Sunil Kumar Prabhakar / Harikumar Rajaguru / Vinoth Kumar Bojan

Feature Extraction and Different Classifiers Applied for Detection of Abnormalities in Computer Tomography (CT) Images

Scientific Study

GRIN

Bibliographic information published by the German National Library:

The German National Library lists this publication in the National Bibliography; detailed bibliographic data are available on the Internet at http://dnb.dnb.de .

Imprint:

Copyright © 2014 GRIN Verlag
ISBN: 9783656894544

This book at GRIN:

https://www.grin.com/document/287937

Sunil Kumar Prabhakar, Harikumar Rajaguru, Vinoth Kumar Bojan

Feature Extraction and Different Classifiers Applied for Detection of Abnormalities in Computer Tomography (CT) Images

GRIN Verlag

FEATURE EXTRACTION AND DIFFERENT CLASSIFIERS APPLIED FOR DETECTION OF ABNORMALITIES IN COMPUTER TOMOGRAPHY (CT) IMAGES

HARIKUMAR RAJAGURU

SUNIL KUMAR PRABHAKAR

VINOTH KUMAR BOJAN

PREFACE

This book has been written to initiate the newcomers who are willing to work in the field of Feature Extraction Concepts for Medical Images which is one of the fastest growing fields in the engineering world. A short Description about the different classifiers applied for the detection of abnormalities of medical images which are at the core of design, implementation, research, and invention of new image processing techniques are presented in this text. This book will be useful for the practicing engineers, as well for researchers, graduate students, and undergraduate students.

ACKNOWLEDGEMENT

The authors are always thankful to the Almighty for perseverance and achievements. They wish to thank the management and authorities of Bannari Amman Institute of Technology, Sathyamangalam, for providing an academic environment and great encouragement given in the successful endeavour of the book. The authors are highly grateful to the Grin publishing group.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This chapter introduces the basic concepts of selecting Region of Interest (ROI), feature extraction and different classifiers applied for detection of abnormalities in Computer Tomography (CT) images. Selecting the region of interest in general cover the preprocessing procedure which involves the better classifier performance.

1.1 MEDICAL IMAGE PROCESSING

Computer aided medical diagnosis is a continuously growing field of research. Image based medical diagnosis techniques mostly relay on proper extraction of features from the input images and their subsequent classification. One of the main difficulties faced by this field is the requirement of huge memory space to store the medical image and computational time needed to process the data. The main objective of feature extraction is the automatic extraction of features from the input, to represent it in a unique and compact form of a single value or matrix vector. Though there are many techniques available in medical image processing and classification, the most prominent method for analysis is that in the wavelet domain. Many image processing applications uses wavelet transforms for data analysis which is an advanced technique in signal and image analysis. This is introduced as an alternative to short term Fourier transforms which suffers from issues related to frequency and time resolution properties.

1.2 STATEMENT OF THE PROBLEM

Abnormality detection using classifiers is one of the recent research areas where much importance is given. It is one of the critical issues where excessive care needs to be taken for better diagnosis. An input image may contain excessive information either wanted or unwanted which depends upon the problem formulation. The problem in this project is to analyze the performance of the classifier in terms of its efficiency in detecting abnormalities in medical images. Any classifier needs to detect the carcinogenesis with respect to the efficiency in time of detection and performance. Here two classifiers are

selected namely Singular Value Decomposition (SVD), and Principle Component Analysis (PCA). Both the SVD and PCA are applied for dual class classification procedure. The performance analysis of all these classifiers are analyzed using the classifier performance measures like, Sensitivity, Selectivity, Average Detection, Perfect Classification, Missed Classification, False Alarm, F-score and Quality Metrics. Here CT images of brain and skull are used for analysis. Two sets of 30 images are taken which contain both normal and abnormal ones. Fig 1.1 shows the general architecture of the image classifier system.

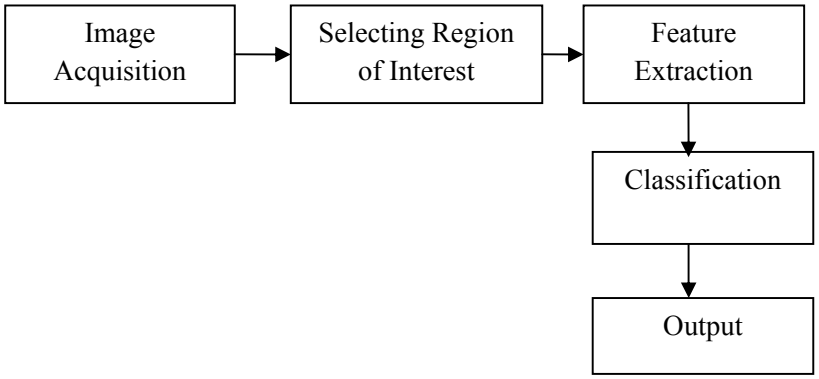

Figure.1 General Architecture of the Image Classifier System

1.3 REGION OF INTEREST

Selection of ROI is an essential preprocessing stage when it comes to abnormality detection. In cancer detection process efficiency of detection is purely based on the region of interest selection. This stage can improve the efficiency and performance of the feature extraction as well as classification stages, as only the required regions are taken for further processing.

1.4 FEATURE EXTRACTION

Maximizing the joint dependency with a minimum size of variables is generally the main task of feature selection. For obtaining a minimal subset, while trying to maximize the joint dependency with the target variable, the redundancy among selected variables must be reduced to a minimum. Feature selection is one of the most crucial steps of many pattern recognition and artificial intelligence problems. There are two general

approaches to feature selection: filters and wrappers. Filter type methods are essentially data pre-processing or data filtering methods. Features are selected based on the intrinsic characteristics which determine their relevance or discriminant powers with regard to the target classes.

In wrapper type methods, feature selection is "wrapped" around a learning method: the usefulness of a feature is directly judged by the estimated accuracy of the learning method. One can often obtain a set with a small number of non-redundant features, which gives high prediction accuracy, because the characteristics of the features match well with the characteristics of the learning method. Wrapper methods typically require extensive computation to search the best features. Wrapper type methods are used here in this project. In this project seven features like mean, variance, entropy and wavelet approximation coefficients at 4-levels are used.

1.5 CLASSIFICATION

Classification of selected features is the next stage in the process. Here we use Singular Value Decomposition (SVD) and Principle Component Analysis (PCA) for the classification purpose. SVD and PCA are common techniques for analysis of multivariate data. PCA is a multivariate statistical technique frequently used in exploratory data analysis and for making predictive models.

