

Cephalometric Analysis using PCA and SVM

M.Arulselvi
Annamalai University
Dept. of Computer Science and
Engineering,
Chidambaram – 608001

V. Ramalingam
Annamalai University
Dept. of Computer Science and
Engineering
Chidambaram – 608001

S. Palanivel
Annamalai University
Dept. of Computer Science and
Engineering
Chidambaram – 608001

ABSTRACT

Cephalometric analysis is the study of dental and skeletal relationship in the head. It depends on cephalometric radiography to study relationships between bony and soft tissue landmarks and can be used for diagnosis of facial growth abnormalities prior to treatment. Skeleton analysis consists of facial skeleton analysis, and mandibular and maxillary base analysis. In this work, landmarks needed for detecting skeletal abnormalities are selected from the digital image; Principal Component Analysis (PCA) is applied to the digital image for dimension reduction to get the desired feature vectors. The normalized feature vectors are trained and tested using an SVM classifier to detect the skeletal abnormalities. The performance measure such as accuracy, sensitivity and specificity were evaluated and the results are found to be satisfactory.

Keywords

Cephalometric analysis, Principal component analysis, Support Vector machine, Saddle, Gonial, Articular, subspinale, supramentale, nasion, sella.

1. INTRODUCTION

Cephalometric analysis is the study of dental and skeletal relationships of the head. It is frequently used by dentists, often orthodontists [1] in particular, as a treatment-planning tool. Cephalometrics helps in orthodontic diagnosis by enabling the study of skeletal, dental and soft tissue structures of the craniofacial region. It helps in classification of the skeletal and dental abnormalities, establishing facial type, planning treatment for an individual, predicting the growth related changes and changes associated with surgical treatment.

Cephalometric analysis has become an important clinical tool in diagnosis, treatment planning, evaluation of growth, or treatment results and research. Cephalometric analysis depends on cephalometric radiography to study relationships between bony and soft tissue landmarks and can be used to diagnose facial growth abnormalities prior to treatment. Skeleton Analysis consists of facial skeleton analysis, and mandibular and maxillary base analysis.

Cephalometric analysis requires an expert system to be developed for computer applications in this field. One can provide a computerized analysis that will measure and compare the anatomy to assist in the treatment plan. So a preliminary research is necessary using cephalometric for making the computer detect the skeletal abnormalities. There is lack of skilled personnel proficient enough in cephalometric analysis and also to assist the doctor. Hence there is a growing need for

the computer as an aid for cephalometric analysis. Calibrating cephalogram is a necessary step in making the computers to detect the skeletal abnormalities.

In this respect, automatic cephalometric analysis is one of the main goals, to be reached in orthodontics in the near future. Accordingly, several efforts have been made to automate assessment of cephalometric analysis.

For automatic assessment, Principal Component Analysis (PCA) is used for dimension reduction. Then dimension reduction data are normalized which are taken as the features. The normalized parameters are used for classification and an SVM classifier is used to classify the skeletal abnormalities.

1.1 Literature review

Traditionally the location of landmarks for a cephalometric analysis was performed manually using a tracing from X-ray film. More recently the film was digitized and semi-automatic systems allowed the orthodontics to plot landmarks directly on the screen. Cephalometric analysis of lateral radiography of the head is an important tool in orthodontics. Manually locating specific landmarks is a tedious, time consuming and error prone task. In the existing system, an automated system based on the use of Active Appearance Models (AAMs) was proposed by Sylvia Rueda and Marino Alca'niz [2].

Two approaches may be used to perform a cephalometric analysis: a manual approach, and a computer-aided approach. The manual approach is the oldest and most widely used. It consists of placing a sheet of acetate over the cephalometric radiograph, tracing salient features, identifying landmarks, and measuring distances and angles between landmark locations [3].

The projection of the software was based on the correct interpretation of the analysis to reach the goals of the Treatment plan. The Orthodontic Folder collects easy-to-complete and consult schemes of essential data, aided by sophisticated graphics that the modern informatic systems can supply. The orthodontist can therefore make use of, in a simple and rapid way, of the consultation to compare the results achieved. The method is commonly called "Orthodontic check-up"[4]. In general, automatic identification of landmarks approaches can be classified into four categories: (i) image filtering plus knowledge-based landmark search [5] (ii) model-based approaches [5], [6], (iii) soft-computing approaches [7] and (iv) hybrid approaches [8].

Data analysis problems, where the data objects have a large number of features are becoming more prevalent in areas such as multimedia data analysis and bioinformatics. In these situations

it is often beneficial to reduce the dimension of the data in order to improve the efficiency and accuracy of data analysis. Statisticians sometimes talk of problems that are "Big p Small n"; these are extreme examples of situations where dimension reduction (DR) is necessary because the number of explanatory variables p exceeds (sometimes greatly exceeds) the number of samples n [9].

In theory the number of examples needs to increase exponentially with the number of features if inference is to be made about the data. In practice this is not the case as real high-dimension data will only occupy a manifold in the input space, so the implicit dimension of the data will be less than the number of features p . Nevertheless, traditional algorithms used in machine learning and pattern recognition applications are often susceptible to the well-known problem of the curse of dimensionality [10][11].

Principal Component Analysis (PCA) was first introduced by Pearson [12] in 1901 and later independently developed by Hotelling [13] in 1933, where the name principle component first appears. In various fields, it is also known as the singular value decomposition (SVD), the Karhunen-Loeve transform, the Hotelling transform, and the empirical orthogonal function (EOF) method [14].

A Support Vector Machine (SVM) performs classification by constructing an N -dimensional hyperplane that optimally separates the data into two categories. SVM models are closely related to neural networks. In fact, a SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. Support Vector Machine (SVM) models are a close cousin to classical multilayer perceptron neural networks. Using a kernel function, SVM's are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training [15].

