

# Dimensionality Reduction Techniques for Improved Diagnosis of Heart Disease

Swati Shilaskar

Vishwakarma Institute of Technology,  
Pune, India

Ashok Ghatol

Director, K.J. Educational Institutes, Pune,  
Former VC BATU, Lonere, Maharashtra

## ABSTRACT

Medical diagnosis is an important task that needs to be executed accurately and efficiently. Medical domain complexities are represented by multidimensional heterogeneous datasets. Computer aided diagnosis must deal with processing and analyzing high dimensional data. Optimization of features in datasets reduces time and memory complexity of learning algorithms. It is necessary to have a tool that gives relationship between features and eliminate redundant ones. Feature selection or feature extraction reduce dimensions and essentially influence the performance of classifier. Many techniques have been used to determine essential features of medical data. We investigate two feature extraction techniques, Principal component analysis (PCA) and common Factor Analysis (FA) techniques for classification of heart disease. These techniques expose the structure, while maintaining the integrity of the data, thus improving diagnosis performance.

## Keywords

Dimensionality reduction, PCA, FA, neural networks, AUC, heart disease

## Abbreviations

PCA Principal Component Analysis,  
FA Factor Analysis,  
AUC Area under Curve,  
DR Dimensionality Reduction.

## 1. INTRODUCTION

The term 'Heart disease' includes various diseases that affect heart. Heart disease is the major cause of casualties in the different countries including India. An effective analysis tool to discover hidden relationships and trends in data is necessary. These tools not only reduce dimensionality but also improve diagnostic ability of the system. Feature selection and feature extraction are the two major approaches to reduce dimensions. Feature selection generates a subset of original features by ranking features by weight or distance measures and then employing iterative search technique. [17]. Feature extraction is based on generation of a completely new feature space through qualitative mapping. This keeps fewer features to represent information in the data [5]. Feature extraction techniques are based on correlation between features whereas feature selection techniques are based on classification ability. Transformation of given data using feature extraction and diagnosis based on neural network is investigated in this paper.

## 1.1 Related Work

Feature extraction techniques are used more popularly for disease diagnosis and classification. Hasmarina and Nooritawati and Mykola et al [5;8] used it for heart disease diagnosis and stated as a promising approach. [1],[2],[6],[10],[12] Performed feature extraction for about 14 features of the heart disease data to develop intelligent decision support system for heart disease diagnosis.

## 2. METHODOLOGY

Features may be analyzed using parametric or non parametric methods. Parametric methods make assumptions that the variables under analysis have normal distribution [9]. Non-parametric methods are referred to as distribution-free methods. We apply parametric methods to medical data under test to obtain resultant subset that represents unique parameters of original features. A further reduced subset of the resultant feature set is subjected to neural network based classifier. Classifier performance averaged over 6 unique test sets is compared in this paper.

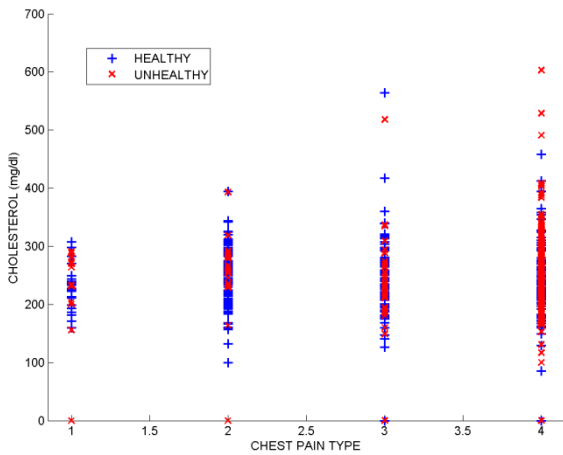
### 2.1 Data set description

Data is obtained from UCI centre for machine learning and intelligent systems [23]. This directory contains 4 databases concerning heart disease diagnosis. All attributes are numeric-valued. The data was collected from four locations Cleveland Clinic Foundation, Hungarian Institute of Cardiology, V.A. Medical Centre and University Hospital, Switzerland giving four datasets. Each database has identical features. Clinical symptoms of heart disease represented by 10 variables are used for analysis. The data set description is given in table 1.

