A PROJECT REPORT
on

NEURO FUZZY CLASSIFICATION FOR DATA MINING TASKS

Submitted to
KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE & ENGINEERING

By

SHALINI JAISWAL   1505062

PRIYANKA MAHATO   1505049

UNDER THE GUIDANCE OF
PROF. HIMANSU DAS

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024
April 2019

A PROJECT REPORT
on

NEURO FUZZY CLASSIFICATION FOR DATA MINING TASKS

Submitted to
KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE & ENGINEERING

BY

SHALINI JAISWAL	**1505062**
PRIYANKA MAHATO	**1505049**

UNDER THE GUIDANCE OF
PROF. HIMANSU DAS

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024
April 2019

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, Odisha 751024

CERTIFICATE

This is certify that the project entitled

NEURO FUZZY CLASSIFICATION FOR DATA MINING TASKS

Submitted By

SHALINI JAISWAL	1505062
PRIYANKA MAHATO	1505049

is a record of bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2018-2019, under our guidance.

Date: /   /

Prof.Himansu Das

Project Guide

Acknowledgments

ABSTRACT

Artificial Neural Network is a popularly used Machine Learning Algorithm both in research and industry.But the output of ANN fluctuates in a large range,to overcome this defect fuzzy logic is used to increase the number of feature and calculate the contribution of each feature in each class this helps in increasing consistency and accuracy especially for datasets that are not very large.

One of the shortcomings of Fuzzy Logic is that it is very time consuming so feature reduction techniques such as Principal Component Analysis ,Linear Discriminant Analysis and Independent Component Analysis is used and the results are recorded.

Finally feature selection algorithm called Differential Evolution algorithm is also used and the a comparative study is performed.

Contents

List of Figures:

Chapter 1

Introduction

With the increasing amount of complexity in real life problems it is very important for machine learning algorithms to satisfy large number of requirements in terms of accuracy, precision ,speed and learning ability.Adaptive Hybrid System models that can be fine tuned and customized for very specific requirements are being used in a very large extent in industries ,research and other academic purposes.

The first model we have implemented is the Neural Network which uses feed forward and back propagation to improve efficiency of results .Neural Networks are a classic machine learning technique which attempts to make a computer model of a brain,this feature of neural networks makes it one of the most researched models of machine learning.

The second model implemented by us is the Neuro-Fuzzy System which is a hybrid between Neural Network and fuzzy logic in which the capability of fuzzy systems to adapt to problems in a way humans perceive it and the learning ability of Neural Networks.Another advantage of fuzzy systems is that they can be easily customized/modified according to user's needs.The distinct feature of fuzzy systems is that it can work efficiently even with incomplete or imperfect datasets compared to other mathematical models.Our main focus on our project will be on Neuro-Fuzz systems and the class belonging parameters of data and how the effect the accuracy of predictions.

The next model that we have analysed is the Neural Network Model along with Principal Component Analysis for speeding up the machine learning algorithm.The overall time required to run the algorithm decreases considerably by using PCA. If the diamensions in a dataset is too high then the processing time will also be large,so PCA which is a diamensionality reduction algorithm plays a key role by only considering the attributes which have high impact on the results.

Then the final model we have analysed is the Fuzzy Neural Network model with Principal Component Analysis.Neuro Fuzzy models are known for yielding accurate predictions but its execution time is very high because it has to perform a large number of repeated operations to arrive on the result.Thus PCA with Neuro-Fuzzy Model not only provides accurate results but also reduces the execution time.

After computing the results using the PCA algorithm we have next used the LDA(Linear Discriminant Analysis) algorithm and implemented it with the ANN and NF models.The purpose of LDA is to reduce the dimensions by removing the redundant and dependent features which do not

contribute much to the final result.Then a comparative analysis is performed between the PCA and LDA models.

Finally the ICA(Independent Component Analysis) algorithm is implemented.Independent Component Analysis is a type of dimensionality reduction algorithm that carries out the transformation of a set of variables to a new set of components, it is performed by maximizing the statistical independence between the new components . The independent component analysis (ICA) of a random vector consists of searching for a linear transformation that reduces the statistical dependence between its components. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing mechanism/system is also unknown. The latent variables are considered to be non-gaussian and mutually independent, and are known as the independent components of the observed data.The common applications for which Independent Component Analysis is used are data analysis and compression, Medical signal processing (fMRI, ECG, EEG),Feature Extraction,Bayesian detection, localization of sources, and blind source identification and facial detection. The result of the ICA modes are recorded for each of the datasets and then evaluated on various performance measuring parameters like root mean square(rms) error,execution time,F measure ,precision , recall and confusion matrix .The observations have been shown using tables,graphs and charts.

Motivation

1. Neural nets are widely used in pattern recognition because of their ability to generalise and to respond to unexpected inputs/patterns. During training, neurons are taught to recognise various specific patterns and whether to fire or not when that pattern is received.

2. New patterns of data can be learned easily with the help of neural networks hence, it can be used to preprocess data in fuzzy systems.

3. Neural network, because of its capability to learn new relationship with new input data, can be used to refine fuzzy rules to create fuzzy adaptive system.

4. Feature Reduction and selection techniques enables the machine learning algorithm to train faster and reduces the complexity of a model and makes it easier to interpret. It improves the accuracy of a model if the right subset is chosen.It reduces overfitting

Challenges

1. Neural Network models tend to overfit if you don't have a lot of data to train them on.

2. Neural Network models have millions, or even billions of parameters with orders of magnitude more relationships between said parameters. We therefore cannot know what relationships or parameters led to a conclusion, or why exactly it came to such a conclusion.

3. Fuzzy Neural Network take more execution time as compared to normal Neural Network because of its increase in number of features involved in processing step.

4. PCA does not say anything about the relevance of the features for classifying the data, because it does not even use the class labels to find the principal components.

5. By contrast, feature selection tries to eliminate candidate features that are irrelevant, thereby decreasing the complexity of the model. Using PCA after feature selection is sometimes useful, often useless.

Chapter 2

Literature Survey

The research work performed by L Zadeh's [7] work on fuzzy sets is regarded as one the benchmark works in the field of fuzzy sets .In his paper he has stated about the importance of fuzzy sets in the way humans perceive a problem especially in the field of language communication,pattern recognition and information abstraction.

Article titled "A novel Neuro-fuzzy classification technique for data mining" by S. Ghosh , S. Biswas , D. Sarkar ,P. Sarkar [4] the authors propose a Neuro-Fuzzy Classification technique for data mining in which they fuzzified the inputs based on the bell shaped membership function.In the fuzzified matrix formed the input features were associated with a degree of membership to different classes. Depending on the value of degree of membership the   specific category or class was determined.In their study they also compared the Neuro-Fuzzy Classification Technique with Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-fuzzy Inference System (ANFIS). Then the performance of all the systems were assessed using parameters such as root mean square error,accuracy precision and recall.In their study they concluded that their proposed Neuro - Fuzzy classification technique performed better on a lot of parameters compared to the other classification techniques (RBFNN and ANFIS).

In the study titled "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease" the authors Kemal Polat , Salih Güne [9] have used the principal component analysis (PCA) and adaptive neuro-fuzzy inference system (ANFIS) techniques .They have used feature reduction to reduce the number of input features of the diabetes data set from 8 to 4 and then have conducted the predictive diagnosis by passing the inputs though the ANFIS model.

In the research work performed by S.Kar ,S. das and P.K. ghosh [6] "Applications of neuro fuzzy systems: A brief review and future outline" have listed about the research trends and developments in the field of NFS in domains of profile modelling system,economic system,medical system ,traffic control, image recognition and processing, feature extraction, manufacturing and system designing, forecasting and predictions, NFS enhancements ,social sciences and political analysis.

"Self -Organisation for Object Extraction Using Multilayer Neural Network and Fuzziness Measures" titled research work performed by the authors A. Ghosh ,N.R. Pal and S.K. Pal stated the use of   feed forward multilayer perceptron (MLP) with back propagation of error .In their proposed architecture MLP not require any supervised learning .The different layers of the neurons are

interconnected and the output status of the neurons are described as a fuzzy set and fuzziness measure are used as a measure of error in their system.

KELE AL and AYT RK KELE [11] in their report on "Extracting fuzzy rules for the diagnosis of breast cancer." aimed to study the strong diagnostic fuzzy rules for diagnosis of   of this study was to extract strong diagnostic fuzzy rules for diagnosis of breast cancer.To achieve this they used, a neuro-fuzzy classification tool called NEFCLASS . The learning algorithm they used was a heuristic tool   and it was   efficient at performance diagnosis and classification tasks. The rule base to be used for diagnosis consists of 9 rules using the Breast Imaging Reporting and Data System (BI-RADS), mass shape, and mass margin attributes.

In the research work "Neural-network-based fuzzy logic control and decision system" the authors C.T. Lin and C.S.G. Lee have proposed a neural-network model for fuzzy logic control and decision systems .The model which is combination of of fuzzy logic controller and neural-network structure .this causes a combination of learning abilities into an integrated neural-network-based fuzzy logic control and decision system.In their model they proposed a combination of unsupervised and supervised learning schemes due to which the learning speed is much faster than the original back propagation neural network model.

PierreComon in his paper Independent component analysis, A new concept?[22] discusses about random vector searching for a linear transformation that minimizes the statistical dependence between its components.In this he proposes efficient algorithm which allows the computation of the ICA of a data matrix within a polynomial time. In his paper he also discusses that ICA can be seen as an extention of PCA which can only impose independence up to the second order and, consequently, defines directions that are orthogonal.

Chapter 3

Basic Methodologies

3.1 Membership Function

Membership functions is a curve that defines the feature of a fuzzy set by assigning to each element the corresponding membership values or degree of membership .The value μ(x)=1 represents complete belonging to a class and μ(x)=0 indicates complete absence from the class. The types of membership function we have used Triangular Function,Trapezoidal Function,Gaussian Function.

3.1.1   Triangular Member Function
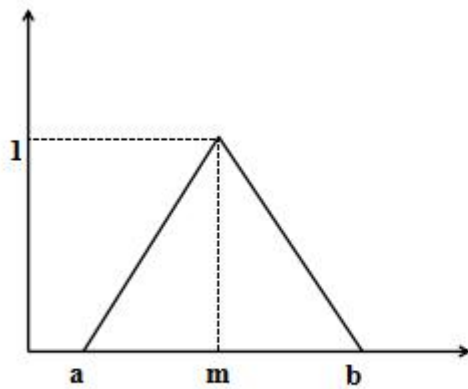
$$\mu_A(x)=\begin{cases} 0 & x\le a \\ \dfrac{x-a}{m-a} & a\le x\le m \\ \dfrac{b-x}{b-m} & m<x\le b \\ 0 & x\ge b \end{cases}$$

Figure 1 : Triangular MF

------(1)

Triangular function is a very commonly used membership function which have lower complexes compare to other membership function while splitting values. In equation 1 and figure 1 a is the lower limit , b is the upper limit and m is the mean of a and b.

3.1.2   Trapezoidal Member Function
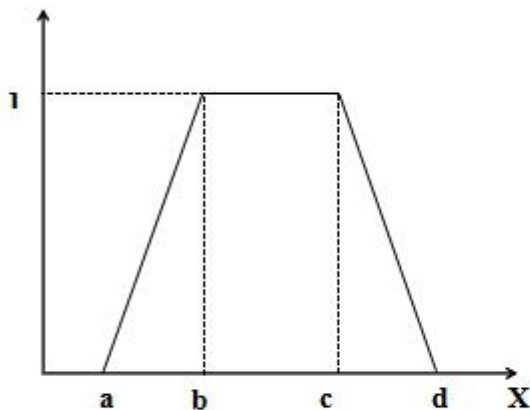
$$\mu_A(x)=\begin{cases} 0 & (x<a)\,\mathrm{or}\,(x>d) \\ \dfrac{x-a}{b-a} & a\le x\le b \\ 1 & b\le x\le c \\ \dfrac{d-x}{d-c} & c\le x\le d \end{cases}$$

Figure 2: Trapezoidal MF

------------(2)

In Trapezoidal membership function where lower limit is a, upper limit d, b is lower support limit and an upper support limit c, where a < b < c < d. Values between b and c have membership degree is 1, whereas the input in a and b increases the membership degree and give's value closer to b while the input between c and d decreases the membership degree and give result closer to d.

