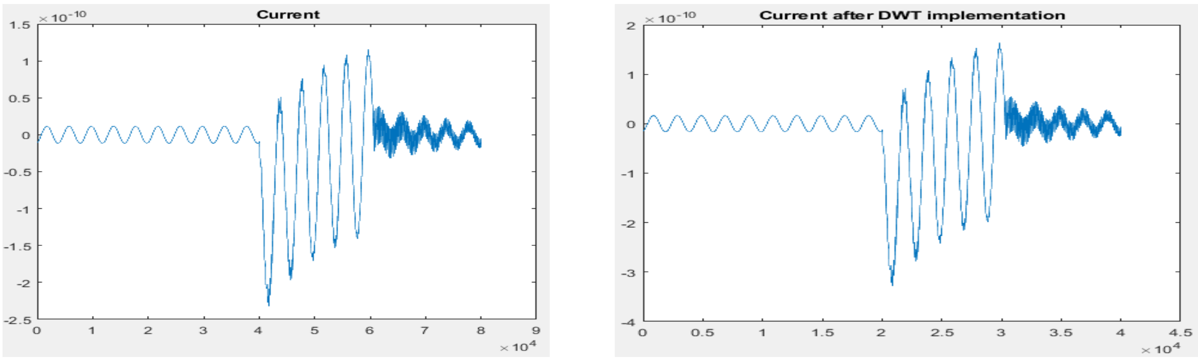


Fig.7.22 Current Waveform of phase A before and after DWT implementation

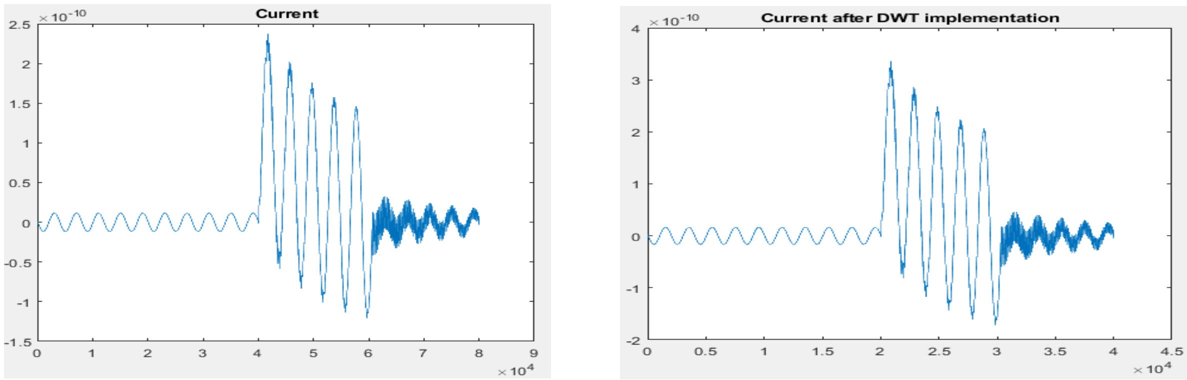


Fig.7.23 Current Waveform of phase B before and after DWT implementation

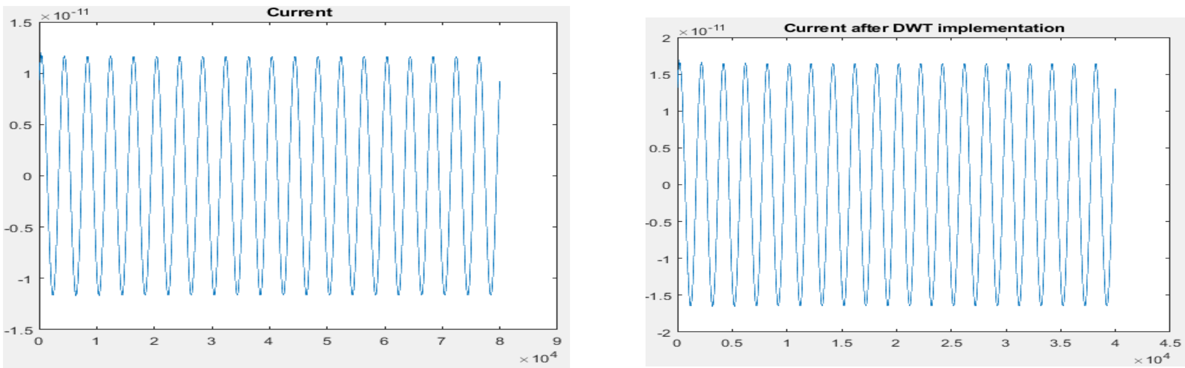


Fig.7.24 Current Waveform of phase C before and after DWT implementation

The performance of the CNN as fault classifier is in terms of accuracy in detection of individual fault is shown in table 7.4. The plot between average accuracy of classification of eleven different kind of shunt type faults at different fault location shows that a reasonable accuracy of 96.63% is achieved in nearly 18 epochs