1.6 EVALUATION MEASURES

The evaluation of the classifier performance is done by using various performance measures like Sensitivity, Selectivity, Average Detection, Perfect Classification, Missed Classification, False Alarm and F-score. The Sensitivity and Specificity specifies the ability of the classifier to classify the data when a correct input is given and the ability of the classifier to classify the objects when a wrong input is given. Perfect Classification shows the number of perfectly classified data and missed classification denotes the number of wrongly classified data. Average detection shows the average performance with respect to the correctly classified data. These measures help in evaluating the

performance of the above mentioned classifiers in cancer detection. Some require considerably more computation or memory than others. Some require a substantial number of training instances to give reliable results. Depending on the situation the user may be willing to accept a lower level of predictive accuracy in order to reduce the run time/memory requirements and/or the number of training instances needed. A more difficult trade-off occurs when the classes are severely unbalanced. The Evaluation measures help in finding the performance of the classifier, so that a specified classifier would be used for a specific problem.

1.7 ORGANIZATION OF THESIS

The Organization of thesis is as follows: Chapter1 already explained introduces the methodologies used in the project based on the problem chosen. Selecting the Region of Interest is described in chapter2 and the features extracted from the selected regions of interest are discussed in Chapter3.The abnormality detection is done by using classification procedure. The classification using SVD and PCA are discussed in Chapter 4. Better the performance better the classification procedure. The Performance Measures and Quality Metrics used to measure the classifier performance are compared and analyzed in Chapter5. Chapter 6 concludes the project and briefs the future scope.

CHAPTER 2

SELECTING THE REGION OF INTEREST

2.1. INTRODUCTION

A *region of interest* (ROI) is a portion of an image that is to be filtered or perform some other operation on. Usually ROI is defined by creating a binary mask, which is a binary image that is the same size as the image you want to process with pixels that define the ROI set to 1 and all other pixels set to 0.

Figure.2 Brain image as input

2.2. SUBDIVIDING INPUT IMAGE AND MASKING

Figure.2 shows an input image of brain used in this analysis. As a preprocessing stage this image is divided into four blocks. Then proper mask are assigned to produce the ROI such that only the required region of the input image is taken for analysis. Among the four blocks, the first and third blocks which constitute the left half of the image are taken for further analysis.

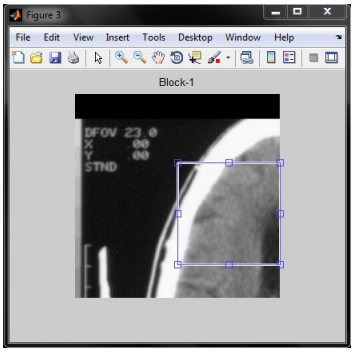

Figure.3.Block-1 of input image

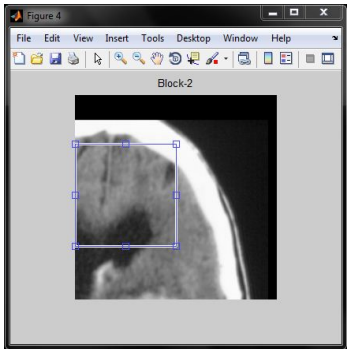

Figure.4.Block-2 of input image

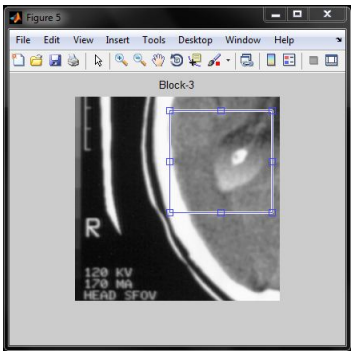

Figure.5.Block-3 of input image

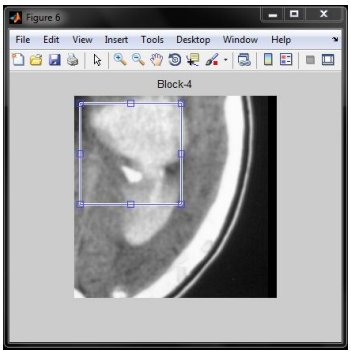

Figure.6.Block-4 of input image

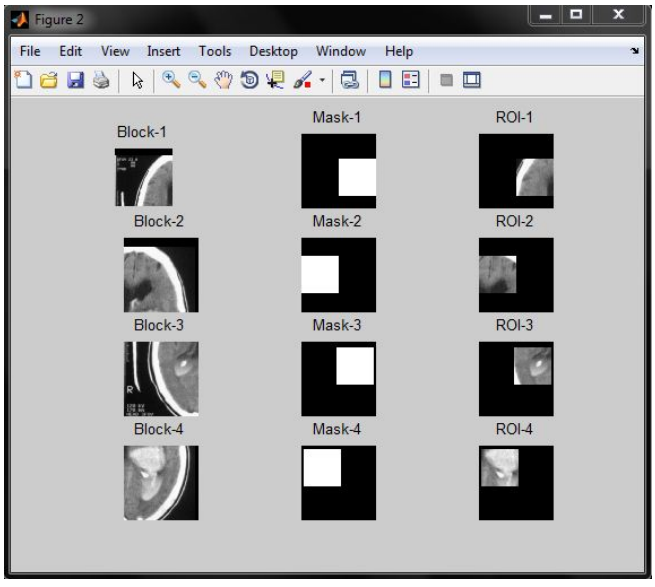

Figure.7 (a). Different ROIs of the input image.

2.3. SUBDIVIDING ROIs

The 1st and 3rd blocks constitute the left half of the input image. These ROIs selected and are subjected to further two level block division to form 16 blocks each. This increase the efficiency of classification as the feature extraction and the analysis of the image can be carried out more deeply. The 32 blocks formed are now used in the feature extraction stage.

2.4 SUMMARY

Selecting proper regions of interest and subdividing those regions improves the overall efficiency of feature extraction and classification. The more detail the analysis, the more efficient is the classification.