3.1.3   ∏ - type Function
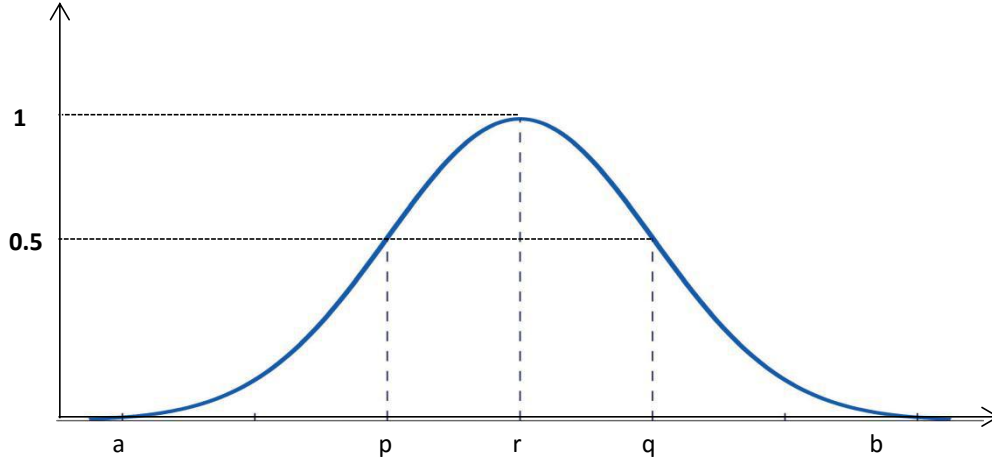
Figure 3: π - type Membership Function

$$\mu\left(F_n;a,r,b\right)=\begin{cases} 0 & F_n \leq a \\ 2^{N-1}\left[(F_n-a)/(r-a)\right]^N & a < F_n \leq p \\ 1-2^{N-1}\left[(r-F_n)/(r-a)\right]^N & p < F_n \leq r \\ 1-2^{N-1}\left[(F_n-r)/(b-r)\right]^N & r < F_n \leq q \\ 2^{N-1}\left[(b-F_n)/(b-r)\right]^N & q < F_n \leq b \\ 0 & F_n \geq b \end{cases}$$

--------------(3)

∏ -type function's smoothness and brief notation makes it more popular as compare to other membership for calculating the fuzzy values. ∏ -type membership function has that quality to modify it's value the of N which is the fuzzifier for membership function . In the above membership function the value of center point is defined as r where r = (p+q)/2 , here p and q are the crossover point in the fuzzified values.

The membership value at r corresponds to 1 and it is 0.5 at points p and q. We have followed the methodology of assigning membership value closer to 1 if the points are closer to r and membership values closer to 0.5 if the point are closer to p and q. Here   p = mean(n) - [max(n)-min(n)]/2   and q = mean(n) + [max(n)-min(n)]/2.

3.2   The Artificial Neural Network Model

An artificial neural network is data processing methodology whose basic building block is a neuron whose name /property /behaviour is adopted from the neurons in the human brain.
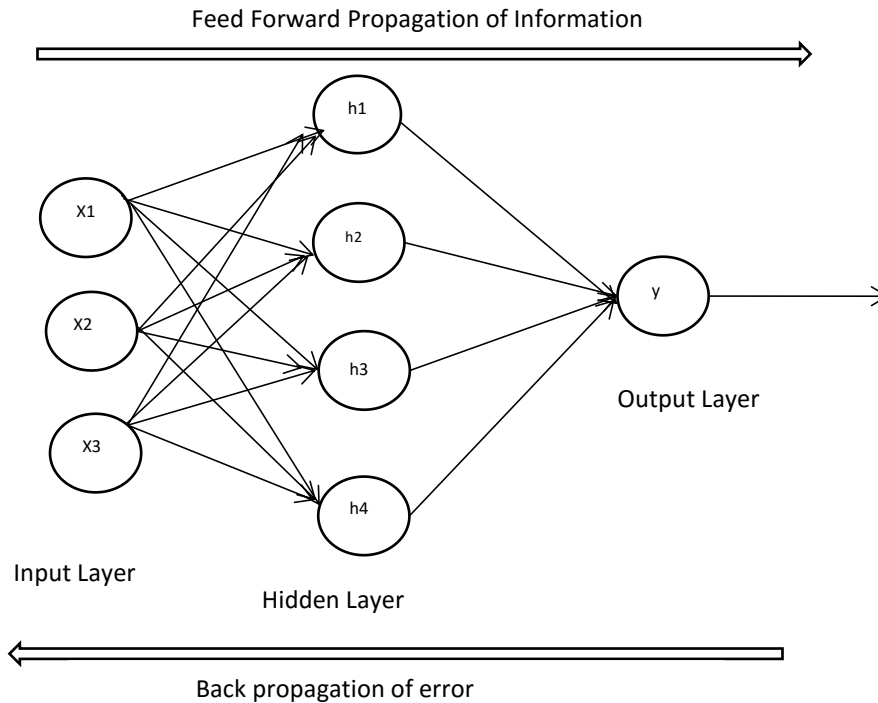
Figure 4: The Artificial Neural Network

Transfer Function :

$$y = f(w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n x_n) = f(\sum_{i=0}^{N} w_i \cdot x_i)$$

--------(4)

Transfer function is the sum of product of input and the weight assigned to each node.

Activation Function :

$$h(x) = \begin{cases} 1 : & \text{if } f(x) = w \cdot x + b > 0 \\ 0 : & \text{otherwise} \end{cases}$$

--------(5)

The processing in the ANN starts with assigning weights to the linkages between the input layer and the hidden layer.The next step would be using the input nodes and linkages to find the activation rates of the hidden layer after which the activation of the output nodes can be found.This entire process is called feed-forward   in Neural Network.Then the error rate at the output node is found and thus the values of the hidden layer nodes ,output nodes and linkages are adjusted to reduce the output error rate.This is called Backpropagation.The process of feed-forward and backpropagation are repeated till the required criteria is met and the error is minimised.

Chapter 4

Proposed Model

4.1 Neuro-Fuzzy Network

With the increasing amount of complexity in real life problems it is very important for machine learning algorithms to satisfy large number of requirements in terms   of accuracy, precision ,speed and learning ability.Adaptive Hybrid System models that can be fine tuned and customized for very specific requirements are being used in a very large extent in industries ,research and other academic purposes.

Neuro-Fuzzy systems are a hybrid between Neural and fuzzy logic in which the capability of fuzzy systems to adapt to problems in a way humans perceive is   utilized. Another advantage of fuzzy systems is that they can be easily customized/modified according to user's needs.The distinct feature of fuzzy systems is that it can work efficiently even with incomplete or imperfect datasets compared to other mathematical models.

The fuzzy systems do not take crisp values as input,they take input in the form of many valued logic in which input can be any real value between 0 and 1.

In the proposed model the output of fuzzy systems is feed through neural network model.Neural Network Model is a powerful Machine Learning model which can be used for both supervised and unsupervised problems.
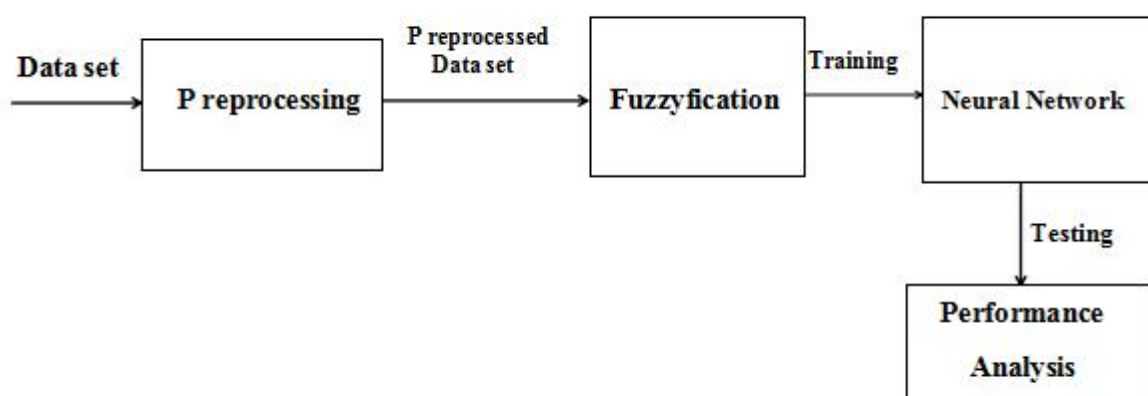
Figure 6:   Fuzzy Neural Network implementation steps

4.1.1   Working

In the proposed neural network model we   have first fuzzified the data values then sent it as input to the neural network.The input features in the dataset are crisp values which are class unrelated.Each

input feature may have   very different ranges of of values which are not scaled hence not suitable for performing prediction operations as the output would be biased.In order to overcome this problem each input feature is mapped to a value between 0-1 using Membership Functions.
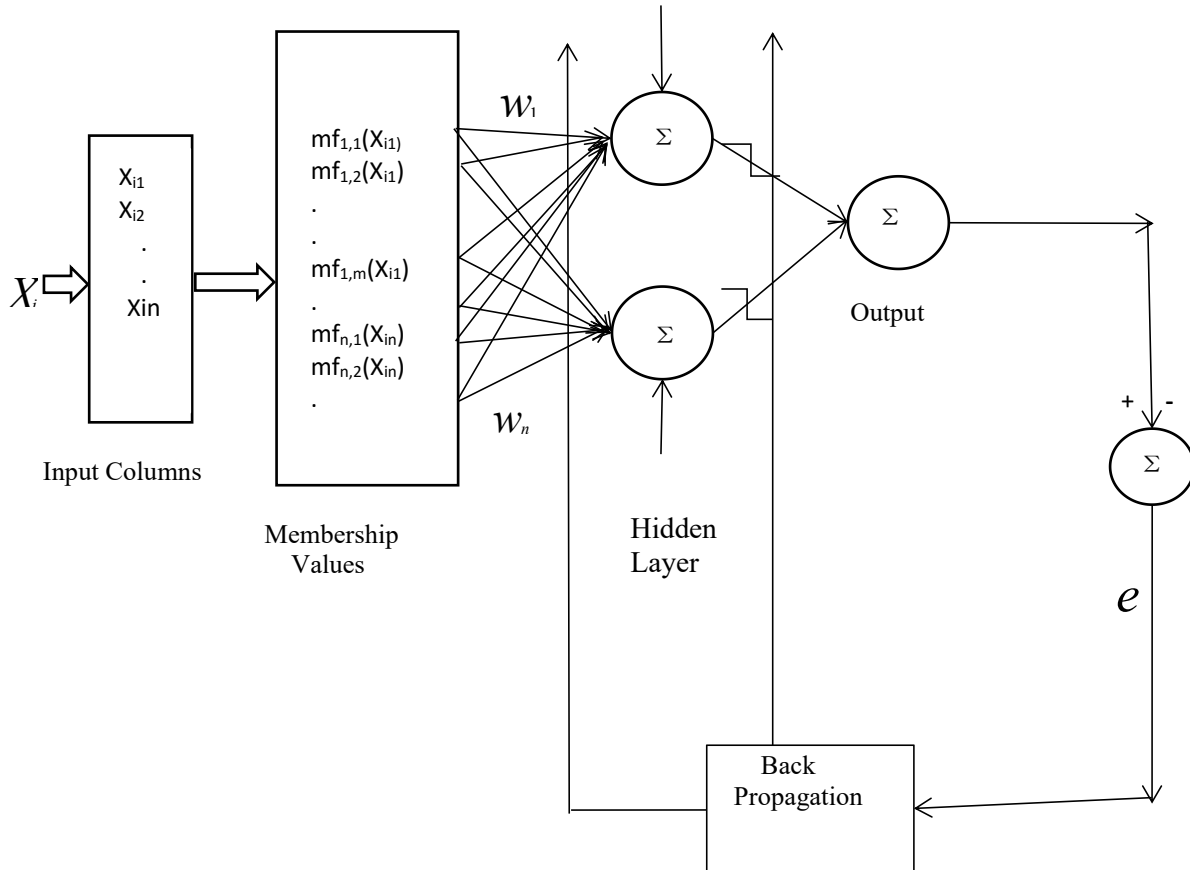
Figure 7: Fuzzy Neural Network Model

4.1.2   Applications of FNN:

The   tasks for which   FNN are   commonly used are:

1)   Automation and Control System Design

2)   Pattern recognition,classification and image processing

3) Modelling and Forecasting: It is used in time series analysis,stock market index prediction,exchange rate prediction

4) FNN are also used in being used for various typical disease diagnoses like brain disorder, cardiac disease, breast cancer,alzheimer, thyroid disorder, leukemia, hypotension, heart disease etc.

4.2   Principal Component Analysis

In today's world large set of data became a problem for every big giant company how to extract the valuable information from huge amount of dataset , here PCA finds the attributes which are strongly co-related.

Principle component analysis is a method of feature extraction which analyses the attributes and drop the least important attributes and uses the most valuable attribute of the dataset without loosing the any information about the dataset . Main motto of PCA is to find the direction with maximum Variance from high dimensional feature space and fit it to low dimension feature space.

Principle Component Analysis (PCA) mainly used for dimensionality reduction in a model which reduces the noise and redundancy and increases the accuracy of the model.

PCA is a tool which is used in exploratory data analysis and to make prediction models. PCA can be implemented by following methods:

- Eigenvalue decomposition of a data covariance matrix
- Singular value decomposition of a data matrix.

4.2..1 Working:

- Standardize the given input features.
- Find the Eigenvectors and Eigenvalues using the covariance matrix.

- The eigenvalues obtained using the covariance matrix is needed to be sorted in descending order and k eigen vectors are to be chosen from them. These eigen vectors should correspond the largest k eigen values where k stands for dimension of new feature subspace (k<d).

- Using the selected k eigenvectors on make the projection matrix X.