amounting to batch accuracy of 99.22% in 30 epochs and the batch loss reduces from 2.4 units to 0.0224 units as shown in table 7.4 which is quite satisfactory.

Table 7.4 Accuracy for shunt type of faults

Type of fault	Target Label	Accuracy (%)
ABCG	0	100%
ABC	1	94.74%
ABG	2	94.74%
AB	3	100%
ACG	4	94.74%
AC	5	94.74%
AG	6	100%
BCG	7	94.74%
BC	8	89.47%
BG	9	89.47%
CG	10	100%
Overall accuracy		**96.63%**

The overall accuracy of the CNN fault classifier is 96.63% for different fault locations as shown in table 7.4

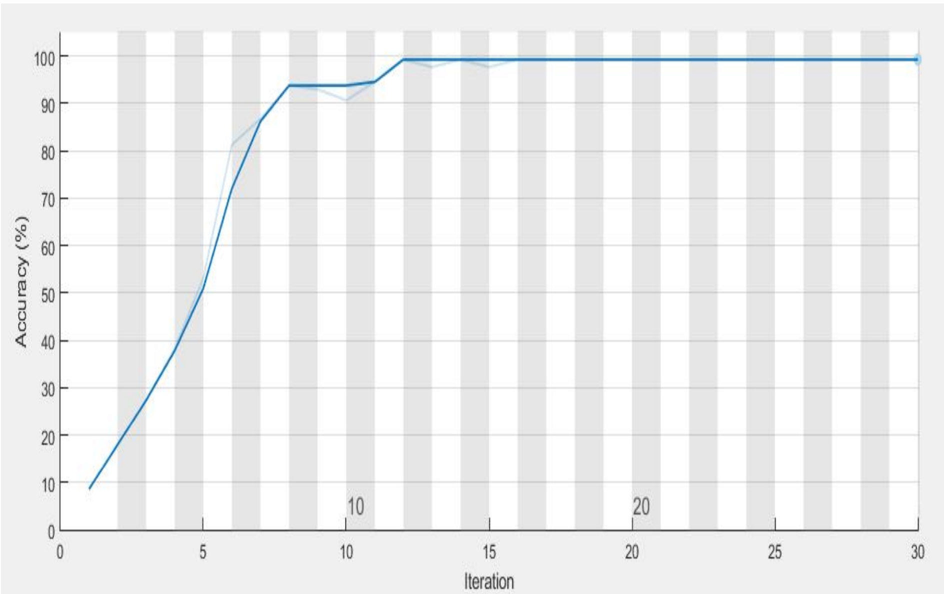


Fig.7.25 Graph showing accuracy with iterations during training of CNN fault classifier at various fault locations