CHAPTER 3

FEATURE EXTRACTION OF MEDICAL IMAGES

3.1. INTRODUCTION

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Maximizing the joint dependency with a minimum size of variables is generally the main task of feature selection. For obtaining a minimal subset, while trying to maximize the joint dependency with the target variable, the redundancy among selected variables must be reduced to a minimum. In order to maximize the classification accuracy various types of features like extraction techniques like Mean, Variance, Entropy and also 1^{st} ,2^{nd} ,3^{rd} and 4^{th} level approximation coefficients of wavelet decomposition are used in this project. Brief discussion of these techniques is as follows:

3.2. MEAN, VARIANCE AND ENTROPY

A data-based relative frequency distribution by measures of location and spread, such as the sample mean and sample variance are common approaches in feature selection [12,16,23,24]. Likewise, we have seen how to summarize *probability distribution* of a random variable X by similar measures of location and spread, the mean and variance parameters. For *location* features of a joint distribution, we simply use the means μx and μy of the corresponding *marginal* distributions for X and Y. Likewise, for *spread*

features we use $\sigma^2 X$ and$\sigma^2 Y$. For *joint* distributions, however, we can go further and explore a further type of feature: the manner in which X and Y are *interrelated* or manifest *dependence*.

One way that X and Y can exhibit dependence is to "vary together" – i.e., the distribution p(x, y) might attach *relatively high probability* to pairs (x, y) for which the deviation of x above its mean, $x - \mu x$ and the deviation of y above its mean,$y - \mu y$, are either *both positive* or *both negative* and relatively large in magnitude. Thus, for example, the information that a pair (x, y) had an x with positive deviation $x - \mu x$ would suggest that, unless something unusual had occurred, the y of the given pair also had a positive deviation above its mean. A natural numerical measure which takes account of this type of information is the sum of terms,

$$\sum_x \sum_y (x - \mu x)(y - \mu y) p(x, y) \qquad (3.1)$$

For the kind of dependence just described, this sum would tend to be dominated by large *positive* terms. *Another way* that X and Y could exhibit dependence is to "vary oppositely," in which case pairs (x, y) such that one of $x - \mu x$ and $y - \mu y$ is positive and the other negative would receive relatively high probability [29, 30]. In this case the sum would tend to be dominated by *negative* terms.

Image entropy is a quantity which is used to describe the amount of information which must be coded for by a compression algorithm. Low entropy images, such as those containing a lot of black shades, have very little contrast and large runs of pixels with the same or similar DN values. An image that is perfectly flat will have entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily scattered areas on the medical image have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images. The equation for entropy of an image is as follows:

$$\text{Entropy} = \sum_i P_i Log_2 P_i \qquad (3.2)$$

3.3. 4-LEVEL APPROXIMATION COEFFICIENTS

For many images the low-frequency content is the most important part. It is what gives the signal its identity. In wavelet analysis, we often speak of *approximations* and *details*. The approximations are the high-scale, low-frequency components of the signal. The details are the low-scale, high-frequency components. The wavelet decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the *wavelet decomposition tree*.

Since the analysis process is iterative, in theory it can be continued indefinitely. In reality, the decomposition can proceed only until the individual details consist of a single sample or pixel. In practice, you'll select a suitable number of levels based on the nature of the signal, or on a suitable criterion such as *entropy*. The figure below shows the general decomposition tree of a input signal or image into approximation coefficients and detailed coefficients.

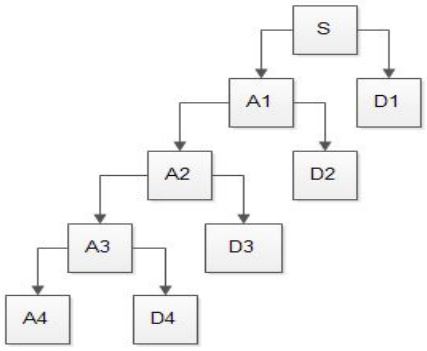

Figure.7 (b).Wavelet decomposition tree

Features extracted for a CT image of brain is shown in Table3.1.

Table.3.1. Features extracted from a CT image of brain

	Entropy	Mean	Variance	a1-	a2-	a3-	a4-
B111	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B112	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B113	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B114	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B121	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B122	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B123	0.729923	0.543557	0.55817	0.399843	0.409285	0.3285	0.399843
B124	0.793336	0.556572	0.952336	0.384302	0.396509	0.335826	0.384302
B131	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B132	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B133	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B134	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B141	0.1	0.968559	0.45473	0.606525	0.523286	0.475444	0.606525
B142	0.1	0.789909	0.308724	0.406239	0.48689	0.534266	0.406239
B143	0.1	0.477827	0.59046	0.364095	0.370501	0.301048	0.364095
B144	0.1	0.435744	0.610952	0.388317	0.412728	0.364816	0.388317
B311	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B312	0.22186	0.119307	0.134632	0.201685	0.232456	0.239458	0.201685
B313	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B314	0.270159	0.143323	0.126371	0.23505	0.285099	0.297094	0.23505
B321	0.901402	0.820321	1	0.460386	0.367618	0.380419	0.460386
B322	0.75393	0.1	0.1	0.557526	0.489772	0.488472	0.557526
B323	1	1	0.63609	0.544641	0.508358	0.468258	0.544641
B324	0.958038	0.918734	0.43056	0.564465	0.557192	0.532271	0.564465
B331	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B332	0.165117	0.124895	0.131944	0.217163	0.289592	0.253309	0.217163
B333	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B334	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B341	0.459728	0.272818	0.456903	0.246425	0.329712	0.3477	0.246425
B342	0.366465	0.443623	0.443643	0.333594	0.396227	0.397834	0.333594
B343	0.1	0.1	0.1	0.1	0.1	0.1	0.1
B344	0.1	0.1	0.1	0.1	0.1	0.1	0.1

3.4 SUMMARY

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. In this section the different features like mean, variance, entropy and four levels of approximation coefficients are explained. These seven features are selected for each block of the images which will yield a 32 x7 feature matrix. This is used for classifying the input image in to normal or abnormal.

CHAPTER 4

FEATURE CLASSIFICATION USING SVD

4.1. INTRODUCTION

Automatic medical image classification is a technique for assigning a medical image to a class among a number of images categorizes. Due to computational complexity its important task in the content-based image retrieval. Various medical image retrieval systems are available today that classify image according to image modalities, orientations, body part or disease. The need for systems that can store, represent and provide efficient retrieval facilities of images CBIR for medical image database dose not aim to replace the physician by predicting the disease of a particular case but to assist the doctor [1,10]. Automatic medical image classification is a technique for assigning a medical image to a class among a number of images categorizes. Due to computational complexity its important task in the content-based image retrieval. Various medical image retrieval systems are available today that classify image according to image modalities, orientations, body part or disease. The need for systems that can store, represent and provide efficient retrieval facilities of images CBIR for medical image database dose not aim to replace the physician by predicting the disease of a particular case but to assist the doctor [1,10].