- Obtain a k-dimensional feature subspace Q by transforming the original dataset P via X .

4.2.2   ANN with PCA

In this model the Principal Component Analysis technique of feature reduction is used to reduce the number of input features given to the Artificial Neural Networks to include only those input features which have a maximum impact on the predicted output.This method is especially useful is the dimension of input feature space is very large or the number instances is very large.This also reduces the computation time for the prediction of output.
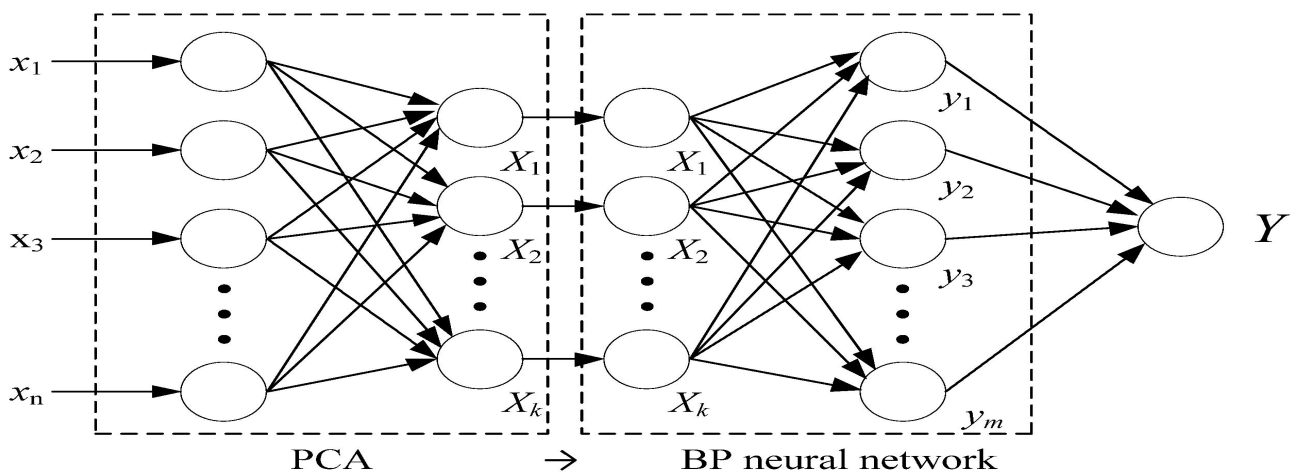
Figure 8 :   Artificial Neural Network With PCA

4.2.3   FNN with PCA

The dimension of input features increases when the inputs are converted into their corresponding membership values.So reduction of dimension of membership values becomes very important for fast and accurate computation of results.The membership values that have low belonging to the output class have not been considered for result computation.
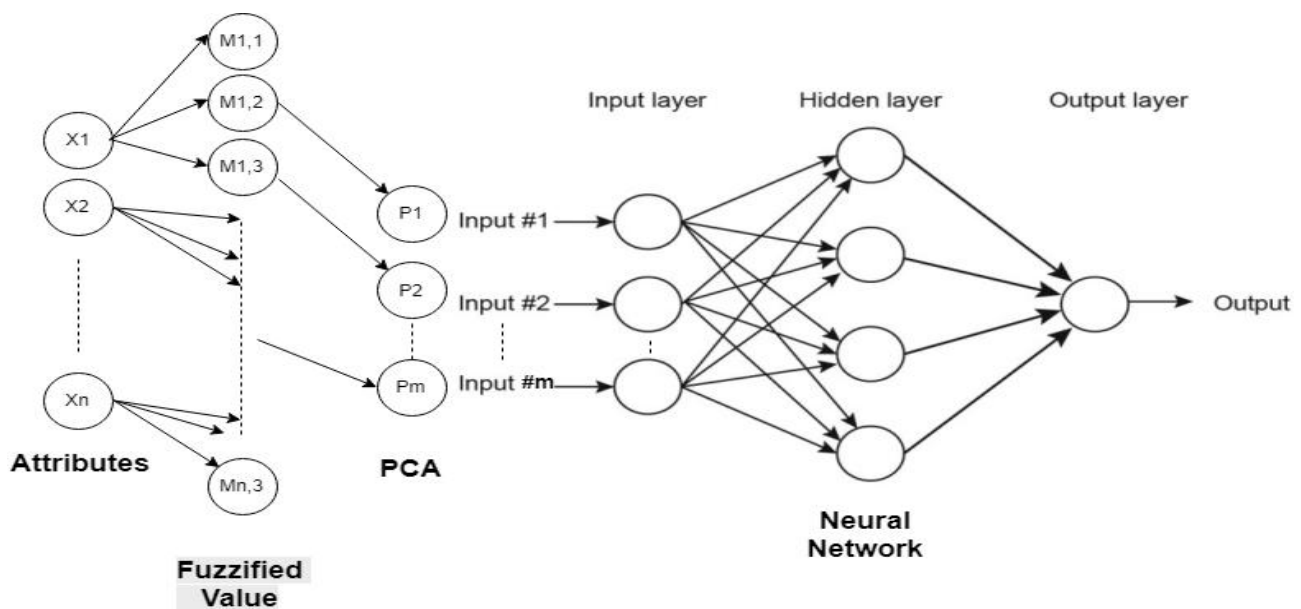

Figure 9: Block Diagram for FNN with PCA

4.3   Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a supervised classification and diamentionality reduction method which was formulated by British mathematician and biostatistian R. A. Fisher in the year 1936.

Linear Discriminant Analysis (LDA) is widely used in dimensionality reduction during pre-processing step for image processing and machine learning applications where the number of attributes are very large in number.

The purpose of dimensionality reduction techniques is to reduce the dimensions by removing the redundant and dependent features which do not contribute much to the final result by transforming the features from higher dimensional space to a space with lower dimensions to reduce processing time, improve accuracy and improve efficiency.

A simple example where LDA can be used is Reducing 2D graph to 1D graph.Linear Discriminant Analysis uses both attributes to create a new axis to project the data into the new axis in a way that maximises the seperation between the two attributes this makes it sutible for classification problems.

For a two attribute problem the method by which LDA creates a new axis is

1. Maximise the distance between the means of the two attributes.

2. Minimise the variation (which LDA calls "scatter" and is represented by s^2)within each category.

   The formula for creating the new axis for two attributes is

$$\frac{(\mu_1 - \mu_2)^2}{s_1^{\,2} + s_2^{\,2}}$$

----------(6)

The main aim is to project a dataset into a lower-dimensional space with good class-separability in order avoid overfitting ("curse of dimensionality") and overlapping among the classes to reduce computational costs.

It is simple, statistically robust method which is commonly used to produce models which have high accuracy and precision as compared to more complex methods.
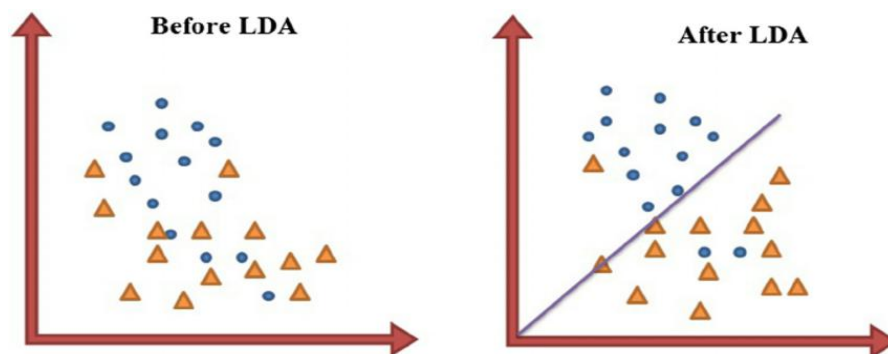

Figure 10 : LDA plain tiring to separate different characteristics most optimally

4.3.1 Computations for Linear Discriminant Analysis

LDA is based upon the concept of searching for a linear combination of variables (predictors) that best separates two classes (targets). To capture the notion of separability, Fisher defined the following score function.

$$Z = \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_d x_d$$

$$S(\beta) = \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \quad \text{Score function}$$

$$\downarrow$$

$$S(\beta) = \frac{\overline{Z}_1 - \overline{Z}_2}{\text{Variance of Z within groups}} \qquad \text{----------(7)}$$

Given the score function, the problem is to estimate the linear coefficients that maximize the score which can be solved by the following equations.

$$\beta = C^{-1}(\mu_1 - \mu_2) \quad \text{Model coefficients}$$

$$C = \frac{1}{n_1 + n_2}(n_1 C_1 + n_2 C_2) \qquad \text{Pooled Covariance Matrix}$$

Where :

β : Linear model coefficients

$C_1 C_2$: Covariance matrices

$\mu_1 \mu_2$: Mean vectors $\qquad\qquad$ -------------(8)

One way of assessing the effectiveness of the discrimination is to calculate the **Mahalanobis distance** between two groups. A distance greater than 3 means that in two averages differ by more than 3 standard deviations. It means that the overlap (probability of misclassification) is quite small.

$$\Delta^2 = \beta(\mu_1 - \mu_2)$$

Δ : Mahalanobis distance between two groups

$\qquad\qquad$ --------------(9)

Finally, a new point is classified by projecting it onto the maximally separating direction and classifying it as *C1* if:

$$\beta^T\left(x - \left(\frac{\mu_1 + \mu_2}{2}\right)\right) > -\log\frac{p(c_1)}{p(c_2)}$$

$\qquad\qquad$ ---------- (10)
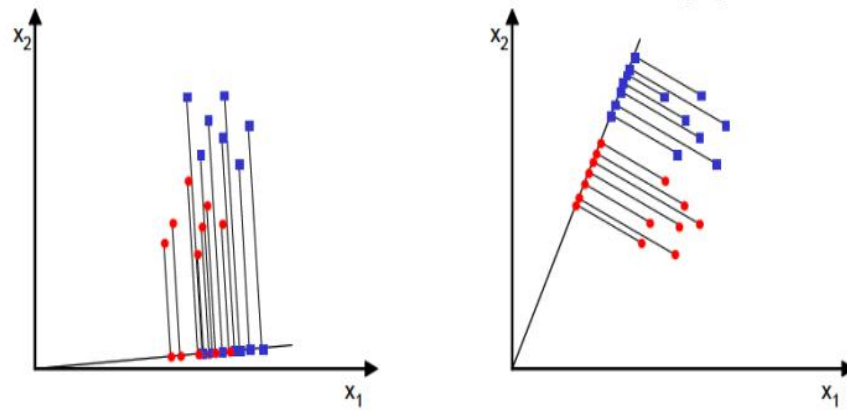
Figure 11:   LDA   find out the best optimal orientation

4.3.2   Application of Linear Discriminant Analysis

Linear Discriminant Analysis have different application in the field of machine learning,data science, image processing and artificial intelligence.

Some of the widely used applications of the fisher linear discriminant are

1. Medical Diagnosis

2. Speech and music classification

3. Face detection

4. Emotion detection

5. Hand Motion Classification

6. Bankruptcy prediction

In bankruptcy prediction based on accounting ratios and other financial variables, linear discriminant analysis was the first statistical method applied to systematically explain which firms entered bankruptcy vs. survived.

7. Face Recognition

In computerized face recognition and image recogintion, each face is represented by a large number of pixel values. Linear discriminant analysis is primarily used here to reduce the number of features to a more manageable number without reducing its details before classification.

8.   Marketing

In the marketing domain, discriminant analysis is used to determine the factors which distinguish different types of customers and/or products services on the basis of surveys or other forms of collected data. This is used to increase productivity and target the right categories.

9. Biomedical studies

The main application of discriminant analysis in medicine is the assessment of severity state of a patient and prognosis of disease outcome.

4.3.3   Limitations of LDA Technique

i) LDA does not work well if the design is not nearly balanced .If the number of objects/attributes in various classes are highly different .

ii) The LDA is sensitive to over-fit and validation of LDA models is at least problematic. However in other methods such as RDA, ANN, SVM etc. The problem is even worse.

iii) LDA is not suitable for non-linear problems.If the distributions are not significantly non gaussian the LDA projections may not preserve the complex structure in the data needed for classification .Example separation of orange- banana shape point clouds.

4.3.4   Comparative Study Between PCA and LDA

There are two broad approaches of dimensionality reduction which are supervised and unsupervised techniques.

Unsupervised techniques comprise of methods such as Principal Component Analysis(PCA), Independent Component Analysis(ICA) and Non-negative Matrix Factorization (NMF) in which the classes of data are not needed to be defined.

Whereas, in supervised dimentionality reduction techniques such as Linear Discriminant Analysis(LDA),Mixture Discriminant Analysis(MDA) class labels have to taken under consideration.