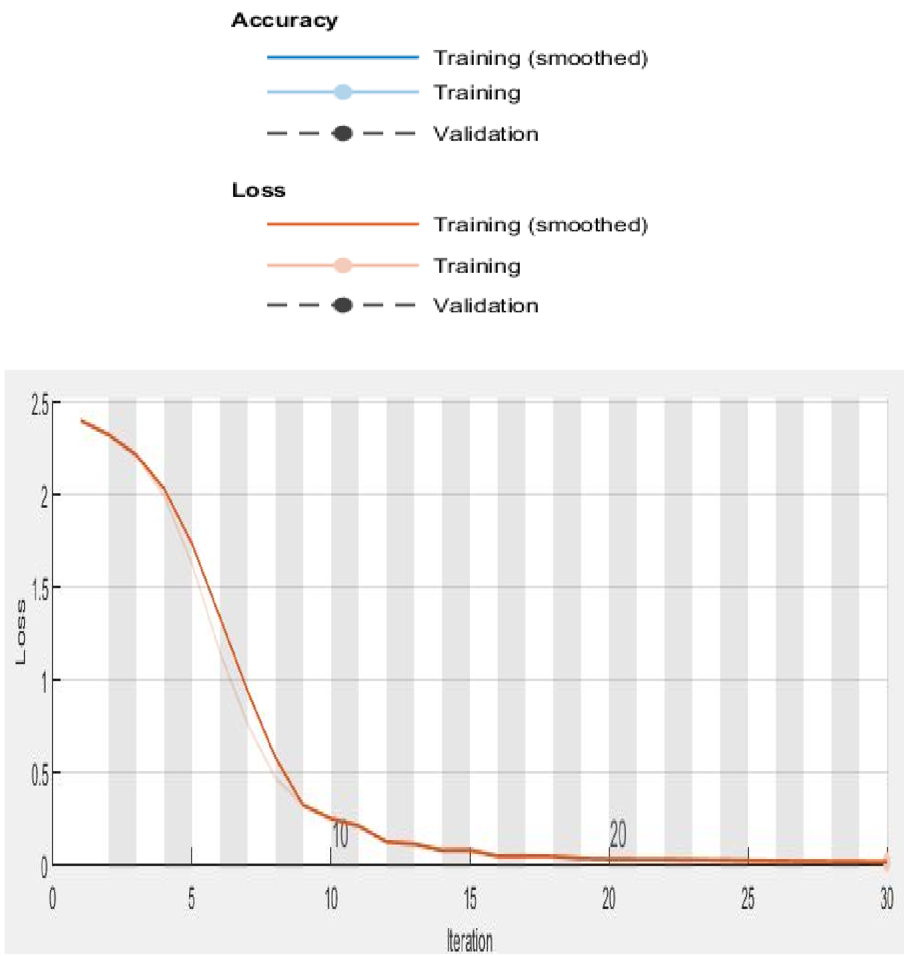


Fig.7.26 Mini-Batch loss during training of CNN fault classifier at various fault locations

The graph shown in figure 7.25 gives the overall accuracy of different types of faults with maximum 30 iterations.

The graph shown in figure 7.26 shows the loss during the training of CNN fault classifier at constant base learning rate 0.01.

Table 7.5 Outputs of the CNN fault classifier during training

Epoch	Time elapsed (seconds)	Mini Batch accuracy (%)	Mini Batch loss
1	10	8.59	2.4042
30	57	99.22	0.0224

The table 7.5 gives the assessment about the time elapsed during training of CNN as fault classifier at various fault locations with mini-batch loss at base learning rate 0.0100.

CHAPTER 8

CONCLUSION AND FUTURE WORK

8.1 CONCLUSION

In this dissertation, problem of fault classification on IEEE 14 Bus system is studied using two approaches namely; artificial neural network and convolutional neural network (CNN). CNN is hybridised with wavelet transform to filter transients and other unwanted signals. The accuracy obtained with both the approaches are high but the accuracy obtained by CNN technique is higher than basic technique based on ANN as CNN isthe best at feature extraction and involve less parameters than ANN. CNN has multiple hidden layers so it leads to reduced computation complexity and gives better performance for the same problem. With CNN fault classifier, the overall average accuracy of 98.18% is obtained in case when variation in fault impedance is considered, however it is 96.63% in case when different loctaions of fault are considered in IEEE 14 bus system.

8.2 FUTURE SCOPE

The analysis of data driven mathematical models using deep learning network may beextended to validation using results from experiment, which can be obtained using laboratory scale AC network set up. The proposed methodology can be applied to investigate the type of disturbances namely; load disturbance, failure of source, failure of synchronised operation of circuit breakers.