4.2. SINGULAR VALUE DECOMPOSITION (SVD)

A mathematical approach that directly reveals the rank and corresponding ideal basis of a dataset is the singular value decomposition (SVD). For a dataset in n dimensional space, for any $k < n$, the SVD will show the ideal basis for representing that data using only k dimensions [6]. If the SVD reveals that the dataset is full rank and no feature reduction is possible along the calculated axes, then no axes exist for which a reduction is

possible. Furthermore, if no reduction is possible, this will be shown by the magnitudes of the singular values revealed by the SVD. An operation such as a classification that would be performed on the entire $m \times n$ matrix A can now be equivalently performed on the entire $k \times n$ matrix where $k < m$, resulting in a reduction in the number of bands present in each vector.

For practical purposes, singular values may in fact be nonzero yet be sufficiently close to zero to reduce the dimension of the data. The singular values $\sigma_{k+1}, \dots, \sigma_m$ represent distances from the subspace spanned by $U_{.1}, \dots, U_{.k}$ and very small distances may not affect the operation that will be performed on the reduced data, such as classification. If none of the singular values on the diagonal are close to zero, then the data is already represented using as few dimensions as possible. Practically speaking, it would be necessary to think of a three dimensional (pixel row, pixel column, data bands) image in two dimensions in order to take advantage of the feature reduction made possible by using the SVD. This would also be an expensive computation as the leading dimension of the matrix would be the number of bands, but the second dimension would be equal to the number of pixels in the image. One of the features of an SVD is that it reveals the basis vectors U that can be used to transform any vector from the original vector space (range of A) to the new vector space (range of U∑). Using this result, it is possible to perform the SVD on a training data matrix that is m × p in dimension, where p is the number of points in the training data set, and use the resulting SVD to transform A. In order to do this, it is necessary to project the columns of the matrix A onto the subspace spanned by the first k columns of U. After the original image has been transformed and the number of bands has been reduced classification algorithm can be applied.

4.3. CLASSIFICATION USING SVD

SVD values are calculated for both Brain and Skull images with 30 images each. A threshold is set based upon which the classification is done. Here the threshold set for Brain images is 4.85 and that for the skull images is 2.48.

Table.4.1. Classification results using SVD

Image No	SVD for Brain images	Classification for Threshold= 4.85	SVD for Skull image	Classification for Threshold= 2.49
Image 1	4.8500	Normal	2.5208	Abnormal
Image2	5.1120	Abnormal	2.5129	Abnormal
Image3	5.0598	Abnormal	2.5104	Abnormal
Image4	4.7051	Normal	2.4965	Abnormal
Image5	4.9946	Abnormal	2.4991	Abnormal
Image6	5.0777	Abnormal	2.4980	Abnormal
Image7	4.9150	Abnormal	2.4855	Normal
Image 8	4.7831	Normal	2.4826	Normal
Image9	5.1058	Abnormal	2.4855	Normal
Image10	4.9837	Abnormal	2.4776	Normal
Image11	4.8131	Normal	2.4807	Normal
Image12	4.7838	Normal	2.4988	Abnormal
Image13	4.6963	Normal	2.5358	Abnormal
Image14	4.7469	Normal	2.5176	Abnormal
Image 15	4.7070	Normal	2.4201	Normal
Image16	5.0921	Abnormal	2.5960	Abnormal
Image17	4.6065	Normal	2.5131	Abnormal
Image18	4.9947	Abnormal	2.4998	Abnormal
Image19	5.0429	Abnormal	2.4756	Normal

Image20	4.9220	Abnormal	2.4807	Normal
Image21	4.9401	Abnormal	2.4657	Normal
Image22	5.1634	Abnormal	2.4720	Normal
Image23	5.0597	Abnormal	2.4808	Normal
Image24	5.0429	Abnormal	2.4653	Normal
Image25	4.9401	Abnormal	2.4892	Normal
I Image26	548641	Abnormal	2.5036	Abnormal
Image27	5.0081	Abnormal	2.4943	Abnormal
Image28	5.0183	Abnormal	2.4631	Normal
Image29	4.8714	Abnormal	2.4960	Normal
Image30	4.8538	Abnormal	2.4606	Normal

4.5. PRINCIPLE COMPONENT ANALYSIS (PCA)

Conventional Principle Component Analysis (PCA) is one of the most commonly used feature extraction techniques. It is based on extracting the axes on which data shows the highest variability [8]. Although PCA "spreads out" data in the new basis, and can be of great help in unsupervised learning, there is no guarantee that the new axes are consistent with the discriminatory features in a classification problem. Another approach is to account class information during the feature extraction process. In parametric and nonparametric eigenvector-based approaches that use the within- and between-class covariance matrices and thus do take into account the class information were analyzed and compared. Both parametric and nonparametric approaches use the simultaneous diagonalization algorithm to optimize the relation between the within- and between-class covariance matrices.

An important issue is how to decide whether a PCA-based feature transformation approach is appropriate for a certain problem or not. Since the main goal of PCA is to extract new uncorrelated features, it is logical to introduce some correlation-based

criterion with a possibility to define a threshold value. In this chapter we are interested in the problem of selecting the best subset of orthogonally transformed features for subsequent classification, i.e. we are not searching for the best transformation but rather try to find the best subset of transformed components, which allow achieving the best classification.

4.4. MINIMUM AND MAXIMUM VALUES

The Maximum and Minimum values aims to classify data instances directly without filling hypothetical missing values, and is therefore different from imputation. Our interpretation is that each data instance resides in a lower dimensional subspace of the feature space, determined by its own existing features [13]. To formulate the classification problem we go back to the geometric intuitions underlying max-margin classification: We try to maximize the worst-case margin of the separating hyper plane, while measuring the margin of each data instance in its own lower-dimensional subspace.

4.5. RESULTS AND DISCUSSION

The results of classification based on the SVD and PCA are as shown in the tables. Based on the classification, a confusion matrix is formed for each of the classifiers.