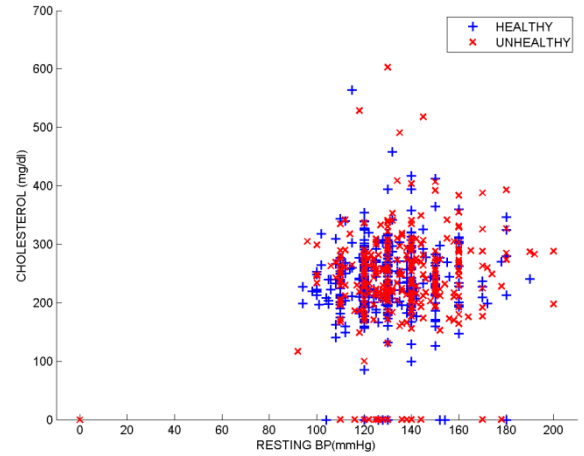
All 4 datasets are combined together to give total 920 samples, with 411 healthy and 509 unhealthy samples. Removal of vectors with missing values reduces dataset to total 686 samples with 352 samples of healthy patients and remaining 334 samples of unhealthy (suffering with heart disease) patients. Removal of rows with missing values may not qualify data as complete or perfect as two features have zero value for few samples. 'Resting BP' has zero value in one sample and Cholesterol level has zero value in 33 cases.

**Table 1. Dataset Description**

Sr No	Attribute	Description	Values
1	Age	In years	Range 28 to 77
2	Gender	Female/Male	0/1
3	CP - Chest Pain Type	Typical /Atypical / Non-angina / Asymptotic	1/2/3/4
4	TRESTBPS	Resting BP (mmHg)	Range 0 to 200
5	CHOL	Cholesterol (mg/dl)	Range 0 to 603
6	FBS	Fasting blood sugar >120mg/dl F/T	0/1
7	RESTECG	Resting ECG results Normal/ST abnormal /left ventricle hypertrophy	0/1/2
8	THALACH	Max Heart rate achieved	Range 69 to 202
9	EXANG	Exercise induced angina No/ Yes	0/1
10	OLDPEAK	ST depression induced by exercise	Range - 0.5 to 6.2
11	CLASS	Healthy/Unhealthy	0/1



**Fig 1. Chest pain type VS Cholesterol levels**



**Fig 2. Resting BP value VS cholesterol levels**

Figure 1 and figure 2 show plots of feature ‘Chest pain type’ VS ‘Cholesterol levels’ and ‘Resting BP’ VS ‘Cholesterol levels’ respectively. The plots clearly show that the feature clusters are not linearly separable. It is evident that the clusters formed are complex and hence accurate diagnosis of heart disease is challenging. These features along with other features as listed in Table 1 are significantly sensitive for discrimination of patient as healthy or unhealthy.

## 2.2 Dimensionality Reduction Techniques

Factor analysis (FA) and principal component analysis (PCA) are feature extraction based data reduction methods. Correlation between variables in a data set is analyzed and the variables are reduced to a smaller number. FA method gives factors whereas PCA gives components. Both methods give a set of ‘loadings’ and a set of ‘scores’ [26]. Loadings are correlations between original variables and extracted factors/components. Scores are values each data item gets on the extracted factors/components after data reduction [26].

Eigen values and eigenvectors of the matrix variables with high correlation are clustered together on components or factors. The eigenvectors (loadings) are sorted according to decreasing eigen values, giving components in the order of significance. Most significant components selected using rules given in section 2.2.1.1, give reduced feature set. PCA analyzes all variance present in the data set, while FA analyzes only common variances that are uncontaminated by error variability. Thus FA is less sensitive to noise in the data [17].

### 2.2.1 Principal Component Analysis

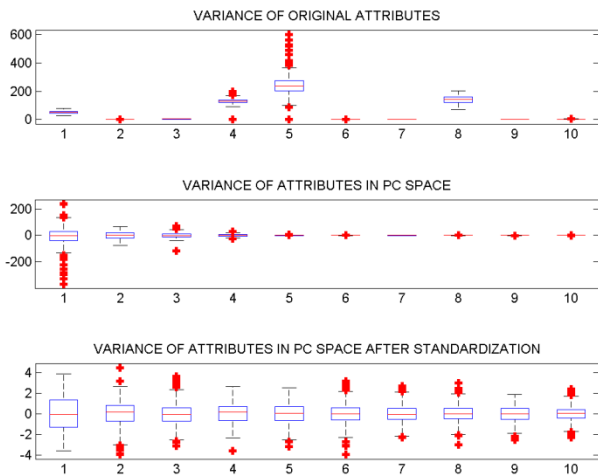
PCA is a model for the covariance structure expanded in sorted set of components with decreasing variance [16]. The intrinsic variability of data is captured by linear extraction of features from the original feature set. [6] [22].