PCA:
component axes that maximize the variance
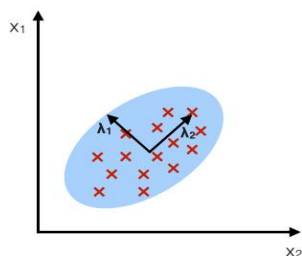
LDA:
maximizing the component axes for class-separation
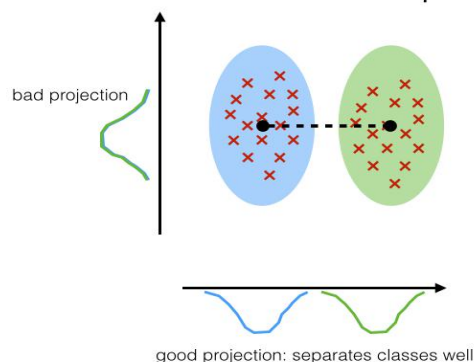
Figure 12 : Comparison between PCA and LDA

4.4   Independent Component Analysis:

Independent component analysis [1] has been established as a fundamental way of analyzing multivariate data. It learns a linear decomposition (transform) of the data, such as the more classical methods of factor analysis and principal component analysis (PCA). However, ICA is able to find the underlying components and sources mixed in the observed data in many cases where the classical methods fail.

ICA attempts to find the original components or sources by some simple assumptions of their statistical properties. Not unlike in other methods, the underlying processes are assumed to be independent of each other, which is realistic if they correspond to distinct physical processes. However, what distinguishes ICA from PCA[2] and factor analysis is that it uses the non-Gaussian structure of the data, which is crucial for recovering the underlying components that created the data.

ICA is an unsupervised method in the sense that it takes the input data in the form of a single data matrix. It is not necessary to know the desired 'output' of the system, or to divide the measurements into different conditions. This is in strong contrast to classical scientific methods based on some experimentally manipulated variables, as formalized in regression or classification methods. ICA is thus an exploratory, or data-driven method: we can simply measure some system or phenomenon without designing different experimental conditions. ICA can be used to investigate the structure of the data when suitable hypotheses are not available, or they are

ICA has many algorithms such as *FastICA*[3], *projection pursuit*[4], and *Infomax*[5]. The main goal of these algorithms is to extract independent components by (1) maximizing the non-Gaussianity, (2) minimizing the mutual information, or (3) using maximum likelihood (ML) estimation method [20]. However, ICA suffers from a number of problems such as over-complete ICA and under-complete ICA.
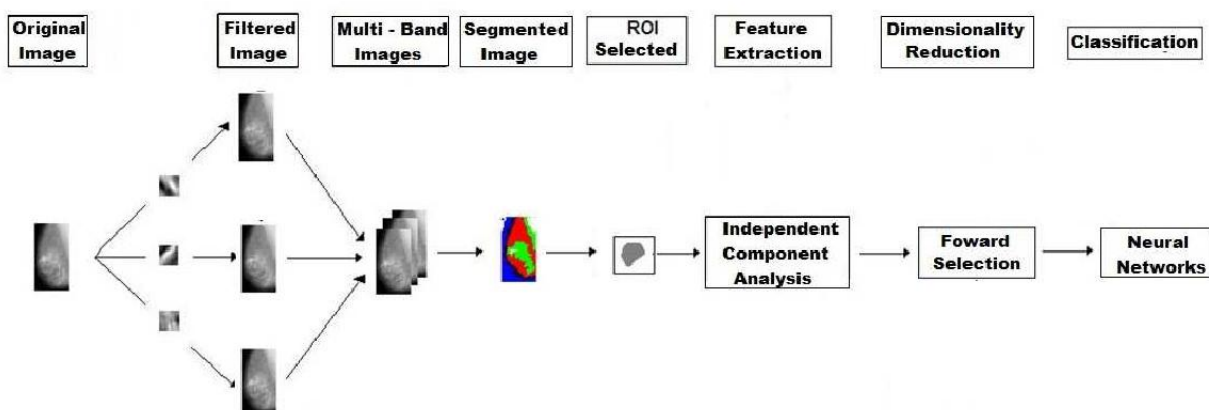

Figure 13 : Working of ICA

4.4.1 Independent Component Analysis Algorithm

Fast ICA

Summarizing the objective functions according to [2] , we see a common goal of maximizing a function $\sum_i E\{G(y_i)\}$, where $y_i = w_i^T x$ is a component of $y = Wx$

$$\sum_i E\{G(y_i)\} = \sum_i E\{G(w_i^T x)\}$$

----------(11)

where w_i^T is the ith row vector in matrix W. We first consider one particular component (with the subscript i dropped). This is an optimization problem which can be solved by Lagrange multiplier method with the objective function

$$O(w) = E\{G(w^T x)\} - \beta(w^T x - 1)/2$$

------------(12)

The second term is the constraint representing the fact that the rows and columns of the orthogonal matrix W are normalized, i.e., $w^T w = 1$. We set the derivative of $O(w)$ with respect to w to zero and get

$$F(w) \triangleq \frac{\partial O(w)}{\partial w} = E\{xg(w^T x)\} - \beta w = 0$$

-----------(13)

where $g(z) = dG(z)/dz$ is the derivative of function $G(z)$. This algebraic equation system can be solved iteratively by Newton-Raphson method:

$$w \Leftarrow w - J_F^{-1}(w)F(w)$$

-----------(14)

where $J_F(w)$ is the Jacobian of function $F(w)$:

$$J_F(w) = \frac{\partial F}{\partial w} = E\{xx^T g'(w^T x)\} - \beta I$$

-----------(15)

The first term on the right can be approximated as

$$E\{xx^T g'(w^T x)\} \approx E\{xx^T\}E\{g'(w^T x)\} = E\{g'(w^T x)\}I$$

----------(16)

and the Jacobian becomes diagonal

$$J_F(w) = [E\{g'(w^T x)\} - \beta]I$$

------------(17)

and the Newton-Raphson iteration becomes:

$$w \Leftarrow w - \frac{1}{E\{g'(w^T x)\} - \beta}\left[E\{xg(w^T x)\} - \beta w\right]$$

-----------(18)

Multiplying both sides by the scaler $\beta - E\{g'(w^T x)\}$, we get

$$w \Leftarrow E\{xg(w^T x)\} - E\{g'(w^T x)\}w$$

-----------(20)

IF we still use the same representation w for the left-hand side, while its value is actually multiplied by a scaler. This is taken care of by renormalization, as shown in the following FastICA algorithm: Note that we still use the same representation w for the left-hand side, while its value is actually multiplied by a scaler. This is taken care of by renormalization, as shown in the following FastICA algorithm:

1. Choose an initial random guess for w
2. Iterate
3. $w \Leftarrow E\{xg(w^T x)\} - E\{g'(w^T x)w\}$
4. Normalize
5. $w \Leftarrow w/\|w\|$
6. If not converged, go back to step 2.

4.4.2   Independent Component Analysis Real World Application and Application

Measurements cannot be isolated from a noise which has a great impact on measured signals. For example, the recorded sound of a person in a street has sounds of footsteps, pedestrians, etc. Hence, it is difficult to record a clean measurement; this is due to (1) source signals always are corrupted with a noise, and (2) the other independent signals (e.g. car sounds) which are generated from different sources . Thus, the measurements can be defined as a combination of many independent sources. The topic of separating these mixed signals is called blind source separation (BSS)[6].The term blind indicates that the source signals can be separated even if little information is known about the source signals[7].

One of the most widely-used examples of BSS is to separate voice signals of people speaking at the same time, this is called cocktail party problem[8]. The independent component analysis (ICA) technique is one of the most well-known algorithms which are used for solving this problem [9]. The goal of this problem is to detect or extract the sound with a single object even though different sounds in the environment are superimposed on one another [31]. Fig. 1 shows an example of the cocktail party problem. In this example, two voice signals are recorded from two different individuals, i.e., two independent source signals. Moreover, two sensors, i.e., microphones, are used for recording two signals, and the outputs from these sensors are two mixtures. The goal is to extract original signals from mixtures of signals. This problem can be solved using independent component analysis (ICA) technique .
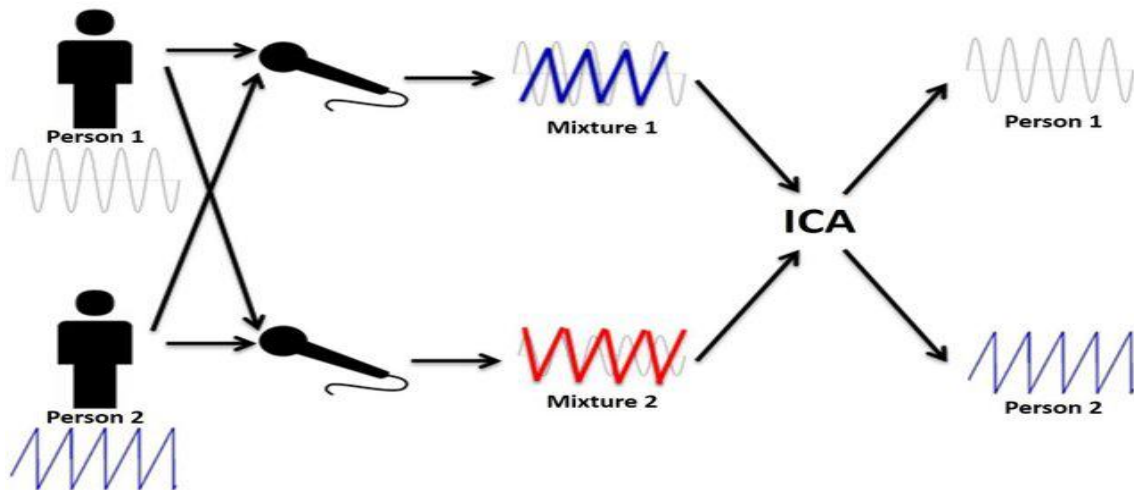
Figure 14 :Cocktail Party Problem

4.4.3   Mathematical and Theoretical Foundations of ICA

A random observerved vector $X = [X_1, X_2, ...X_m]^T$ whose m elements are mixtures of m independent elements of a random vector $S = [S_1, S_2, ...S_m]^T$ given by **X=AS**

In this section we briefly discuss about the basic principles of ICA, such as finding the un-mixing matrix which is an inverse of mixing matrix. And denote 5 assumptions about ICA.

1) The the sources are statistically independent.

2) The mixing matrix is square and full rank.

3) There should be no external noise

4) Data is zero mean

5) source signals must not have a Gaussian probability density function. (At least one of them.)

Statistical independence :

Let $x_1, x_2, ..., x_m$ be in random variables with function $f(x_1, x_2, ..., x_m)$ then x_i are mutually independent if

$$f(x_1, x_2, ..., x_m) = f_1(x_1) f_2(x_2).........f_m(x_m)$$

--------------(21)

When we have two random variables x_i and x_j we can define uncorrelatedness between those two variables by the equation below.

$$E[x_i x_j] - E[x_i]E[x_j] = 0 \quad \text{for} \quad i \neq j \quad \text{where} \quad E[.] \text{ is a exception}$$

On the other hand we can define the statistical independence as equation seen below.

$$E[g(x_i)g(x_j)] - E[g(x_i)]E[g(x_j)] = 0 \quad \text{for} \ i \neq j$$

And in a specific case when the joint pdf is gaussian uncorrelatedness is equivalent to independence. And the authors introduce two methods to measure independence which are minimizing mutual information, or maximization of non-Gaussianity. (The are the same solution.)

Minimization of mutual information :

When we have two variable X and Y, mutual information can be seen as the reduction of uncertainty regarding variable X after the observation of Y. Therefore by having an algorithm that seeks to minimize mutual information, we are searching for components (latent variables) that are maximally independent. (InfoMax is the name of the algorithm).

Maximization of Non-Gaussianity :

When we have two variables X and Y, we can achieve independence by forcing each of them to be as far from the normal distribution as possible. And to do this we measure the non-gaussianity by using negentropy. (which is a positive measure of gaussianity.). And we calculate the approximated negentropy rather than direct calculation.

4.4.4   Comparative study PCA,LDA and ICA

The concept of ICA can be seen as an extension of the principal component analysis (PCA)[23], which can only impose independence up to the second order and defines directions that are orthogonal. This is similar to Principle Component Analysis (PCA), which maps a collection of variables to statistically uncorrelated components, except that ICA goes a step further by removing correlation and higher order independence.

PCA does low rank matrix factorization for compression.ICA performs full rank matrix factorization to remove dependency between the rows ,here number of features remain the same.

Linear Discriminant Analysis is a supervised learning algorithm which class labels have to be taken under consideration. It is a way to reduce 'dimensionality' while at the same time preserving as much of the class discrimination information as possible. LDA[24] helps you find the boundaries around clusters of classes. It projects your data points on a line so that your clusters are as separated as possible, with each cluster having a relative (close) distance to a centroid.

4.5   Differential Evolution Algorithm

In evolutionary computation, **differential evolution** (**DE**) is a method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. Such methods are commonly known as metaheuristics as they make few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. However, metaheuristics such as DE do not guarantee an optimal solution is ever found.