Table.4.2. Classification results using PCA (max-max)

Image No	PCA (max-max) Brain images	Classification for Threshold= 0.8	PCA (max-max) Skull image	Classification for Threshold= 0.74
Image 1	0.7914	Normal	0.7071	Normal
Image2	0.8358	Abnormal	0.7071	Normal
Image3	0.7925	Normal	0.8727	Abnormal
Image4	0.7989	Normal	0.7202	Normal
Image5	0.7071	Normal	0.7071	Normal

Image6	0.7308	Normal	0.8899	Abnormal
Image7	0.7553	Normal	0.7607	Abnormal
Image 8	0.7661	Normal	0.7406	Abnormal
Image9	0.8327	Abnormal	0.8245	Abnormal
Image10	0.8362	Abnormal	0.7071	Normal
Image11	0.7149	Normal	0.7393	Normal
Image12	0.7878	Normal	0.7648	Abnormal
Image13	0.8443	Abnormal	0.7710	Abnormal
Image14	0.9038	Normal	0.7071	Normal
Image 15	0.7733	Normal	0.7327	Normal
Image16	0.8283	Abnormal	0.7071	Normal
Image17	0.8432	Abnormal	0.7934	Abnormal
Image18	0.7735	Normal	0.7350	Normal
Image19	0.8327	Abnormal	0.7546	Abnormal
Image20	0.8283	Abnormal	0.7315	Normal
Image21	0.8227	Abnormal	0.7421	Abnormal
Image22	0.7071	Normal	0.7564	Abnormal
Image23	0.8874	Abnormal	0.8091	Abnormal
Image24	0.8489	Abnormal	0.8564	Abnormal
Image25	0.8760	Abnormal	0.7767	Abnormal
I Image26	0.7071	Normal	0.8263	Abnormal
Image27	0.7071	Normal	0.7869	Abnormal
Image28	0.7126	Normal	0.8180	Abnormal
Image29	0.7643	Normal	0.7772	Abnormal
Image30	0.7812	Normal	0.7667	Abnormal

Table.4.3. Classification results using PCA (max-min)

Image No	PCA (max-min) Brain images	Classification for Threshold= 0.28	PCA (max-min) Skull image	Classification for Threshold= 0.36
Image 1	0.3098	Abnormal	0.3930	Abnormal
Image2	0.3055	Abnormal	0.3753	Abnormal
Image3	0.3035	Abnormal	0.4089	Abnormal
Image4	0.2963	Abnormal	0.2904	Normal
Image5	0.2506	Normal	0.2431	Normal
Image6	0.2302	Normal	0.4123	Abnormal
Image7	0.3845	Abnormal	0.3230	Normal
Image 8	0.2707	Normal	0.3312	Normal
Image9	0.2440	Normal	0.3565	Normal
Image10	0.1978	Normal	0.3616	Abnormal
Image11	0.3674	Abnormal	0.3376	Normal
Image12	0.2271	Normal	0.4018	Abnormal
Image13	0.3021	Abnormal	0.3937	Abnormal
Image14	0.3591	Abnormal	0.3145	Normal
Image 15	0.2889	Abnormal	0.2713	Normal
Image16	0.2807	Abnormal	0.2335	Normal
Image17	0.2634	Normal	0.3049	Normal
Image18	0.2616	Normal	0.2162	Normal
Image19	0.2440	Normal	0.2207	Normal
Image20	0.2807	Abnormal	0.2104	Normal
Image21	0.2175	Normal	0.1844	Normal
Image22	0.2979	Abnormal	0.2456	Normal
Image23	0.3232	Abnormal	0.2837	Normal
Image24	0.3162	Abnormal	0.2736	Normal

Image25	0.2729	Abnormal	0.3224	Normal
Image26	0.3628	Abnormal	0.2961	Normal
Image27	0.2955	Abnormal	0.2906	Normal
Image28	0.2839	Abnormal	0.2642	Normal
Image29	0.2855	Abnormal	0.3155	Normal
Image30	0.3144	Abnormal	0.2545	Normal

4.5. CONFUSION MATRIX

To assess the accuracy of an image classification, it is common practice to create a confusion matrix. In a confusion matrix, the classification results are compared to additional ground truth information. The strength of a confusion matrix is that it identifies the nature of the classification errors, as well as their quantities. The results of confusion matrix highly depend on the selection of ground truth / test set pixels.

In the field of machine learning, a confusion matrix, also known as a contingency table or an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another).

Table 4.4.: Confusion Matrix of SVD Classification for CT images of Brain

No Of Input Images 30 (Normal-15,Abnormal-15)

ACTUAL CLASSES	*PREDICTED CLASSES*		
		Normal	Abnormal
	Normal	8	7
	Abnormal	1	14

Table 4.5. Confusion Matrix of SVD Classification for CT images of Skull

No Of Input Images 30 (Normal-15,Abnormal-15)

ACTUAL CLASSES	*PREDICTED CLASSES*		
		Normal	Abnormal
	Normal	7	8
	Abnormal	4	11

Table4.6. Confusion Matrix of PCA (max-max value) Classification for CT images of Brain

No Of Input Images 30 (Normal-15,Abnormal-15)

ACTUAL CLASSES	*PREDICTED CLASSES*		
		Normal	Abnormal
	Normal	8	7
	Abnormal	3	12

Table4.7. Confusion Matrix of PCA (max-max value) Classification for CT images of Skull

No Of Input Images 30 (Normal-15,Abnormal-15)

ACTUAL CLASSES	*PREDICTED CLASSES*		
		Normal	Abnormal
	Normal	8	7
	Abnormal	3	12

Table4.8: Confusion Matrix of PCA (max-min value) Classification for CT images of Brain

No Of Input Images 30 (Normal-15,Abnormal-15)

ACTUAL CLASSES	*PREDICTED CLASSES*		
		Normal	Abnormal
	Normal	6	9
	Abnormal	5	10

Table4.9: Confusion Matrix of PCA (max-min value) Classification for CT images of Skull

No Of Input Images 30 (Normal-15,Abnormal-15)

ACTUAL CLASSES	*PREDICTED CLASSES*		
		Normal	Abnormal
	Normal	8	7
	Abnormal	6	9

4.6. SUMMARY

Confusion matrix or contingency table is used to compare the results of classification with the ground truth. Thus from a confusion matrix, the classification error rate can be well understood. Based on the confusion matrix the performance analysis can be done, which is explained in the next chapter.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 INTRODUCTION

The experimental results are taken for CT images of brain and skul, 30 images each. In order to study the performance of these classifiers in cancer detection various classifier performance measures like perfect classification, missed classification, false alarm, sensitivity, specificity, average detection, f-score are used.