Principal components based on correlations are determined using mean centered data calculated by (1).

$$z = \frac{(x - \text{mean}(x))}{sd(x)} \quad (1)$$

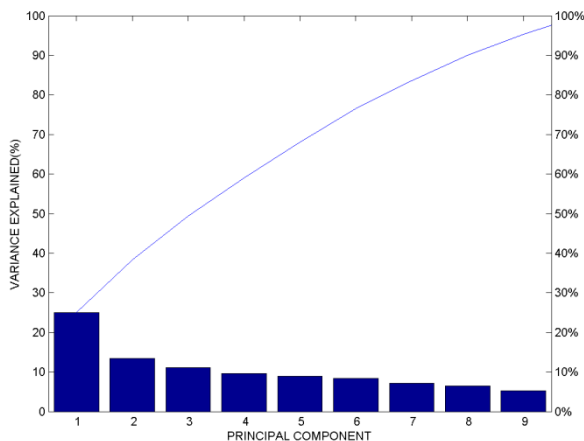
$x$  - feature value

$sd$  - standard deviation



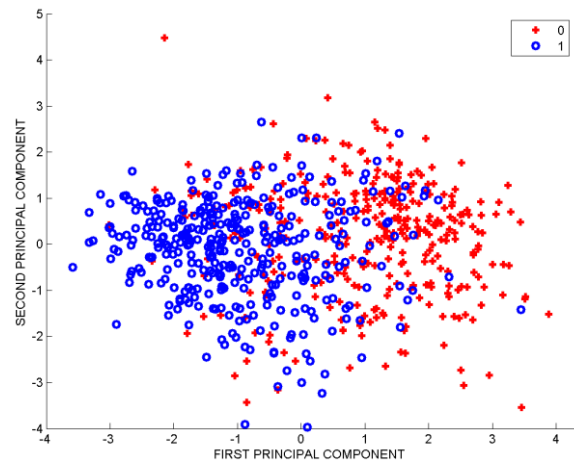
**Fig 3. Variance of attributes**

Figure 3 shows the variance of attributes in principal component space. First principal component is in the direction of greatest variability (covariance) in the data, second is in the next orthogonal (uncorrelated) direction of greatest variability. Components equal to the number of features were generated. The projections of these components correspond to the eigen values of the input covariance matrix. Cumulative sum of variance of principal components is shown in figure 4.



**Fig 4. Variance of attributes added cumulatively**

Principal components are ordered as per their eigen values. First VS second principal components plot shows two classes distinctly in the principal component space. (Ref figure 5)



**Fig 5. Scatter plot showing two classes**

### 2.2.1.1 Retention of Components

The variance from correlated variables is packaged into  $p$  uncorrelated components by PCA. Some of the components have large eigen values and some have small eigen values. Following rules suggest retention of components in subset.

- Kaiser's rule: Suggests to retain only components with eigen value equal to one or more [4], [5].
- Elbow test: Elbow rule hints to plot the eigen values graphically and look for the elbow i.e. for the point after which the remaining eigen values decrease in linear fashion and retain only those above the elbow [8]. Scree test is a graphical device for deciding on the number of components to retain [4]. Pareto function based scree plot of variability of each principal component is given in figure 4.

Kaiser rule based subset (ref. Table 2) has first three principal components, Elbow rule allows components above 0.9 Eigen value and picks five components. The number of components extracted is equal to the number of features of the transformed dataset which completely explains the variance in each feature. This gives us PCA3 [686x3] and PCA5 [686x5] matrices using Kaiser and Elbow rule respectively.

**Table 2. Principal component analysis**

Principal Component	Eigen values	Progressive Contribution
1	2.50	24.94
2	1.35	38.40
3	1.11	49.44
4	0.96	59.03
5	0.90	68.01
6	0.84	76.44
7	0.71	83.55
8	0.65	90.04
9	0.53	95.32
10	0.47	100.00

### 2.2.2 Factor Analysis

Medical dataset features are based on measured parameters as well as symptoms. There are chances that the variables overlap, in a sense that symptoms may be dependent on measured parameters or some other symptom. Thus a group of variables may be dependent giving redundant variables. Factor analysis estimates interdependence of attributes in multivariate data [18]. PCA extracts all the factors underlying a set of variables whereas FA analyzes only the shared variance i.e. common factor. The factor analysis model for  $k$  common factors (with  $k < p$ ) is given by (2).

$$x = \mu + \Lambda f + u \quad (2)$$

Where

- $k$  - Number of common factors
- $p$  - Number of original attributes
- $x$  - Vector of observed variables,
- $\mu$  - Constant vector of means,
- $\Lambda$  - Constant d-by-m matrix of factor loadings,
- $f$  - Standardized common factors of length m,
- $u$  - Vector of independent specific factors.

Length of  $x$ ,  $\mu$  and  $u$  is  $d$ .