DE is used for multidimensional real-valued functions but does not use the gradient of the problem being optimized, which means DE does not require the optimization problem to be differentiable, as is required by classic optimization methods such as gradient descentand quasi-newton methods. DE can therefore also be used on optimization problems that are not even continuous, are noisy, change over time, etc

DE optimizes a problem by maintaining a population of candidate solutions and creating new candidate solutions by combining existing ones according to its simple formulae, and then keeping whichever candidate solution has the best score or fitness on the optimization problem at hand. In this way the optimization problem is treated as a black box that merely provides a measure of quality given a candidate solution and the gradient is therefore not needed.
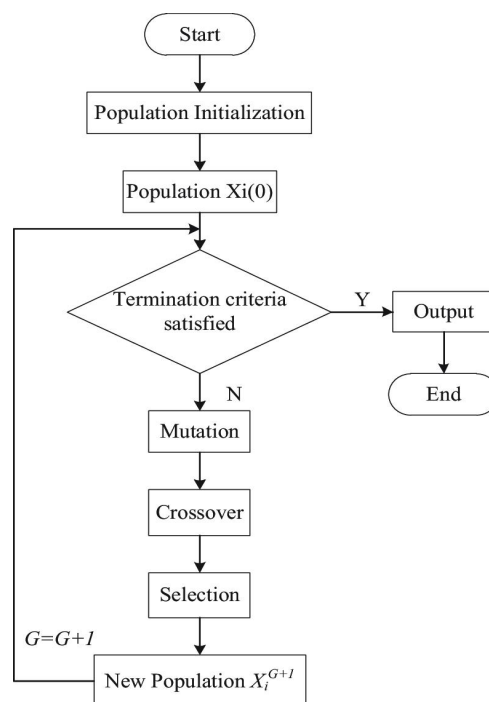
Figure : Flowchart of Differential Evolution Algorithm
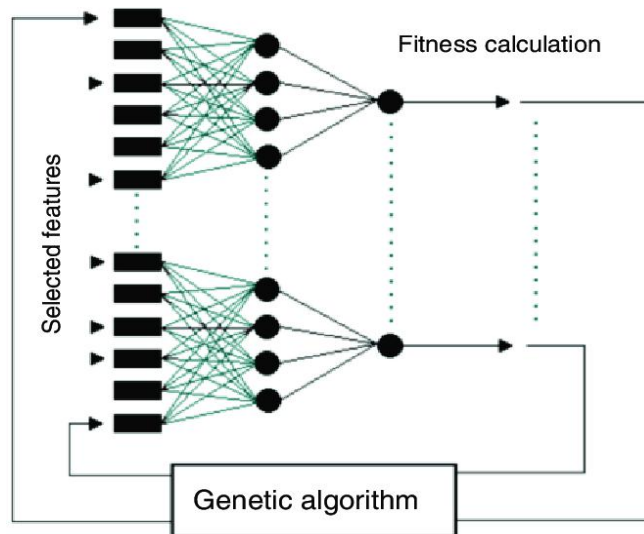
4.5.1 Working

Figure : Feature Selection With Optimization Algorithm

Neural networks have many advantages, but they require very large numbers of parameters to be estimated. When you have a hidden layer (or more than one hidden layer), and multiple input variables, the number of parameters (link weights) that need to be accurately estimated explodes. But every parameter to be estimated consumes some of your informational budget. You are essentially guaranteed to overfit this model.

A DEA optimises a fitness function and determines the best subset of features. Feature selection-based method involving DEA which addressed the classification system's three stages: feature extraction, feature selection and classification.

Performing all the steps of the above flow chart, at the end of the latest generation, your algorithm will have found the population containing a NN with the optimum hyper-parameters. It will be the fittest NN of all in the population.

Chapter 5

Result Analysis

5.1 Performance Measures

5.1.1 Root Mean-Square Error (RMSE)

RMSE is used to measure the performance of the classifier. It is the difference between the predicted values of the classifier and the actual value discovered by the classifier.

$$\text{RMSE} = \sqrt[2]{\frac{\sum_{i=1}^{n} E_{Actual} - E_{Predicted}}{n}}$$

--------------(22)

where E_{Actual} is the actual error and the $E_{Predicted}$ is the predicted error of the model for i = 1, 2, ..., n.

5.1.2 Confusion Matrix:

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

- **true positives (TP):** These are cases in which positive values are predicted as positive

- **true negatives (TN):** These are the cases in which actual negative values are predicted as negative

- **false positives (FP):** These are the cases in which false values are predicted as true

- **false negatives (FN):** These are the cases in which true values are predicted as false

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

Figure 15: Configuration Matrix

Precision : Precision is the ratio between the number of True Positive and the total number of sample our model recognize.

$$\text{Precision} = TP/(TP+FP)$$

Recall: Recall is the ratio between the true positive and the total number of actual positive.

$$\text{Recall} = TP/(TP+FN)$$

F-Measure : F-measure gives the relationship between Precision and recall .F-Measure is computed as harmonic between precision and recall.

$$\text{F-Measure} = 2*Precision*Recall/(Precision+Recall)$$

The result Analysis has been been performed in 3 stages

PCA: In the first stage the comparative study has been performed between ANN,FNN(NF),ANN-PCA,NF-PCA models.

LDA: In the second stage comparative study has been performed between ANN,ANN-PCA,ANN-LDA,NF,NF-PCA,NF-LDA.

ICA: In the third stage the results for 14 datasets using ANN,ANN-LDA,ANN-ICA ,NF,NF-LDA,NF-ICA has been recorded.

5.2 Result Analysis for PCA

Name	Total rows	Total Columns	No. of classes	No. of rows in 1st class	No. of rows in 2nd class	No. of rows in 3rd class
Iris	150	4	3	50	50	50
Mammographic Mass	830	5	2	403	427	-
Breast Cancer	699(683)	9	2	444	239	-
Pima Indian Diabetes	768	8	2	268	500	-
Hayes-Roth	132	6	3	51	51	30
Thyroid	215	6	3	150	35	30
Titanic	1310(1043)	14	2	425	618	-
Wine	178	14	3	59	71	48
Haberman	306	4	2	225	81	-
Blood Tansfusion Service Center	748	5	2	178	570	-

Table 1: Information about dataset
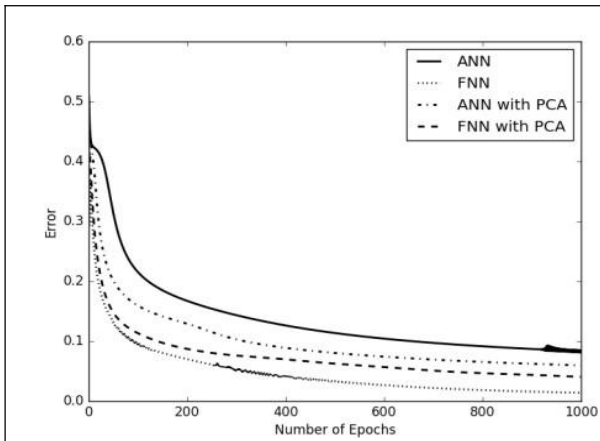
Iris
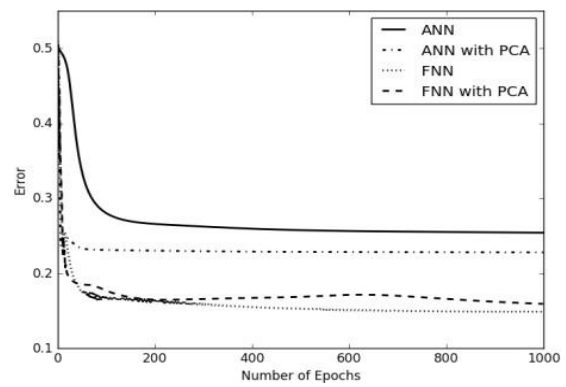
Mammographic
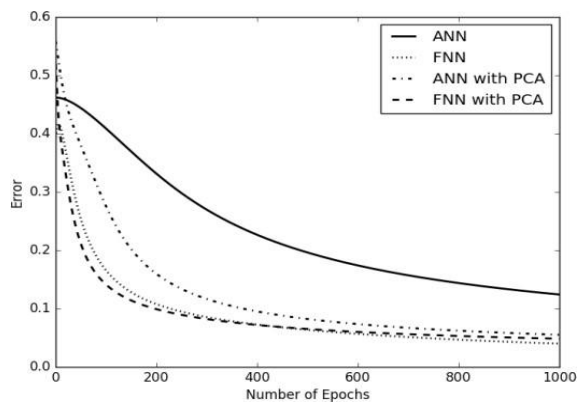
Breast Cancer
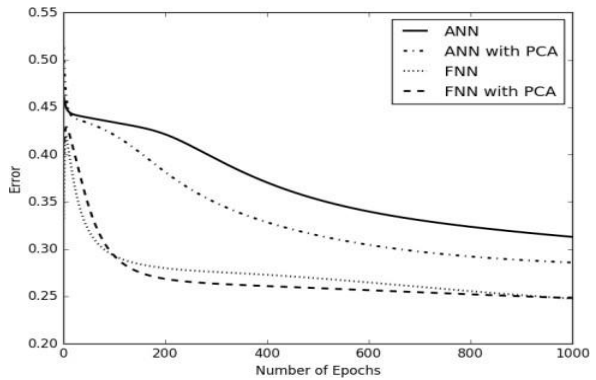
Pima Indian

Hayes-Roth

Thyroid

Titanic

Wine

Haberman

Blood Tansfusion Service Center
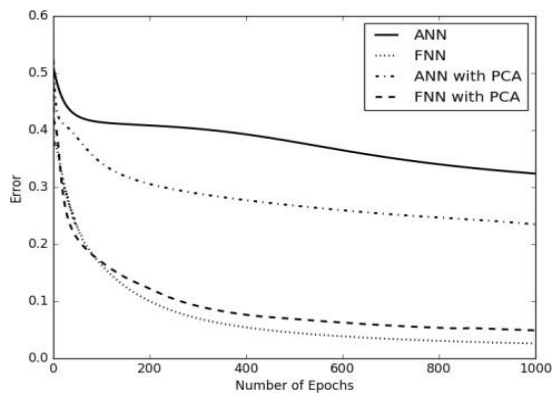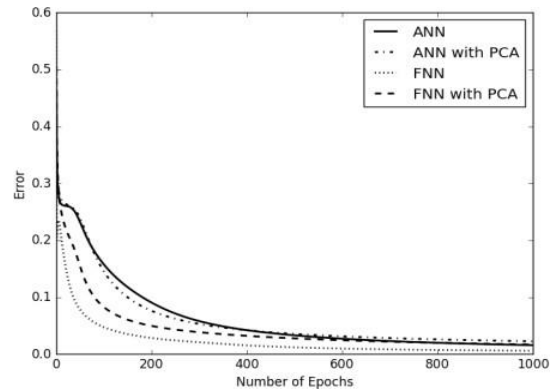
Figure 15. Error plots of all the data sets