5.2. PERFORMANCE QUALITY MEASURES

5.2.1 Perfect Classification

Perfect Classification is the ability of the classifier to identify the data correctly. The error rate will be totally zero, where no true negatives and false positives are found [26,27]. This can't be fully 1 because there will be some errors called bayes error resulting in wrong classification of data. The equation used for measuring perfect classification is given below:

$$\text{Perfect Classification} = \frac{\text{No of TP+No of TN}}{\text{Total no of inputs}} \tag{5.3}$$

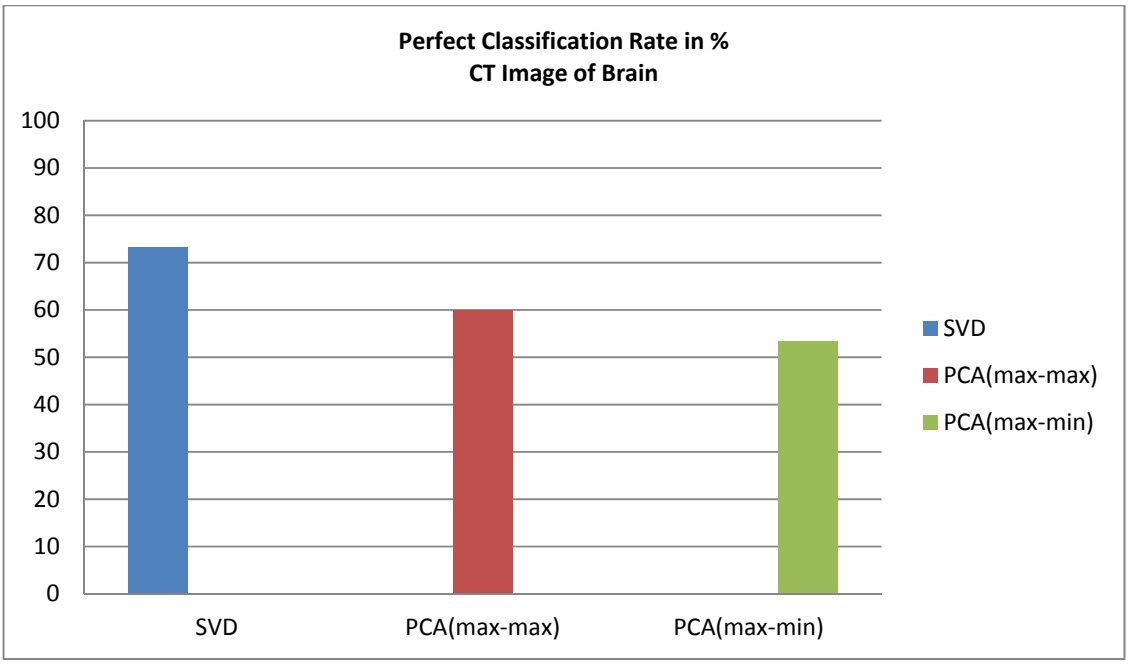

Figure.8. Perfect Classification Rate of CT image of Brain.

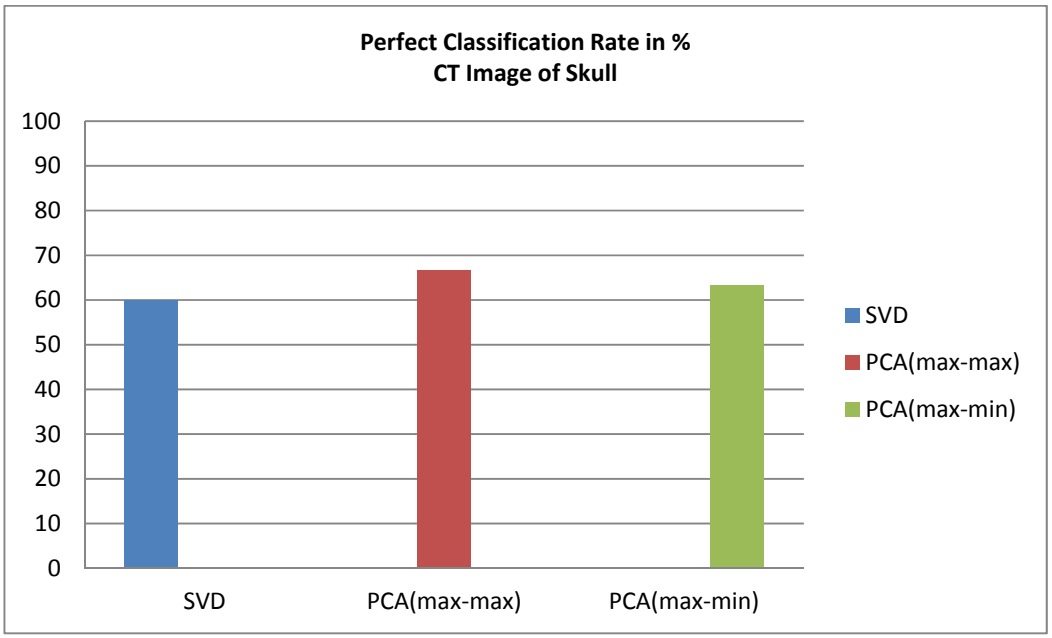

Figure.9. Perfect Classification Rate of CT image of Skull.

5.2.2 Missed Classification

Missed Classification is opposite to the perfect classification [23, 24]. Missed classification is the total number of misclassified data with respect to the total number of inputs given. Equation for missed classification is given below:

$$\text{Missed Classification} = \frac{\text{No of TN+No of FP}}{\text{Total no of inputs}} \tag{5.4}$$

5.2.3 False Alarm

False alrm is similar to missed classification. It is the total number of normal images that are classified abnormal a with respect to the total number of inputs given. Equation for false alarm is given below:

$$\text{False Alarm} = \frac{\text{No. FP}}{\text{Total number of Inputs}} \tag{5.5}$$