In a factor analysis model, the measured variables depend on few dormant factors. Each of the factors affects several variables in common, hence they are known as *common factors* (referred to as 'Factors 3' in this paper). Each variable can be represented by a linear combination of the common factors. The coefficients of this linear equation are known as *loadings*. Each measured variable includes a component due to independent random variability, known as *specific variance* because it is specific to one variable. Factor analysis models the correlation structure in terms of  $k$  factors including measurement errors. Unlike PCA the factors are not sorted by any criterion.

The '*loadings*' of all 10 attribute on three common factors is calculated. The biplot shows the common factors together with the projections of the axes of the original variables. Figure 6 shows biplot of 10 features' vector position and values displayed as vectors, with respect to the three common factors. FA defines a  $k$ -dimensional subspace. With  $k = 3$ , three columns of common factor matrix (Factors3) now represent 10 attributes reducing feature set dimension to [686 x 3] matrix.

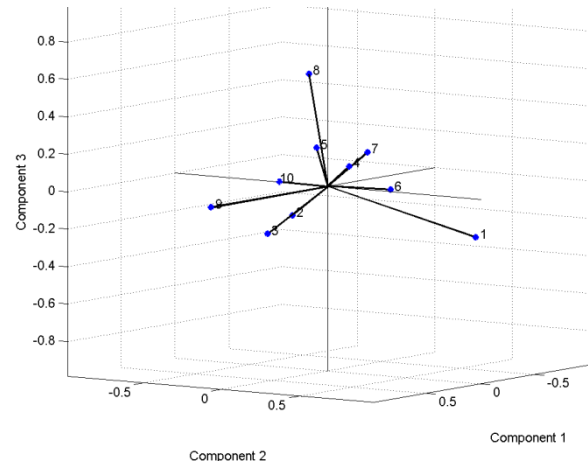


Fig 6. Feature values with respect to common factors

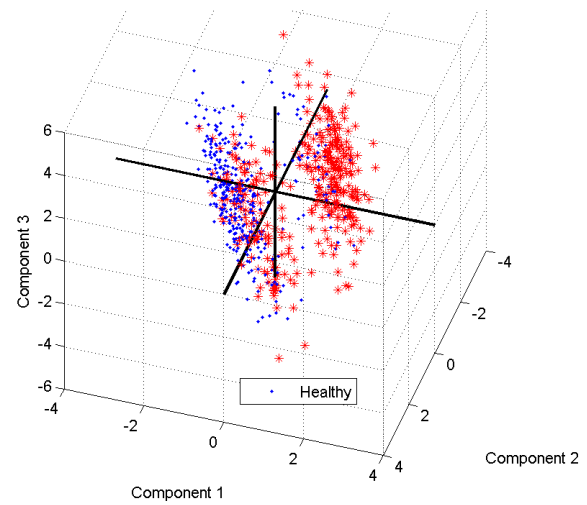


Fig 7 (a). Three common factors

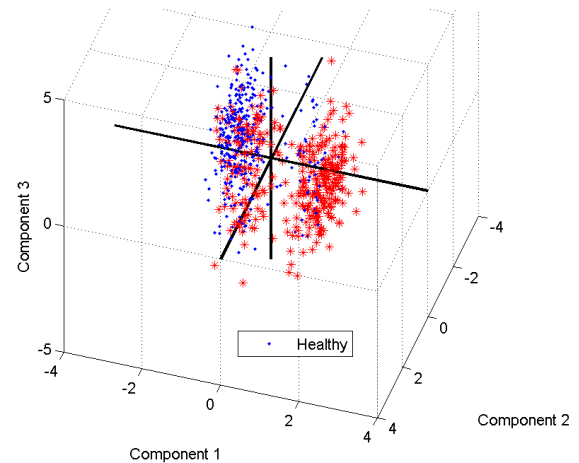


Fig 7 (b). Rotated axes of common factors

Figure 7 (a) shows the samples plotted on the three common factor axes. If the axes pass through the clusters more nearly, interpretability increases. The axes are rotated to improve interpretability as shown in figure 7 (b). After rotating the axes the loading matrix is recomputed. Interpretability can be improved by rotation of factor loadings and scores. Maximum likelihood estimation (MLE) of each original attribute is obtained with all extracted common factors. Table 3 shows MLE of the rotated factor loadings matrix.