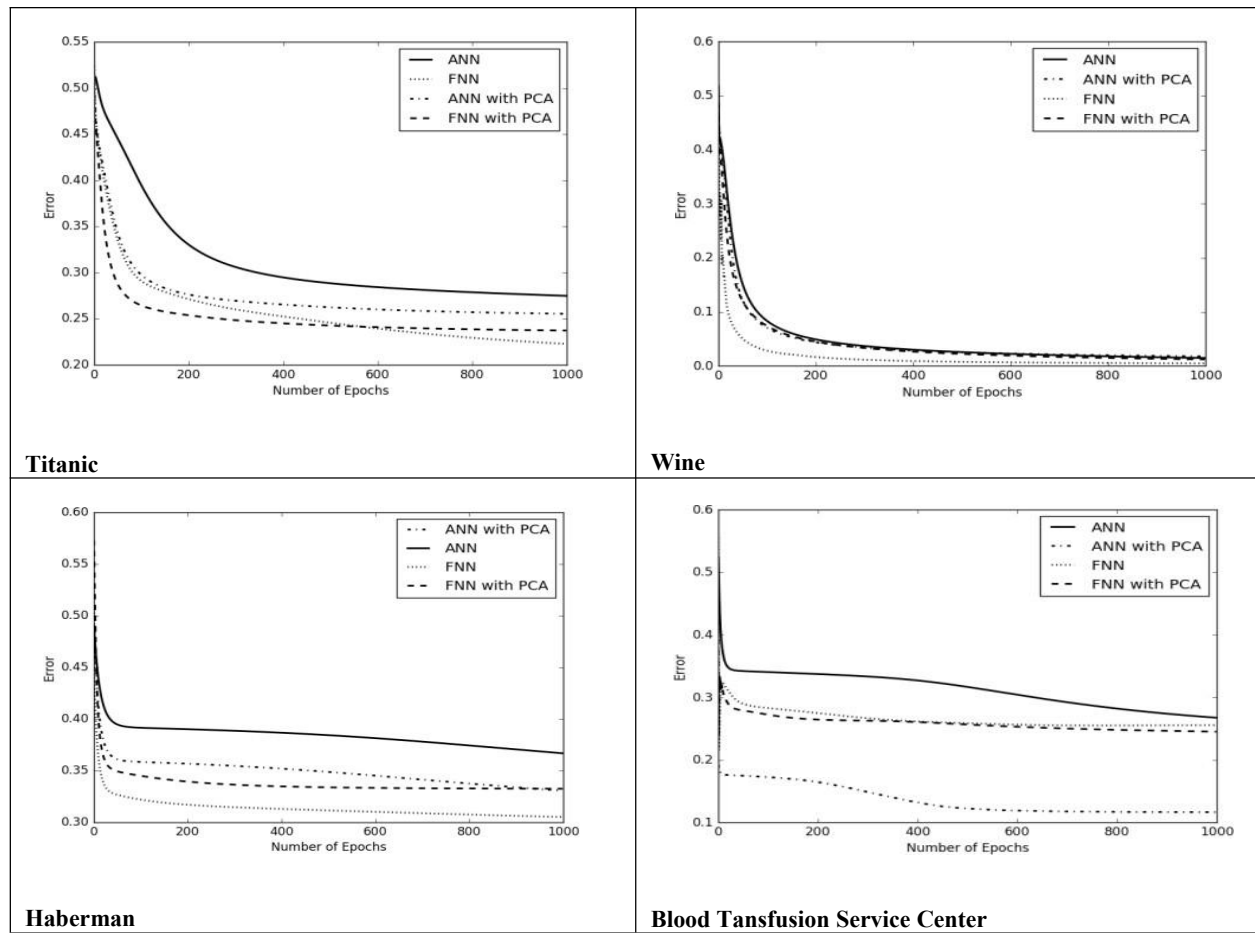| Datasets / Models	ANN			FNN			ANN-PCA			FNN-PCA		
	Worst	Avg	Best	Worst	Avg	Best	Worst	Avg	Best	Worst	Avg	Best
Iris	85.71	91.56	94.44	90.00	92.63	97.06	90.63	94.02	98.02	91.43	95.15	99.26
Mammographic Mass	73.91	79.21	84.62	79.99	84.19	87.35	82.53	84.50	86.36	83.05	85.31	87.72
Breast Cancer Wisconsin	85.42	91.07	95.00	91.18	93.98	96.48	93.66	95.39	97.74	93.86	95.04	97.04
Pima Indian Diabetes	65.52	72.96	78.41	71.76	78.78	86.84	76.00	81.66	86.49	76.74	82.55	85.94
Hayes-Roth	56.25	66.34	76.00	60.00	73.85	83.33	69.23	75.20	83.87	70.59	77.15	86.95
Thyroid	83.72	86.76	92.85	86.04	92.01	96.07	88.46	92.53	97.29	92.98	95.60	99.55
Titanic	75.13	77.64	81.87	76.00	79.25	83.74	77.00	80.54	85.00	77.10	81.47	85.94
Wine	85.29	91.24	94.12	88.88	93.1	97.22	90.90	93.94	98.07	90.62	95.59	99.88
Haberman	70.21	75.42	78.79	75.00	78.05	82.81	76.60	80.29	83.93	77.97	81.02	86.89
Blood Tansfusion Service Center	73.37	75.86	80.00	73.61	76.79	80.41	76.00	80.14	84.50	76.58	80.22	84.91

Table 2. Comparison of classification accuracy of ANN, FNN, ANN-PCA and FNN-PCA

Datasets / Models	Precision, Recall, F-Measure of Classifiers											
	ANN			FNN			ANN-PCA			FNN-PCA		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Iris	0.918	0.910	0.911	0.926	0.928	0.923	0.947	0.942	0.940	0.950	0.953	0.948
Mammographic	0.733	0.523	0.610	0.846	0.545	0.663	0.808	0.491	0.610	0.852	0.532	0.653
Breast Cancer	0.972	0.359	0.501	0.914	0.343	0.499	0.942	0.327	0.485	0.925	0.324	0.480
Pima Indian	0.393	0.181	0.246	0.903	0.750	0.819	0.633	0.263	0.368	0.936	0.723	0.815
Hayes-Roth	0.683	0.706	0.673	0.781	0.754	0.751	0.750	0.751	0.734	0.801	0.758	0.756
Thyroid	0.953	0.751	0.803	0.915	0.868	0.875	0.933	0.831	0.863	0.932	0.944	0.933
Titanic	0.762	0.389	0.512	0.781	0.681	0.728	0.840	0.324	0.465	0.816	0.686	0.740
Wine	0.9155	0.92899	0.9186	0.9361	0.94199	0.9359	0.8512	0.9437	0.9363	0.9629	0.9544	0.9569
Haberman	0.6233	0.08007	0.1417	0.635	0.06999	0.1245	0.7014	0.074	0.132	0.5233	0.0308	0.0576
Blood Tansfusion Service Center	0.5015	0.08399	0.0574	0.7283	0.99923	0.8703	0.129	0.1015	0.1011	0.8105	0.9741	0.8843

Table 3. Comparison of different performance parameters (Precision, Recall and F-Measure) of

ANN, FNN,

5.3 Result Analysis LDA

Dataset Name	Number of Samples	Number of Features	No. of Classes
Pima Indian diabetes	768	8	2
Breast Cancer Wisconsin (Diagnosis)	569	30	2
Breast-Cancer-Wisconsin	699	9	2
Mammographic Mass	961	5	2
Thyroid	215	5	3
Blood Tansfusion Service Center	748	4	2
Heart-statlog	270	13	2
Lung-cancer	32	56	3
SPECTF heart	267	44	2
Haberman	306	3	2
Liver	345	6	2
Hepatitis	155	19	2
Titanic	1310	14	2
Wine	178	14	3
Iris	150	4	3
Hayes Roth	132	6	3

Table 4 :Information about the datasets

| Dataset	RMSE of ANN	RMSE of ANN-PCA	RMSE of ANN-LDA	RMSE of NF	RMSE of NF-PCA	RMSE of NF-LDA
Pima Indian diabetes	0.41819	0.28248	0.25462	0.29887	0.30353	0.305590
Breast Cancer Wisconsin (Diagnosis)	0.0809876	0.01082	0.05601	0.0451097	0.0669	0.067951
Breast-Cancer-Wisconsin	0.41819	0.21637	0.07655	0.17052	0.02546	0.06669
Mammographic Mass	0.30163	0.26163	0.24898	0.1914	0.18031	0.23288
Thyroid	0.19373	0.07685	0.04365	0.28599	0.34237	0.02402
Blood Transfusion Service Center	0.31176	0.31284	0.170960628	0.13597	0.29339	0.30307
Heart-statlog	0.17989	0.22469	0.20753	0.2253	0.1985	0.240866
Lung-cancer	0.03312	0.5406	0.1671	0.04837	0.08123	0.2181
SPECTF heart	0.291	0.1211	0.199	0.129	0.169	0.059
Haberman	0.25329	0.35470	0.3358	0.33007	0.35749	0.33005
Liver	0.33576	0.4430	0.4039	0.41518	0.40686	0.35402
Hepatitis	0.03701	0.060	0.1860	0.172	0.178	0.2100
Titanic	0.19373	0.07685	0.270	0.28599	0.34237	0.2385
Wine	0.05249	0.35470	0.04920	0.03854	0.10774	0.082694
Iris	0.18721	0.38248	0.08524	0.10258	0.1170	0.10645
Hayes Roth	0.32242	0.35525	0.2824	0.14770	0.1622	0.1369

Table 5 : Comparison of RMS Error

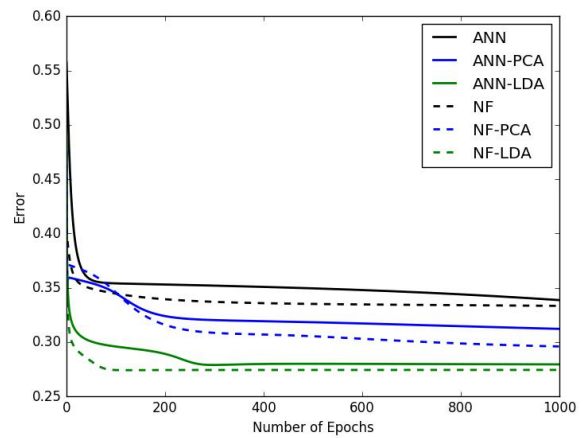Figure 16:   Error plots of all the data sets
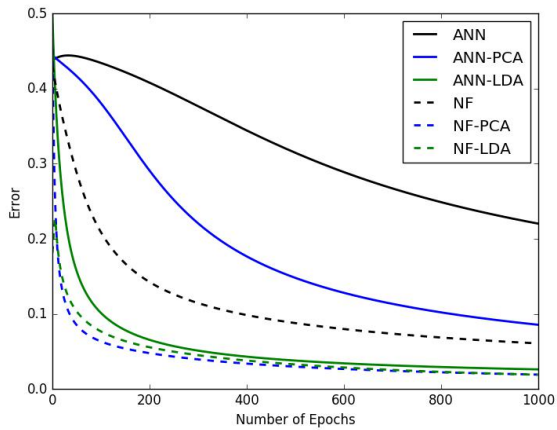
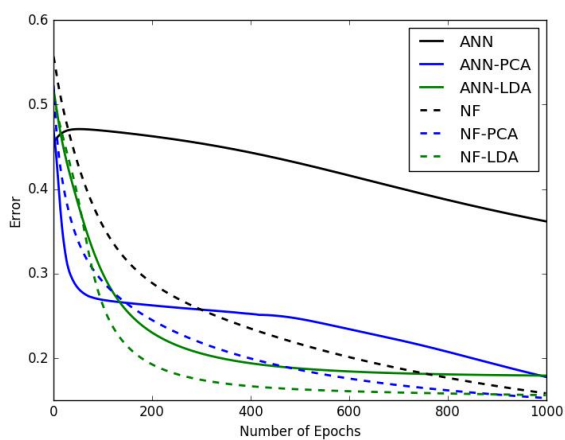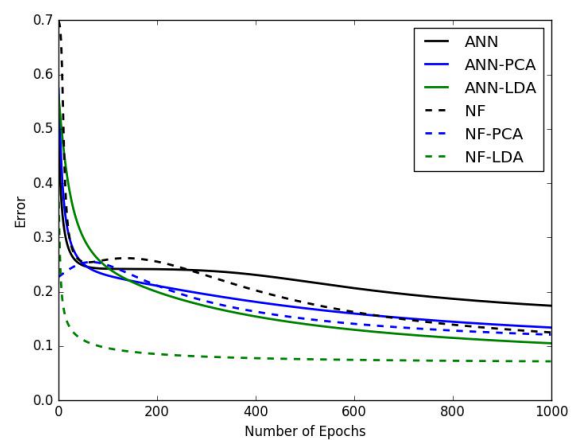Breast Cancer Wisconsin (Diagnosis)
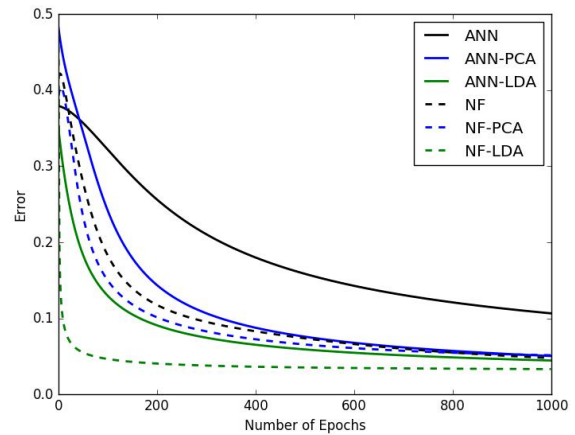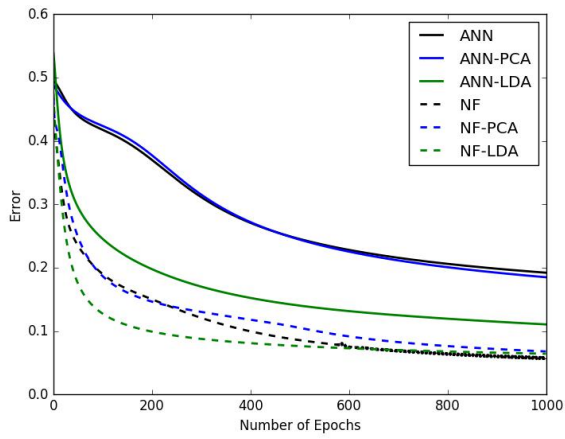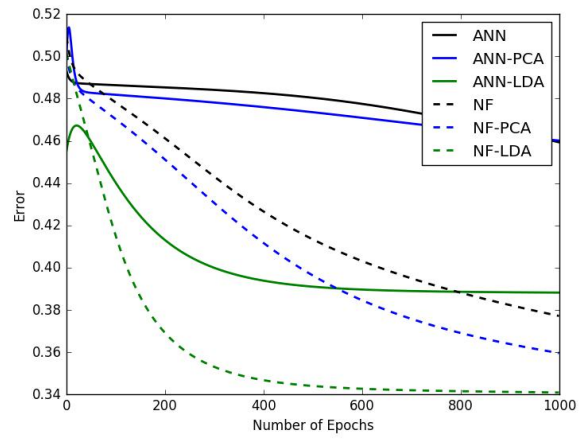