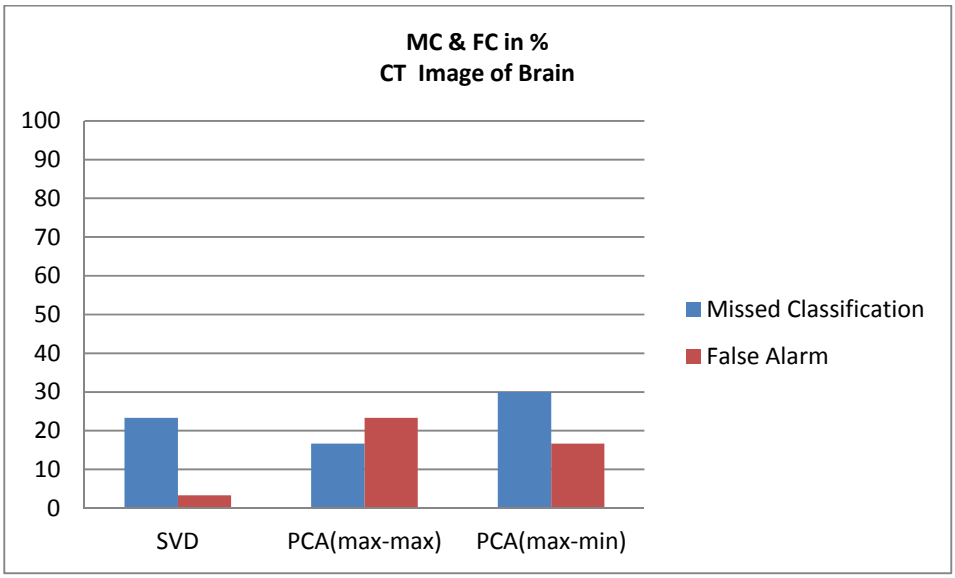

Figure.10.Missed Classification and False Alarm Rate of CT image of Brain

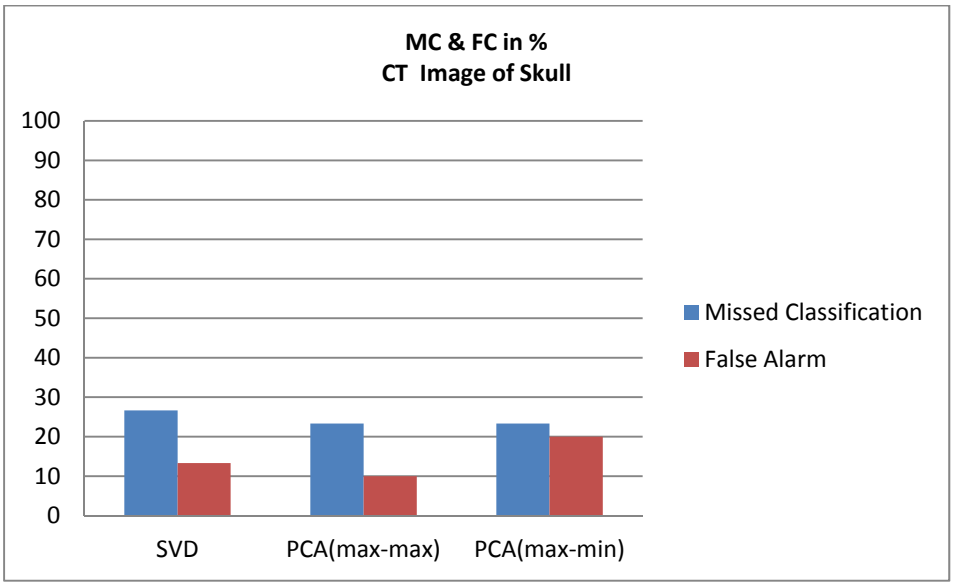

Figure.11.Missed Classification Rate of CT image of Skull

5.5.4 Sensitivity

Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as classification function. Sensitivity (also called the *true positive rate*, or the recall rate in some fields) measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition) [29, 30]. Specificity measures the proportion of negatives which are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition, sometimes called the *true negative rate*). These two measures are closely related to the concepts of type I and type II errors. A perfect predictor would be described as 100% sensitive (i.e. predicting all people from the sick group as sick) and 100% specific (i.e. not predicting anyone from the healthy group as sick); however, theoretically any predictor will possess a minimum error bound known as the Bayes error rate. *Sensitivity relates to the test's ability to identify positive results.*

$$Sensitivity = \frac{No\ of\ TP}{\substack{No\ of\ TP+No\ of\ FN \\ =probablity\ of\ a\ negative\ test,given\ that\ patient\ is\ ill}} \tag{5.1}$$

5.5.5 Specificity

Specificity relates to the test's ability to identify negative results. However, highly specific tests rarely miss negative outcomes, so they can be considered reliable when their result is *positive*. Therefore, a positive result from a test with high specificity means a high probability of the presence of disease [26,21]. A test with a high specificity has a low type I error rate.

$$Specificity = \frac{No\ of\ TN}{\substack{No\ of\ TN+No\ of\ FP \\ =probablity\ of\ a\ negative\ test,given\ that\ patient\ is\ well}} \tag{5.2}$$

5.5.6. Average Detection

The missed classification is calculated using sensitivity and specificity. The equation for Average Detection is given as:

$$\text{Average Detection} = \frac{\text{Sensitivit y} + \text{Specificit y}}{2} \tag{5.5}$$