**Table 3. Common factor MLE loadings on attributes**

Factor Analysis with 3 common factors			
Attribute Number	Common Factor 1	Common factor 2	Common factor 3
1	-0.08	0.90	-0.22
2	0.24	-0.07	-0.14
3	0.53	-0.02	-0.20
4	0.14	0.24	0.14
5	0.05	-0.04	0.21
6	-0.08	0.35	0.00
7	0.00	0.26	0.20
8	-0.30	-0.33	0.55
9	0.87	-0.14	-0.03
10	0.53	0.06	0.08

### 2.3 Neural network model as classifier

The classifiers used here are multilayer neural networks with linear and sigmoid outputs transfer functions. For all the classifiers we used randomly selected 412 samples (60%) for training and 136 (20%) each for validation and testing. Numerical optimization technique finds a weight vector  $w$  that minimizes given energy function. Starting with random generated weights, Levenberg-Marquardt (LM) algorithm is used for back propagation learning of neural networks. Network weights are updated as given (3)

$$w(k+1) = w(k) - [J_k^T J_k + \lambda_k]^{-1} J_k^T e_k \quad (3)$$

For very small value of  $\lambda_k$  the equation approaches Newtons algorithm as given in (4). For increased value of  $\lambda_k$  the second term inside square bracket becomes dominant and updated equation can be written as (5) which is steepest descent gradient method. LM algorithm represents transition between steepest descent and Newton's method of optimization [11], [21], [24] and [25].

$$w(k+1) = w(k) - [J_k^T J_k]^{-1} J_k^T e_k \quad (4)$$

$$w(k+1) = w(k) - [\lambda_k]^{-1} J_k^T e_k \quad (5)$$

$$J = \begin{bmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \dots & \frac{\partial e_1}{\partial w_N} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \dots & \frac{\partial e_2}{\partial w_N} \\ \dots & \dots & \dots & \dots \\ \frac{\partial e_p}{\partial w_1} & \frac{\partial e_p}{\partial w_2} & \dots & \frac{\partial e_p}{\partial w_N} \end{bmatrix} \quad (6)$$

Where

$w$  – weight vector

$\lambda$  - scalar number to control number of iterations

$k$  – 1,2,3... $N$  discrete time index

$I$  –  $N \times N$  identity matrix

$e$  - error

$J$  – jacobian matrix represented by (6)

We are using three types of neural classifiers PURELIN, TANSIG, CAS2. PURELIN and TANSIG are feedforward neural networks with linear and sigmoid output layer transfer functions respectively. Linear function allows the network outputs to take on any value from 0 to 1. Sigmoid transfer function limits outputs of network to a small range. Sigmoid functions are characterized by the fact that their slopes approach zero as the input gets large. The gradient can have a very small magnitude and, therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values. Thus time and number of epoch for training is more as compared to linear transfer function based networks.

CAS2 are cascade-forward network with multiple layers. First layer has weights coming from the input. Each subsequent layer has weights coming from the input and all previous layers. All layers have biases. The last layer is the network output with linear transfer function.

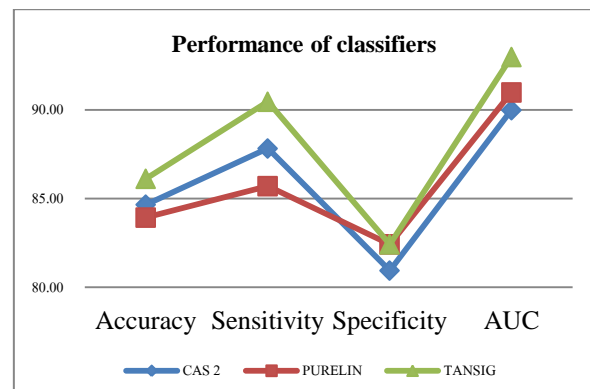
### 2.4 Performance Measures

In classification problems, the purpose of the network is to assign each case to one of the classes. In two-state representation, a single node corresponds to the class, and a value of 0 is interpreted as one state, and a value of 1 as the other state. During training and evaluation, neural networks take samples as input and generate output. The output units have continuous values between 0.0 and 1.0. Closeness of outputs to 0.0 or 1.0 decides the output class. If they are not, the class is regarded as inconclusive or undecided.

**Table 5. Confusion Matrix**

	Classified healthy	Classified Unhealthy
Healthy	TP (True Positive)	FN (False Negative)
Unhealthy	FP (False Positive)	TN (True Negative)

While assessing the classification ability of the network, most important indicator is the classification summary spreadsheet i.e. confusion matrix as shown in Table 5. It shows correct and incorrect classification [3]. Table 6 lists performance measures and individual calculation formula [18]-[20].



**Fig 8. Classifier Performance with PCA5 technique**

Accuracy is the most popularly used standard for comparison of performance. The sensitivity criterion of a diagnostic test is the proportion of patients for whom the outcome is positive that are correctly identified by the test. The specificity is the

proportion of patients for whom the outcome is negative that are correctly identified by the test. These parameters are plotted for classifiers in figure 8.