Breast-Cancer-Wisconsin
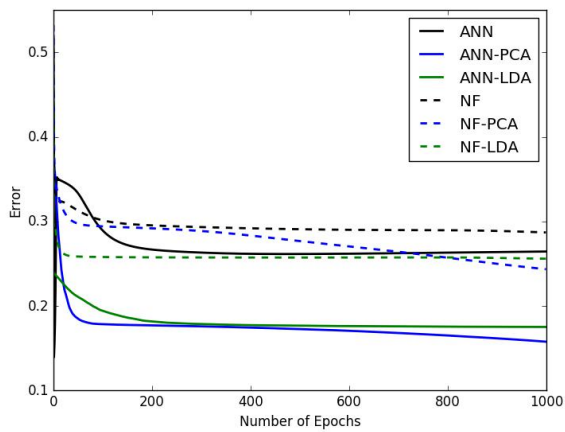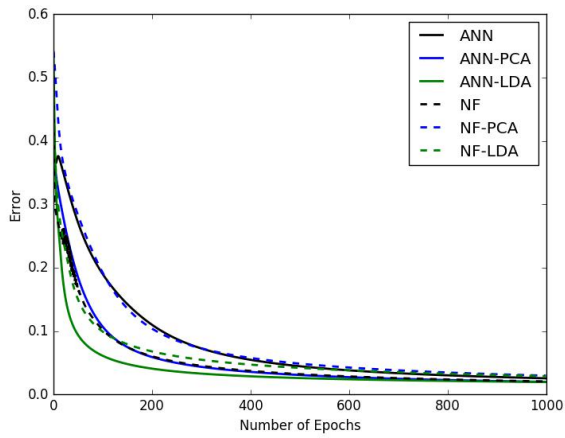

Blood Transfusion Service Center
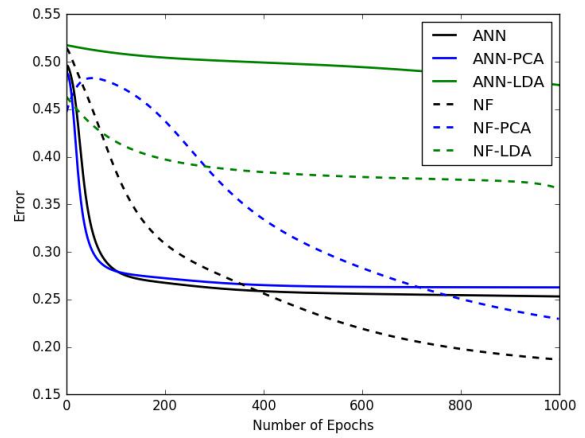

Haberman

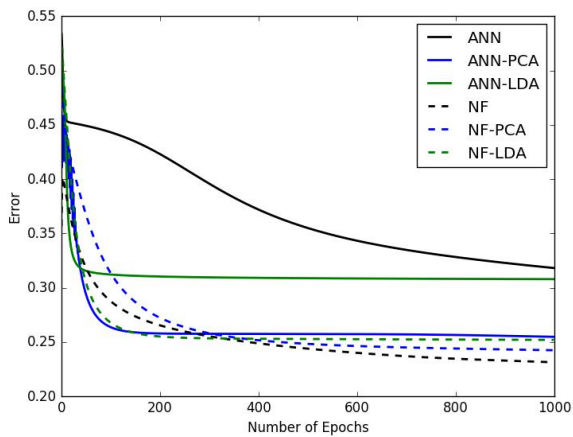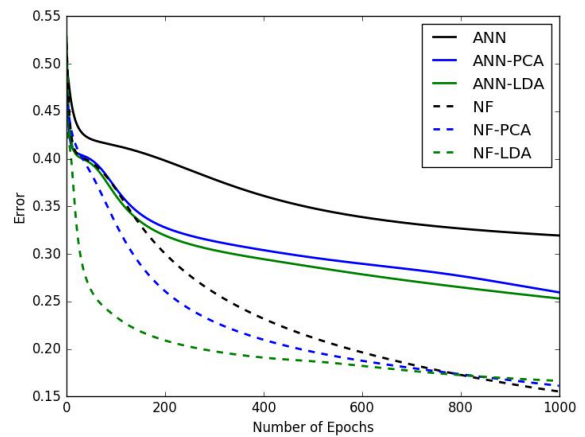Heart-statlog

Hepatitis

Iris

Liver

Lung Cancer

Mammographic (not correct)

Pima Indian diabetes

Hayes Roth
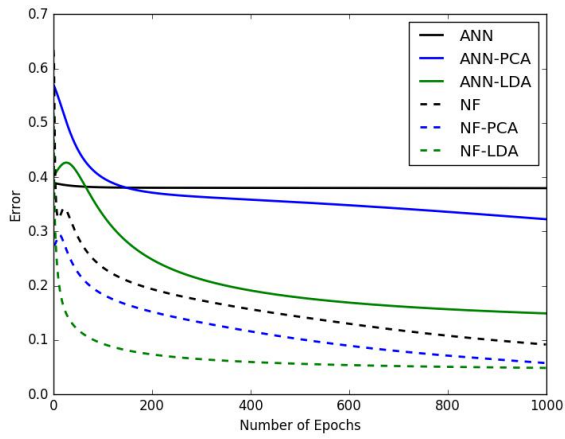
SPECTF heart
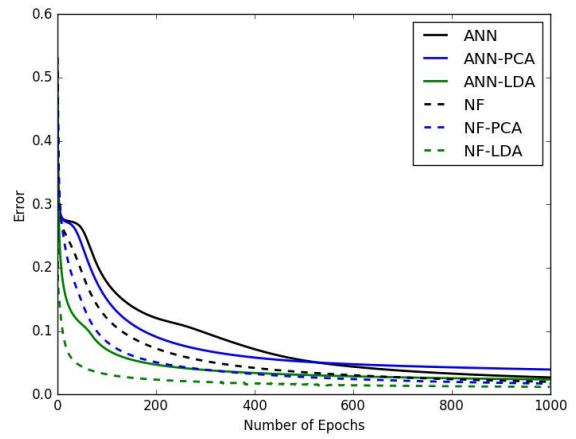
Thyroid
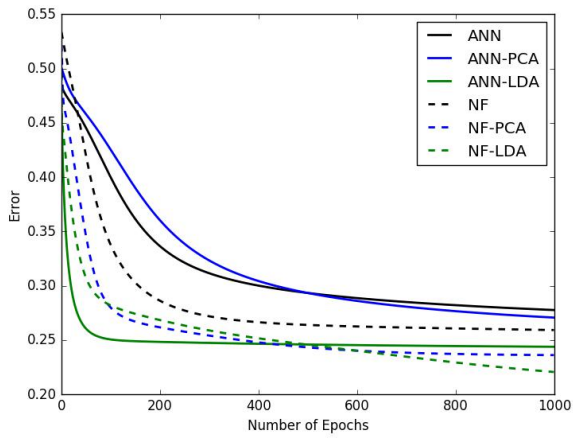
Titanic
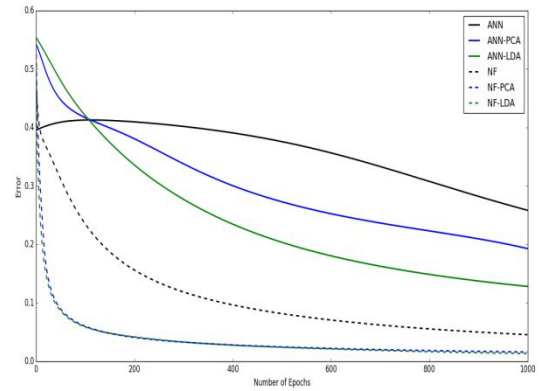
Wine

Table 6:   Comparison between precision of ANN,ANN-PCA,ANN-LDA,NF,NF-PCA,NF-LDA models

	ANN	ANN-PCA	ANN-LDA	NF	NF-PCA	NF-LDA
Pima Indian diabetes	0.393	0.633	0.566	0.903	0.936	0.8
Breast Cancer Wisconsin (Diagnosis)	0.95712	0.9649	0.9628	0.9810	0.9465	0.9808
Breast-Cancer-Wisconsin	0.972	0.942	0.934	0.914	0.925	0.920
Mammographic Mass	0.733	0.808	0.787	0.846	0.852	0.8505
Thyroid	0.953	0.933	0.95098	0.915	0.932	0.95833
Blood Transfusion Service Center	0.5015	0.129	0.486	0.7283	0.8105	0.781
Heart-statlog	0.5480	0.6182	0.8354	0.79003	0.8142	0.8647
Lung-cancer	0.599206	0.54605	0.6671	0.60837	0.5993	0.6676
SPECTF heart	0.9011	0.8713	0.8507	0.797	0.7623	0.6913
Haberman	0.6233	0.7014	0.7333	0.6335	0.5233	0.4
Liver	0.7365	0.72086	0.63461	0.70422	0.765	0.6724
Hepatitis	0.8427	0.9184	0.8996	0.596	0.5261	0.5364
Titanic	0.762	0.840	0.7356	0.718	0.816	0.7878
Wine	0.9155	0.8512	0.9135	0.9361	0.9629	0.764
Iris	0.918	0.942	0.8583	0.926	0.950	0.9111
Hayes Roth	0.683	0.750	0.557	0.781	0.801	0.8055

Table 7:   Comparison between Recall of ANN,ANN-PCA,ANN-LDA,NF,NF-PCA,NF-LDA models

	ANN	ANN-PCA	ANN-LDA	NF	NF-PCA	NF-LDA
Pima Indian diabetes	0.181	0.263	0.2654	0.750	0.723	0.684
Breast-Cancer-Wisconsin	0.359	0.327	0.366	0.343	0.324	0.330
Mammographic Mass	0.523	0.491	0.	0.545	0.532	0.5068
Thyroid	0.751	0.831	0.7555	0.868	0.944	0.9333
Blood Transfusion Service Center	0.0839	0.1015	0.8996	0.9923	0.97414	0.9255
Heart-statlog	0.74658	0.7908	0.79509	0.2525	0.2307	0.3005
Lung-cancer	0.6190	0.5992	0.666	0.65833	0.5701	0.6444
SPECTF heart	0.74658	0.7908	0.8507	0.797	0.7623	0.6913
Haberman	0.0807	0.074	0.2291	0.0699	0.0308	0.0416
Liver	**0.64022**	**0.70594**	0.66	**0.7346**	**0.6349**	0.6393
Hepatitis	0.885	0.9002	0.8391	0.0784	0.0683	0.0986
Titanic	0.389	0.324	0.3926	0.681	0.686	0.62275
Wine	0.9289	0.9437	0.9299	0.9419	0.9544	0.943
Iris	0.910	0.942	0.85833	0.928	0.953	0.88690
Hayes Roth	0.706	0.751	0.7103	0.754	0.758	0.6338

	ANN	ANN-PCA	ANN-LDA	NF	NF-PCA	NF-LDA
Pima Indian diabetes	0.2457	0.36807	0.361	0.8172	0.8147	0.720
Breast Cancer Wisconsin (Diagnosis)	0.7441	0.7590	0.7723	0.766	0.7766	0.7576
Breast-Cancer-Wisconsin	0.5005	0.49876	0.4997	0.485	0.4796	0.4361
Mammographic Mass	0.498	0.532	0.549	0.532	0.660	0.6351
Thyroid	0.8060	0.856	0.82544	0.904	0.9116	0.9407
Blood Transfusion Service Center	0.535	0.532	0.675	0.746	0.668	0.712
Heart-statlog	0.806	0.8281	0.820	0.38011	0.3545	0.4167
Lung-cancer	0.5202	0.5837	0.592	0.5796	0.577	0.6099
SPECTF heart	0.8623	0.9068	0.8665	0.1354	0.1117	0.1652
Haberman	0.432	0.883	0.3492	0.503	0.231	0.0754
Liver	0.6816	0.71286	0.64705	0.71804	0.6904	0.65546
Hepatitis	0.8623	0.9068	0.8665	0.1354	0.117	0.165
Titanic	0.492	0.486	0.512	0.601	0.655	0.6956
Wine	0.9186	0.936	0.9202	0.9352	0.956	0.9240
Iris	0.91057	0.9401	0.8518	0.923	0.9481	0.886
Hayes Roth	0.6726	0.73354	0.5643	0.751	0.7558	0.6628

Table 8:Comparison between f-measure of ANN,ANN-PCA,ANN-LDA,NF,NF-PCA,NF-LDA

5.4 Result Analysis ICA

	ANN-ICA	NF-ICA
Breast Cancer Wisconsin (Diagnosis)	0.351896944	0.212544297
Breast-Cancer-Wisconsin	0.090416162	0.126503211
Mammographic Mass	0.486073001	0.318301849
Thyroid	0.152491631	0.193272043
Blood Transfusion Service Center	0.314975183	0.25635183
SPECTF heart	0.283452936	0.339696851
Liver	0.471867005	0.481731111
Hepatitis	0.21330813	0.176598428
Titanic	0.356606866	0.29955751
Wine	0.078640561	0.212026417
Iris	0.141162016	0.15947706
Hayes Roth	0.377100898	0.172625415