PUBLICATIONS

[1] H. Kaur, M. Kaur, "Fault classification in a transmission line using Levenberg Marquardt Algorithm based Artificial Neural Network,"*Computing, Power and Communication technologies,* 2019, Springer (Accepted)

[2] H. Kaur, M. Kaur , "Classification of Shunt Type Faults using Convolutional Neural Network," 2019 International Workshop on Artificial Intelligence and Deep Learning methods for Human Centric Systems (AIDL-HCSy),Oct.28 - 30, 2019 DUBLIN, IRELAND (Submitted)

REFERENCES

[1] Zhu Yongli, Y. H. Yang, B. W. Hogg, W. Q. Zhang, and S. Gao, “An expert system for power systems fault analysis,” *IEEE Trans. Power Systems*, vol. 9, no. 1, pp. 503–509, Feb. 1994.

[2] X. Lin, P. Mao, H. Weng, B. Wang, Z. Q. Bo, and A. Klimek, “Study on Fault Location for High Voltage Overhead Transmission Lines Based on Neural Network System,”, *International Conference on Intelligent Systems Applications to Power Systems*, 2007, pp. 1–5, 2007

[3] Eisa Bashier M Tayeb , “Faults Detection in Power Systems Using Artificial Neural Network,” *American Journal of Engineering Research*,vol. 2, pp. 69-75, 2013.

[4] A. Nag and A. Yadav, “Fault classification using Artificial Neural Network in combined underground cable and overhead line,” in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, 2016, pp. 1–4.

[5] Omar A. S. Youssef, "Fault classification based on wavelet transforms", *Transmission and Distribution Conference and Exposition, IEEE/PES*, vol. 1, pp. 531-536, 2001.

[6] Omar A. S. Youssef, "New algorithm to phase selection based on wavelet transforms", *IEEE Transactions on Power delivery*, vol. 17, no. 4, pp. 908-914, 2002.

[7] D. Chanda, N. K. Kishore, and A. K. Sinha, “Application of wavelet multiresolution analysis for classification of faults on transmission lines,” in *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*, vol. 4, pp. 1464–1469, 2003.

[8] K.Saravanababu , P.Balakrishnan , K.Sathiyasekar, “Transmission line fault detection, classification, and location using Discrete Wavelet Transform, ” *IEEE In Power, Energy and Control (ICPEC) International Conference*, pp.233-238 , 2013.

[9] M. Saini, A. Mohd Zin, M. Mustafa, A. Sultan, and R. Nur, “Algorithm for Fault Location and Classification on Parallel Transmission Line using Wavelet based on Clarke’s Transformation,” *Int. Journal of Electrical & Computer Engineering.*, vol. 8, pp. 983-993, 2018.

[10] B. Das and J. V Reddy, “Fuzzy-logic-based fault classification scheme for digital distance protection,” *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 609–616, Apr. 2005.

[11] C. Cecati and K. Razi, “Fuzzy-logic-based high accurate fault classification of single and double-circuit power transmission lines,” in *International Symposium on Power Electronics Power Electronics, Electrical Drives, Automation and Motion*, 2012, pp. 883–889.

[12] Majid Jamil, RajveerSingh, Sanjeev and Kumar Sharma,“Fault identification in electrical power distribution system using combined discrete wavelet transform and fuzzy logic,” *Journal of Electrical Systems and Information Technology.*, vol. 2, no. 2, 2015.

[13] M. Jamil, A. Kalam, A. Q. Ansari, and M. Rizwan, “Generalized neural network and wavelet transform based approach for fault location estimation of a transmission line,” *Applied Soft Computing*, vol. 19, pp. 322–332, 2014.

[14] Ashish Maheshwari, Vinesh Agarwal, Sanjeev Kumar Sharma, “ Transmission line fault classification using Artificial Neural Network based fault classifier,” *International Journal of Electrical Engineering & Technology*, vol. 9 , pp. 170-181, 2018

[15] J. Zhang, Z. He, S. Lin, Y. Zhang, and Q. Qian, "An ANFIS-based fault classification approach in power distribution system," *International Journal of Electrical Power & Energy Systems*, vol. 49, pp. 243-252, 2013

[16] Parikha UB, Dasb B, Maheshwarib R,“Fault classification technique for series compensated transmission line using support vector machine,” *International Journal of Electical Power& Energy Systems*, vol. 32, pp. 629-636, 2010.

[17] H. Livani and C. Y. Evrenosoğlu, “A fault classification method in power systems using DWT and SVM classifier,” in *PES : Transmission & Distribution in 2012*, pp. 1–5, 2012

[18] Pushkar Tripathi, G.N.Pillai , H.O. Gupta , “New method for fault classification in TCSC compensated transmission line using GA tuned SVM,” *IEEE International Conference on Power Electronics*, pp.1-6 , October 2012.