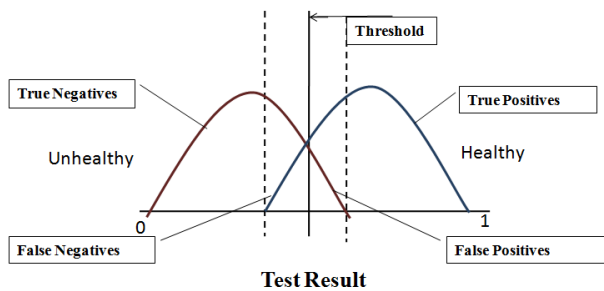


Fig 9. Classifier output VS number of samples

False Positive Rate and True Positive Rate show relative trade-off between true positive and false positive values. As shown in figure 9, when plotted with varying threshold they exhibit receiver operating characteristic curve [13], [15]. The value for the area under the ROC curve (AUC) can be interpreted as follows: an area of 0.80 means that a randomly selected individual from the positive group has a correct test value 80% of the time [14]. The area is 0.5 and the ROC curve coincides with the diagonal when there is no difference between positive and negative distributions. When there is a perfect separation of the values of the two groups, the area under the ROC curve equals 1. ROC curves value when equals 1 they reach upper left corner of the graph and are classified as excellent.

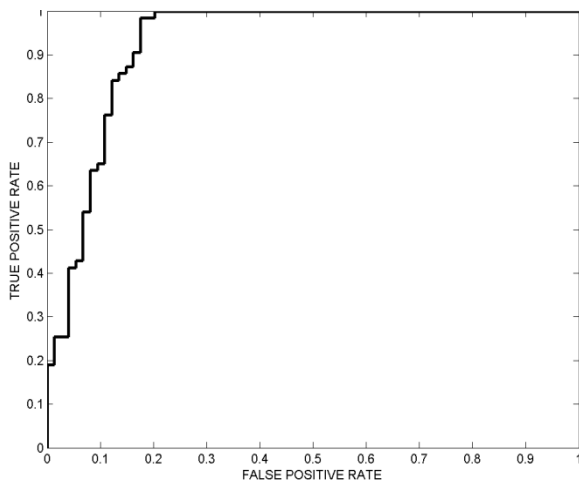


Fig 10. ROC curve [PCA 5, TANSIG (3 PE), AUC 93%]

The area under the ROC curve (AUC) is widely recognized as the measure of a diagnostic test's discriminatory power [7]. It is important as well as nontrivial to achieve high value sensitivity and specificity both. Incorrect positive diagnosis of disease may subject a patient to stress, unnecessary tests or medication. Whereas incorrect negative diagnosis deprives the patient of much needed medical attention and may put his life at risk. The cost of misclassification is a significant area of research and is out of scope of the investigations in this paper.

High sensitivity indicates power of positive predictions. Negative predictions are indicated by specificity parameter (ref table 6). Figure 10 exhibits ROC for PCA5 subset using neural network with TANSIG transfer function.

AUC for this curve is 93%, which is indicative of improved positive as well of negative diagnosis.

Table 6. Performance Measures

Measure	Formula	Intuitive Meaning
Accuracy	$\frac{(TP + TN)}{(Total\_Input)}$	Correct predictions
Sensitivity - True Positive Rate (TPR)	$\frac{TP}{(TP + FN)}$	Positive predicted positive labels
Specificity - True Negative Rate (TNR)	$\frac{TN}{(TN + FP)}$	Negative predicted negative labels
False Positive Rate (FPR) 1-Specificity	$\frac{FP}{(TN + FP)}$	Positive predicted negative labels
Geometric Mean (GM)	$(TPR * TNR)^{1/2}$	Geometric mean of sensitivity and specificity
Area Under Curve (AUC)	Receiver operating characteristics FPR vs. TPR	Trapezoidal approximation of area under ROC
Mean Squared Error (MSE)	$\frac{1}{Q} \sum_{k=1}^a (t_k - o_k)^2$	Q - total instances, t - target value, o - output of classifier
Precision	$\frac{TP}{(TP + FP)}$	Correct positive predictions
Recall	$\frac{TP}{(TP + FN)}$	Positive predicted positive labels
F - Measure	$\frac{2 * Recall * precision}{(Recall + Precision)}$	Harmonic mean between precision and recall
Max Epoch	Number of iterations required for training	

### 3. RESULT AND DISCUSSION

Technology plays a crucial role in decision support system for medical diagnosis. Due to advances in the field of medical instrumentation, a large number of measured attributes are available for diagnosis. PCA and FA feature extraction techniques are implemented on heart disease database to find non correlated features. Performance of neural network classifier with and without dimensionality reduction techniques is investigated in this paper.