Table 9 : RMS Error comparison
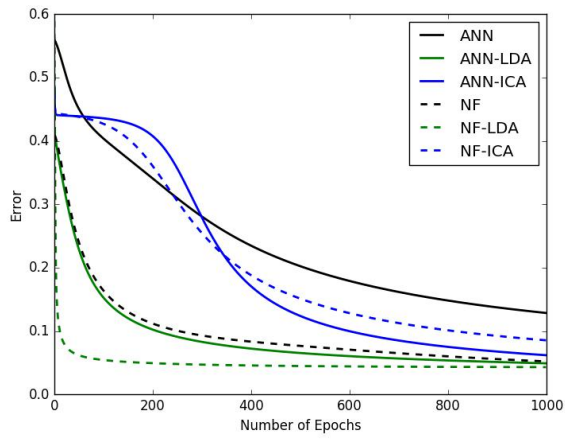
Breast-Cancer-Wisconsin
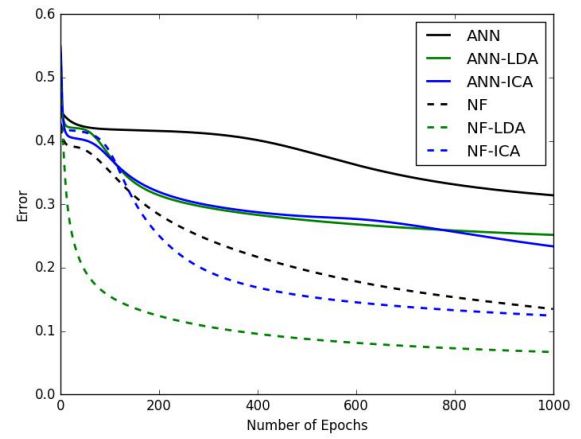
Hayes Roth

Breast Cancer Wisconsin
(Diagnosis)

Heart Statlog

Titanic

Wine

Figure 17 : Error plots of ANN,ANN-LDA,ANN-ICA ,NF,NF-LDA,NF-ICA

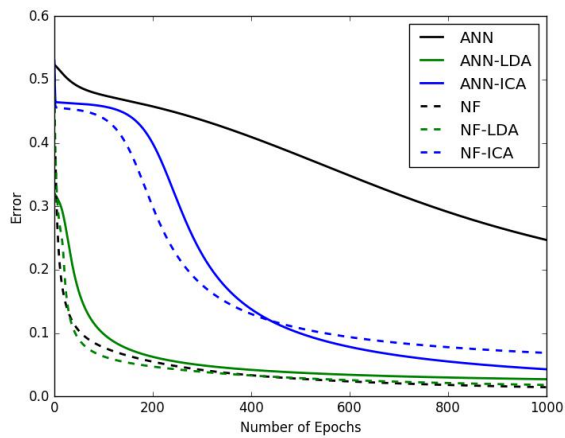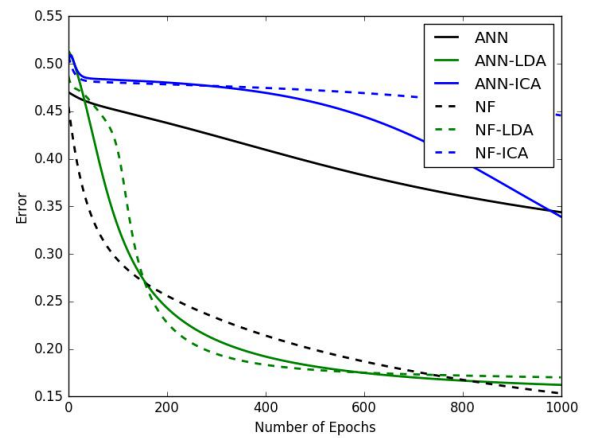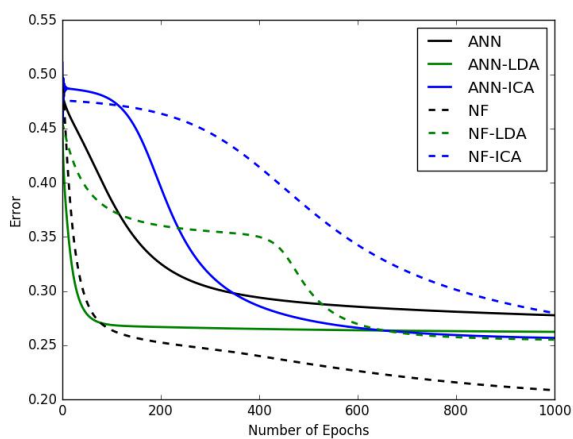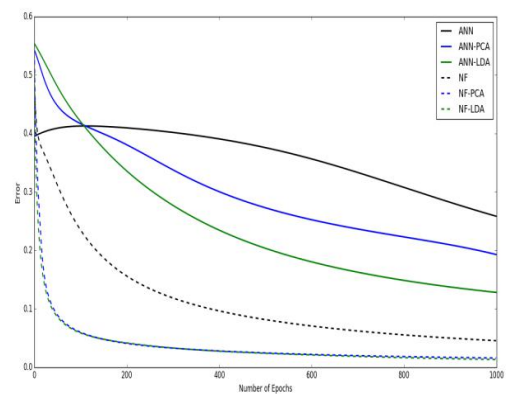|	ANN-ICA	NF-ICA
Breast Cancer Wisconsin (Diagnosis)	0.901846523	0.955765657
Breast-Cancer-Wisconsin	0.937619345	0.938488256
Mammographic Mass	0.768499224	0.83299485
Thyroid	0.963502533	0.97979
Blood Transfusion Service Center	0.696825397	0.813082905
SPECTF heart	0.866018928	0.867646
Liver	0.637609306	0.639341111
Hepatitis	0.87416035	0.72047619
Titanic	0.86072925	0.802584922
Wine	0.938173591	0.938495389
Iris	0.916395965	0.92781074
Hayes Roth	0.61715815	0.779251304

Table 9: Precision Comparison between ANN - ICA and FNN - ICA

Table 10: Recall Comparison between ANN - ICA and FNN - ICA

	ANN-ICA	NF-ICA
Breast Cancer Wisconsin (Diagnosis)	0.689054909	0.633723158
Breast-Cancer-Wisconsin	0.345550492	0.338000156
Mammographic Mass	0.499873625	0.527896668
Thyroid	0.830391214	0.906895949
Blood Transfusion Service Center	0.034326362	0.946149983
SPECTF heart	0.81631545	0.101779641
Liver	0.83175078	0.936032222
Hepatitis	0.888135845	0.090986924
Titanic	0.220265235	0.657350283
Wine	0.932594667	0.927825678
Iris	0.916583944	0.928096196
Hayes Roth	0.637246807	0.757745906

Table 11: F-Measure Comparison between ANN - ICA and FNN - ICA

	ANN-ICA	NF-ICA
Breast Cancer Wisconsin (Diagnosis)	0.779094261	0.761555105
Breast-Cancer-Wisconsin	0.50409261	0.496431786
Mammographic Mass	0.603822529	0.645920975
Thyroid	0.874945615	0.914262172
Blood Transfusion Service Center	0.064698304	0.873459649
SPECTF heart	0.8383003	0.170153446
Liver	0.720848372	0.759458778
Hepatitis	0.878750242	0.158912173
Titanic	0.317991706	0.721710771
Wine	0.930099196	0.92959701
Iris	0.907126745	0.923123214
Hayes Roth	0.589415261	0.749980961

5.5   Result Analysis of DE Algorithm

For Pima Indian diabetes Datatset

Accuracy	RMS Error	Precision	Recall	Fmeasure	Time(s)
83.10%	0.360748396	0.941176471	0.768	0.845814978	10.09336877
89.64%	0.372294078	0.711111111	0.820512821	0.761904762	10.342393
80.86 %	0.358852	0.92857	0.850467	0.8878048	12.8721482
86.59%	0.380096	0.90526315	0.796296296	0.84729064	11.6486775
78.24%	0.346030	0.90721649	0.8073394	0.8543689	9.99334669

Table 12 : Performance Measures recording for FNN with DE Algorithm

Accuracy	RMS Error	Precision	Recall	Fmeasure	Time(s)
77.38%	0.46747777	0.586956522	0.771428571	0.666666667	2.975111485
69.64%	0.472294078	0.711111111	0.820512821	0.761904762	3.854628563
78.90%	0.476177342	0.625	0.813953488	0.707070707	2.872148752
74.74%	0.459433028	0.620689655	0.782608696	0.692307692	2.848561525
76.22%	0.485747106	0.652777778	0.959183673	0.776859504	1.104115486

Table 13 : Performance Measures recording for ANN with DE Algorithm

As Pima Indian Diabetes Dataset is very small it perform same as feature reduction algorithm (FNN with PCA, LDA, ICA) but for large datasets this feature selection technique with DE algorithm perform better than the feature selection techniques.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

As per our objective we compared Fuzzy Neural Network with Principal component Analysis and without PCA to Artificial neural network with PCA and without models on the basis of different performance measures like Accuracy, RMS Error , Precision, Recall and F-Measure. It is tested on the four classic Dataset from UCI Repository . Proposed Model FNN with PCA classifier found is the finest among all classification models. Results drawn from the above study is completely based on our   implementation of powerful classification models.

In the Phase 2 results   among the six models 16 biomedical and popular datasets from UCI machine learning repository it was found that NF-LDA classifier performed best for 13 datasets.The datasets for which the model did not perform the best include the Pima Indian,Breast Cancer Wisconsin and Mammographic Mass datasets.

Due to some shortcomings of LDA we have also focused on techniques of feature selection and feature extraction (ICA and Optimisation Techniques)to test the accuracy of our proposed model. Due to high demand of soft computing and machine learning techniques, this proposed model can be used in various real life problems such as gene expression classification, document classification, and satellite image classification,biometric classification and identification,gesture detection.

6.2 Future Scope :

Fuzzy Neural Network have a huge application in the field various typical disease diagnoses like brain disorder, cardiac disease, breast cancer,alzheimer, thyroid disorder, leukemia, hypertension, heart disease . And FNN with PCA ,LDA,ICA and other optimisation techniques have that power to improve performance in the above field.

This is an ongoing research work and we will try to further improve the accuracy using factor analysis techniques.Fuzzy Neural Networks have already been used in a lot real life applications and also in a lot of research domains .Fuzzy Systems have performed well in their tasks thus making their use more extensive.

References

[1]Abraham, A. (2001, June). Neuro fuzzy systems: State-of-the-art modeling techniques. In *International Work-Conference on Artificial Neural Networks* (pp. 269-276). Springer, Berlin, Heidelberg.

[2]Ghosh, Ashish, B. Uma Shankar, and Saroj K. Meher. "A novel approach to neuro-fuzzy classification." *Neural Networks* 22, no. 1 (2009): 100-109.

[3]Pal, S. K., & Ghosh, A. (1996). Review Neuro-fuzzy computing for image processing and pattern recognition. *International Journal of Systems Science*, *27*(12), 1179-1193.

[4]Ghosh, Soumadip, Sushanta Biswas, Debasree Sarkar, and Partha Pratim Sarkar. "A novel Neuro-fuzzy classification technique for data mining." *Egyptian Informatics Journal* 15, no. 3 (2014): 129-147.

[5]Pal, S. K., & Mitra, S. (1999). *Neuro-fuzzy pattern recognition: methods in soft computing*. John Wiley & Sons, Inc..

[6]Kar, Samarjit, Sujit Das, and Pijush Kanti Ghosh. "Applications of neuro fuzzy systems: A brief review and future outline." *Applied Soft Computing* 15 (2014): 243-259.

[7]Sets, Fuzzy. "L. Zadeh." *Information and Control.–NY* 8/3 (1965): 338-353.

[8] Polat, Kemal, and Salih Güneş. "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease." *Digital Signal Processing* 17, no. 4 (2007): 702-710.

[9]Ghosh, Ashish, Nikhil R. Pal, and Sankar K. Pal. "Self-organization for object extraction using a multilayer neural network and fuzziness mearsures." (1993).

[10]KELE AL   and AYT RK KELE. "Extracting fuzzy rules for the diagnosis of breast cancer."*Turkish Journal of Electrical Engineering & Computer Sciences* 21, no. 5 (2013): 1495-1503.

[11]Kuncheva, Ludmila. *Fuzzy classifier design*. Vol. 49. Springer Science & Business Media, 2000.

[12]Wold, Svante, Kim Esbensen, and Paul Geladi. "Principal component analysis." *Chemometrics and intelligent laboratory systems* 2, no. 1-3 (1987): 37-52.

[13]Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." *Wiley interdisciplinary reviews: computational statistics* 2, no. 4 (2010): 433-459.

[14]Delen, Dursun, Glenn Walker, and Amit Kadam. "Predicting breast cancer survivability: a comparison of three data mining methods." *Artificial intelligence in medicine* 34, no. 2 (2005): 113-127.

[15]Haykin, Simon S., Simon S. Haykin, Simon S. Haykin, and Simon S. Haykin. *Neural networks and learning machines*. Vol. 3. Upper Saddle River, NJ, USA:: Pearson, 2009

[16] Lin, C-T., and C. S. George Lee. "Neural-network-based fuzzy logic control and decision system." *IEEE Transactions on computers* 40, no. 12 (1991): 1320-1336.