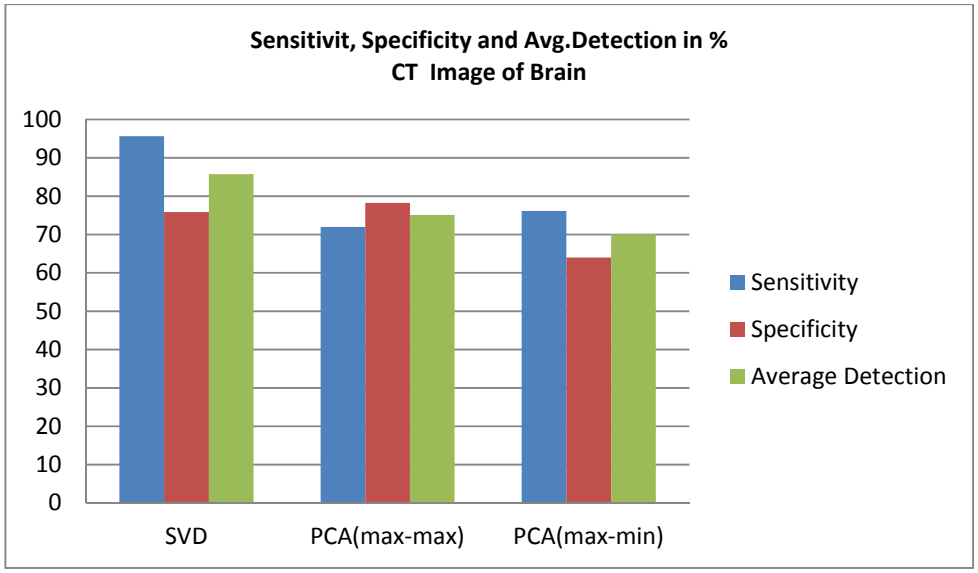

Figure.12. Sensitivity, Specificity and Average Detection of CT image of Brain

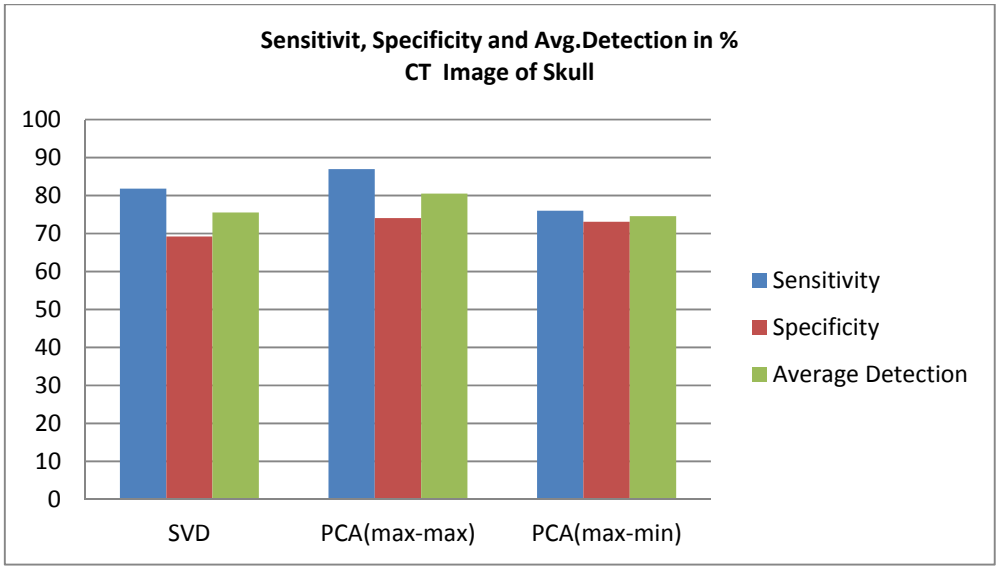

Figure.13. Sensitivity, Specificity and Average Detection of CT image of Brain

5.5.5. F-Score

F-score is another parameter which can be used to analyze the quality of classification. The F-Score is calculated by using the equation.

$$F\_score = \frac{\text{Sensitivity X Specificity}}{\text{Average Detection}} \tag{5.6}$$

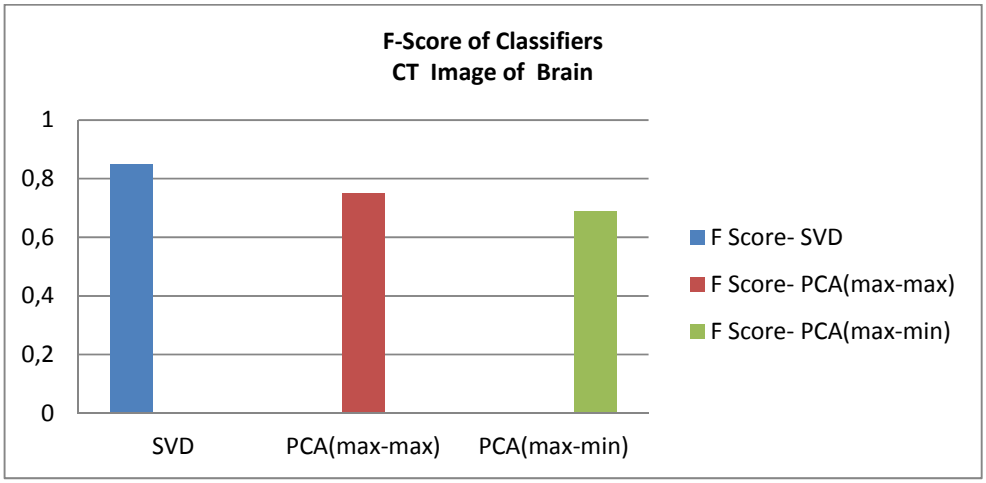

Figure.13.F-Score of CT Image of Brain

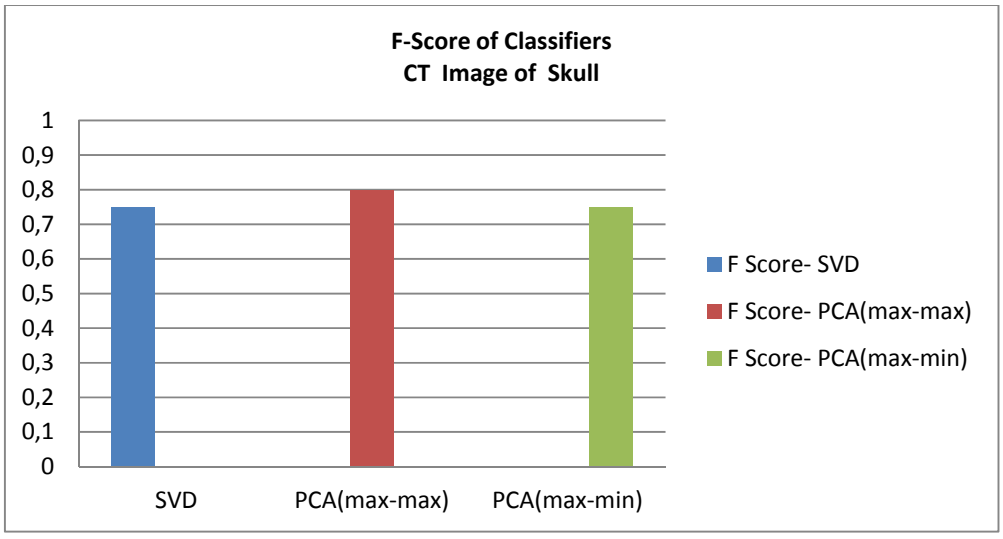

Figure.14.F-Score of CT Image of Skull

5.3 CONCLUSION

The performance analysis was done using various classifier performance measures and Quality Metrics. The results are compared for the three classifier techniques used namely: SVD and PCA. The results shows that both the SVM and PCA classifiers performs good with less number of missed classifications and False alarms .

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

Conclusion

Performance analysis of SVD and PCA classifiers is done in CT images and is analyzed using various performance measures and quality metrics like perfect classification, missed classification, false alarm, sensitivity, specificity, average detection, f-score.

Future Scope

The whole project is based on the analysis of general abnormality detection and classification process. This could be extended to the specified process concerned over a specific region or specific type of cancer like mammography or brain tumor detection. Every cancer detection process involves more details with improved accuracy. This could be found out for specified type of problems