We have used LM optimization for neural network classifier training. Hessian-based algorithms allow the network to learn more subtle features of a complicated mapping. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method which is slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a Gauss-Newton method. This improves classifier performance.

Along with typical measurement indices, interpretation of AUC is discussed. Techniques used for dimensionality

reduction show improved performance compared to original all features. The performances of classifiers are tabulated in table 7. PCA subset exhibits improvement in classification performance as compared to original feature set and FA based subset.

Feature selection techniques that do not involve transformation of data, work well on data with missing values. Parametric approach for feature extraction limits this flexibility. Immunity to three factors viz. noise, rotation and redundancy are the major advantages of PCA and FA based feature extraction techniques. Thus the benefits of these methods surmount their limitations. We conclude that

dimensionality reduction technique should be seen as an approach to improve classification performance and not just as a pre-processing technique.

Discrimination power of components may be explored further by using wrapper approach. Component selection not just by variability but also by classification performance is expected to improve diagnostic performance with reduced dimensions. Finally, developing a wrapper approach could be possible direction for future research.

**Table 7. Classification results**

Performance	CAS2				PURELIN				TANSIG			
	Original 10	Factor 3	PCA 5	PCA 3	Original 10	Factor 3	PCA 5	PCA 3	Original 10	Factor 3	PCA 5	PCA 3
<b>DR technique</b>												
<b>%Accuracy</b>	82.48	82.48	83.9	<b>84.6</b>	81.02	80.29	83.2	<b>83.9</b>	78.83	81.02	<b>86.1</b>	83.2
<b>%Sensitivity</b>	85.71	77.78	85.5	<b>87.8</b>	84.85	87.10	84.1	<b>85.7</b>	85.07	88.89	<b>90.4</b>	87.1
<b>%Specificity</b>	79.10	87.69	82.3	<b>80.9</b>	77.46	74.67	82.4	<b>82.4</b>	72.86	74.32	<b>82.4</b>	80.0
<b>%FPR</b>	20.90	12.31	17.6	<b>19.0</b>	22.54	25.33	17.5	<b>17.5</b>	27.14	25.68	<b>17.5</b>	20.0
<b>%Geometric mean</b>	42.32	30.94	38.8	<b>40.9</b>	43.73	46.97	38.4	<b>38.8</b>	48.05	47.77	<b>39.8</b>	41.7
<b>AUC</b>	0.87	0.91	0.89	<b>0.90</b>	0.89	0.88	0.89	<b>0.91</b>	0.88	0.88	<b>0.93</b>	0.89
<b>MSE</b>	0.01	.001	.001	<b>0.02</b>	0.01	0.03	0.01	<b>0.03</b>	0.01	0.01	<b>0.01</b>	0.01
<b>Precision</b>	0.81	0.88	0.83	<b>0.84</b>	0.78	0.74	0.80	<b>0.81</b>	0.75	0.75	<b>0.81</b>	0.78
<b>Recall</b>	0.86	0.78	0.86	<b>0.88</b>	0.85	0.87	0.84	<b>0.86</b>	0.85	0.89	<b>0.90</b>	0.87
<b>F-measure</b>	0.83	0.82	0.84	<b>0.86</b>	0.81	0.80	0.82	<b>0.83</b>	0.80	0.81	<b>0.86</b>	0.82
<b>Max no of epoch</b>	11	9	12	<b>8</b>	10	15	11	<b>10</b>	14	48	<b>14</b>	28

Note: The columns in bold italics represent best performance of classifier with reduced dimensionality