[19] A. K. Sinha and Kranthi Kiran Chowdoju, “Power system fault detection classification based on PCA and PNN,” *International Conference on Emerging Trends in Electrical and Computer Technology*in *2011*, pp. 111–115, 2011

[20] S. Mishra and A. Baral, “Classification of power system faults using voltage Concordia pattern feature aided PNN,” *IEEE 6th International Conference on Power Systems (ICPS in 2016,* pp. 1–6, 2016.

[21] Ray, Papia, Panigrahi, B.K., Senroy, N., "Extreme learning machine based fault classification in a series compensated transmission line," *IEEE Conference on Power Transmission Publications held in Dec. 2012*, pp. 1–6, 2012

[22] K. Chen, J. Hu, and J. He, "Detection and Classification of Transmission Line Faults Based on Unsupervised Feature Learning and Convolutional Sparse Autoencoder," *IEEE Transactions on Smart Grids*, vol. 9, no. 3, pp. 1748–1758, May 2018.

[23] M. M. Rahman, M. F. Rabbi, M. K. Islam, and F. M. M. Rahman, "HVDC over HVAC power transmission system: Fault current analysis and effect comparison," in *2014 International Conference on Electrical Engineering and Information Communication Technology*, 2014, pp. 1–6, 2014.

[24] A. Sharma, S. R. Ola, and O. P. Mahela, "Analysis of faults in power system network in the presence of series compensated EHV transmission lines," *IEEE 7th Power India International Conference (PIICON)*in *2016*, pp. 1–6, 2016.

[25] M. A. Yalcin, M. Turan, and Z. Demir, "Effects of transmission line faults on dynamic voltage stability," in *PowerTech Budapest 99. Abstract Records. (Cat. No.99EX376)*, 1999, p. 80.

[26] A. Cochocki and R. Unbehauen, "*Neural Networks for Optimization and Signal Processing*," 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1993.

[27] Haykin S*, "Neural Networks. A comprehensive foundation",*2nd ed. New York,Macmillan Collage Publishing Company, Inc., 1994.

[28] Zurada, Jacek M."*Introduction to artificial neural systems"*. vol. 8. St. Paul: West, 1992.

[29] N. Ampazis and S. J. Perantonis, "Levenberg-Marquardt algorithm with adaptive momentum for the efficient training of feedforward networks," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, 2000, vol. 1, pp. 126–131 vol.1.

[30] Morlet, J.; Arens, G.; Fourgeau, E. and Giard, D., "Wave propagation and sampling theory", Part1: Complex signal land scattering in multilayer media. Journal of Geophysics, 47: 203-221, 1982.

[31] P. M. Bentley and J. T. E. McDonnell, "Wavelet transforms: an introduction," *Electron. Commun. Eng. J.*, vol. 6, no. 4, pp. 175–186, Aug. 1994

[32] M. K. Wali, M. Murugappan, R. B. Ahmad, and Bong Siao Zheng, "Development of Discrete Wavelet Transform (DWT) toolbox for signal processing applications," in

2012 International Conference on Biomedical Engineering (ICoBE), 2012, pp. 211–216.

[33] Y. Wang, Z. Li, C. Wang, L. Feng, and Z. Zhang, "Implementation of discrete wavelet transform," *12th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT*in *2014)*, pp. 1–3, 2014

[34] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1316–1322

[35] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 2017, pp. 721–724.

[36] N. Jmour, S. Zayen, and A. Abdelkrim, "Convolutional neural networks for image classification," in *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, 2018, pp. 397–402.

[37] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015, pp. 2359–2364.

[38] A. H. Namin, K. Leboeuf, H. Wu, and M. Ahmadi, "Artificial neural networks activation function HDL coder," in *2009 IEEE International Conference on Electro/Information Technology*, 2009, pp. 389–392.

[39] Y. Tang, X. Zhu, and M. Cui, "Convolutional neural network with gradient information for image super-resolution," in *2016 IEEE International Conference on Information and Automation (ICIA)*, 2016, pp. 1714–1719.

[40] S. Lee, T. Chen, L. Yu, and C. Lai, "Image Classification Based on the Boost Convolutional Neural Network," *IEEE Access*, vol. 6, pp. 12755–12768, 2018.

[41] S. Razavi and B. A. Tolson, "A New Formulation for Feedforward Neural Networks," *IEEE Trans. Neural Networks*, vol. 22, no. 10, pp. 1588–1598, Oct. 2011.

APPENDIX

A.1 THE DATA FOR IEEE 14-BUS SYSTEM

Table A1.1 Bus Data for IEEE 14-Bus System

Bus No.	Bus Voltage		Generation		Load	
	Magnitude Per Unit	Phase Angle Degrees	Real MW	Reactive MVAR	Real MW	Reactive MVAR
1	1.060	0.0	232.4	-16.9	0.0	0.0
2	1.045	-4.98	40.0	42.4	21.7	12.7
3	1.010	-12.72	0.0	23.4	94.2	19.0
4	1.019	-10.33	0.0	0.0	47.8	3.9
5	1.020	-8.78	0.0	0.0	7.6	1.6
6	1.070	-14.22	0.0	12.2	11.2	7.5
7	1.062	-13.37	0.0	0.0	0.0	0.0
8	1.090	-13.36	0.0	17.4	0.0	0.0
9	1.056	-14.94	0.0	0.0	29.5	16.6
10	1.051	-15.10	0.0	0.0	9.0	5.8
11	1.057	-14.79	0.0	0.0	3.5	1.8
12	1.055	-18.07	0.0	0.0	6.1	1.6
13	1.050	-15.16	0.0	0.0	13.5	5.8
14	1.036	-16.04	0.0	0.0	14.9	5.0

Table A1.2 Regulated Bus Data (P-V Buses)

Bus No.	Voltage Magnitude Per unit	Reactive Power Limits	
		Minimum MVAR	MaximumMVAR
2	1.045	-40.0	50.0
3	1.010	0.0	40.0
6	1.070	-6.0	24.0
8	1.090	-6.0	24.0

Table A1.3: Line Data for IEEE 14-Bus System

Line No.	Between buses	Line impedance		Half Line Charging Susceptance Per unit	Branch MW Limit Per unit
		R per unit	X per unit		
1	1-2	0.01938	0.05917	0.02640	0.6
2	2-3	0.04699	0.19797	0.02190	0.7
3	2-4	0.05811	0.17632	0.01870	0.8
4	1-5	0.05403	0.22304	0.02460	0.5

5	2-5	0.05695	0.17388	0.01700	0.4
6	3-4	0.06701	0.17103	0.01730	0.3
7	4-5	0.01335	0.04211	0.0064	0.2
8	5-6	0.0	0.25202	0.0	0.5
9	4-7	0.0	0.20912	0.0	0.4
10	7-8	0.0	0.17615	0.0	0.2
11	4-9	0.0	0.55618	0.0	0.2
12	7-9	0.0	0.11001	0.0	0.2
13	9-10	0.03181	0.08450	0.0	0.2
14	6-11	0.09498	0.19890	0.0	0.3
15	6-12	0.12291	0.25581	0.0	0.2
16	6-13	0.06615	0.13027	0.0	0.2
17	9-14	0.12711	0.27038	0.0	0.2
18	10-11	0.8205	0.19207	0.0	0.2
19	12-13	0.22092	0.19988	0.0	0.2
20	13-14	0.17093	0.34802	0.0	0.2

Table A1.4: Transformer Data

Transformer	**Between Buses**	**Tap Setting**
1	4-7	0.978
2	4-9	0.969
3	5-6	0.932

Table A1.5: Shunt Capacitor Data

Bus Number	**Susceptance Per Unit**
9	0.190

ancn

ORIGINALITY REPORT

15%	8%	11%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.matlabcodes.com Internet Source	1%
2	Submitted to SASTRA University Student Paper	1%
3	en.wikipedia.org Internet Source	1%
4	dspace.wul.waseda.ac.jp Internet Source	1%
5	baadalsg.inflibnet.ac.in Internet Source	1%
6	S. Mishra, A. Baral. "Classification of power system faults using voltage Concordia pattern feature aided PNN", 2016 IEEE 6th International	<1%