#### 4. REFERENCE

- [1] Alok Sharma, Godfrey C. Onwubolu, 2007. A Hybrid Approach for Modeling High Dimensional Medical Data .International Workshop on Inductive Modelling. Prague, 39-45.
- [2] Asha Gowda Karegowda, M.A.Jayaram, A.S .Manjunath, 2011 . Feature Subset Selection using Cascaded GA & CFS: A Filter Approach in Supervised Learning. International Journal of Computer Applications . 23–2, 10.
- [3] Asha.T , Dr. S. Natarajan, Dr. K.N.B. Murthy, 2010 . Diagnosis of Tuberculosis using Ensemble methods. IEEE Catalog Number: ISBN: CFP1057E-PRT . 978-1-4244-5537-9, 409-412.
- [4] Donald Jackson, 1993 . Stopping rules in Principal Component Analysis: A Comparison of Heuristical and Statistical Approaches . Ecology. 74, 2204-2214.
- [5] Hasmarina Hasan, Nooritawati Md Tahir , 2010. Feature Selection of Breast Cancer Based on Principal Component Analysis. 6th International Colloquium on Signal Processing & Its Applications (CSPA), 242-245.
- [6] J.Padmavathi 2012 . A Comparative Study on Logistic Regression Model and PCA-Logistic Regression Model in Medical Diagnosis. International Journal of Engineering and Technology . 2 - 8, 1372-1378 .
- [7] Jerome Fan, Suneel Upadhye, Andrew Worster 2006. Understanding receiver operating characteristic (ROC) curves Pedagogical Tools and Methods. Can J Emerg Med 8(1):19-20 .
- [8] Mykola Pechenizkiy, Alexey Tsymbal, Seppo Puuronen , 2004. PCA-based Feature Transformation for Classification: Issues in Medical Diagnostics. 17th IEEE Symposium on Computer-Based Medical Systems .
- [9] Philip Sedgwick ,2012. Statistical Question Parametric v non-parametric statistical tests. BMJ Journal. 344 doi: <http://dx.doi.org/10.1136/bmj.e1753>
- [10] Ranjana Raut, S.V.Duddul, 2010 . Intelligent Diagnosis of Heart Diseases using Neural Network Approach. International Journal of Computer Applications .1 – 2, 97-102.
- [11] Sudarshan Nandy, Partha Pratim Sarkar, Kalyani, Nadia, Achintya Das. 2012 An Improved Gauss-Newton's Method based Back-propagation Algorithm for Fast Convergence. International Journal of Computer Applications (0975 – 8887) vol39– No.8, February 2012. 1-7.

- [12] Sunila Godara, Nirmal . Intelligent and Effective Decision Support System Using Multilayer Perceptron. *International Journal of Engineering Research and Applications*. 1-3, 513-518.
- [13] Yonghong Peng, Zhiqing Wu, Jianmin Jiang ,2010 . A novel feature selection approach for biomedical data classification. *Journal of Biomedical Informatics*, 43. No. 1,15-23.
- [14] Zweig MH, Campbell G., 1993 . Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem. Apr* 39-4:561-77
- [15] Charles X. Ling Jin Huang Harry Zhang, 2003 . AUC: a Statistically Consistent and more Discriminating Measure than Accuracy . 18th International Conference On Artificial Intelligence .
- [16] Gang Zheng, Ya-Lou Huang, Peng-Tao Wang , Guang-Fu Shu, 2005. Study Of Classification Rules On Weighted Coronary Heart Disease Data. Fourth International Conference On Machine Learning And Cybernetics, Guangzhou, 18-21 , 1845-1850. ISBN
- [17] Kyungim Baek and Bruce A. Draper, 2002. Factor Analysis for Background Suppression . International Conference on Pattern Recognition Colorado.
- [18] Naeem Seliya, Taghi M. Khoshgoftaar, Jason Van Hulse, 2009. Aggregating Performance Metrics for Classifier Evaluation . IEEE IRI , July 10-12,, Las Vegas, Nevada, USA, 35-40.
- [19] Shivajirao M. Jadhav, Sanjay L. Nalbalwar, Ashok A. Ghatol, 2010. Generalized Feedforward Neural Network based Cardiac Arrhythmia Classification from ECG Signal Data. *Advanced Information Management and Service (IMS)*’, 6th International Conference, 351-356.
- [20] V. Garc’ia, R.A. Mollineda and J.S. S’anchez, 2010. Theoretical Analysis of a Performance Measure for Imbalanced Data . International Conference on Pattern Recognition Pub in IEEE Computer society, 617-620.
- [21] Deepak Mishra, Abhishek Yadav, Sudipta Ray, and Prem K. Kalra . Levenberg-Marquardt Learning Algorithm for Integrate-and-Fire Neuron Model. *Neural Information Processing - Letters and Reviews* Vol.9, No.2, 41-51
- [22] Jonathon Shlens , 2009. A Tutorial on Principal Component Analysis .April 22 . Version 3.01, tutorial available at <http://www.sn1.salk.edu/~shlens/pca.pdf>.
- [23] UCI ML Repository. <http://archive.ics.uci.edu/ml/>
- [24] Fredric M. Ham, Ivica Kostanic , 2002. Principles of Neurocomputing for science and Engineering. Tata McGraw – Hill Edition.
- [25] Vojislav Kecman , 2004. Learning and Soft Computing. The MIT Press, Cambridge.
- [26] Principal Component Analysis and Factor Analysis. Therese Leinonen. Seminar in Statistics and Methodology, 25th February, 2009 <http://www.let.rug.nl/~nerbonne/teach/rema-stats-meth-seminar/presentations/leinonen-pcfa-2009.pdf>