1.2 Outline of the work

Section 2 deals with the cephalometric landmarks used in this work. Section 3 explains the need for dimension reduction and PCA approach. Section 4 describes the SVM technique. Section 5 explains the proposed method. Section 6 deals with the Experimental results and Section 7 concludes with conclusion.

2. CEPHALOMETRIC LANDMARKS

Cephalometric makes use of certain landmarks or points on the skull, which are used for quantitative analysis and measurements are shown in Fig 1. These landmarks represent actual anatomic structures of the skull and are obtained secondarily from anatomic structures in a cephalogram. The landmarks should also permit valid quantitative measurements of lines and angles projected from them.

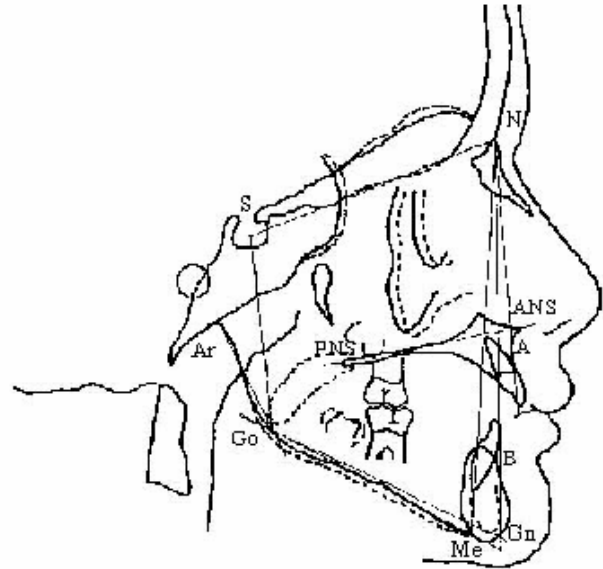


Fig 1: Cephalometric Landmarks

The landmarks used in this work are:

Nasion (N): The most anterior point midway between the frontal and nasal bones on the fronto-nasal suture.

Sella (S): The point representing the midpoint of the pituitary fossa or sella turcica. It is a constructed point in the mid-sagittal plane.

Articulare (Ar): This point was introduced by Bjork (1947). It provides radiological orientation, being the point of intersection of the posterior margin of the ascending ramus and the outer margin of the cranial base.

Gonion (Go): A constructed point, the intersection of the lines tangent to the posterior margin of the ascending ramus and the mandibular base.

Menton (Me): According to Krogman and Sassouni, Menton is the most caudal point in the outline of the symphysis; it is regarded as the lowest point of the mandible and corresponds to the anthropological gnathion.

Point A, subspinale (A): The deepest midline point in the curved bony outline from the base to the alveolar process of the maxilla, i.e. at the deepest point between the anterior nasal spine and prosthion. In anthropology, it is known as subspinale.

Point B, supramentale (B): Most anterior part of the mandibular base. It is the most posterior point in the outer contour of the mandibular alveolar process, in the median plane. In anthropology, it is known as supramentale, between infradentale and pogonion.

3. DIMENSION REDUCTION

Developments in data collection and storage techniques have led to information overload. Researchers working on diversified domains come across a high dimensional datasets which present many mathematical challenges to handle due to curse of dimensionality [10].

Due to the increases in dimensionality of the dataset, its performance decreases and if the dimensionality of the input space is higher, more feature vectors are needed for training. The major problem with these high dimensional data sets is that all the measured data are not important for understanding the underlying phenomena of interest. Thus, for effective analysis of high dimensional data sets dimension reduction is necessary.

According to statistical point of view it is desirable that the number of examples in the training set should significantly exceed the number of features used to describe those examples which is shown in Fig 2(a). For this reason data sets as depicted in Fig 2(b) can still be analyzed [11]. However, traditional algorithm used in machine learning and pattern recognition applications are often susceptible to the well-known problem of the curse of dimensionality, which refers to the degradation in the performance of a given learning algorithm as the number of feature increases. To cope with this, dimension reduction techniques are often applied as a data pre-processing step or as part of the data analysis to simplify the data model.

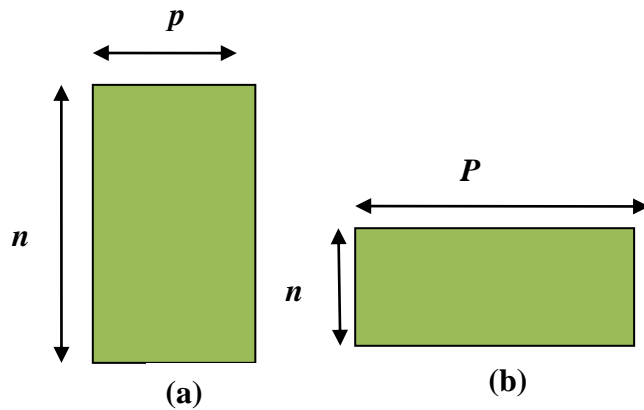


Fig 2: Big p small n problems are problems where the number of features in a dataset is large compared with the number of objects: (a) This is how statisticians would like your data to look, (b) This is what we call the big p small n problem''.

Dimension reduction technique involves the identification of a suitable low-dimensional representation for the original high-dimensional data set. By working with this reduced representation, tasks such as classification or clustering can often yield more accurate and readily interpretable results, while computational costs may also be significantly reduced.

The inspiration for dimension reduction can be summarised as follows:

- The identification of a reduced set of features that are predictive of outcomes can be very useful from a knowledge discovery perspective.
- For many learning algorithms, the training and/or classification time increases directly with the number of features.

- Noisy or irrelevant features can have the same influence on classification as predictive features so they will impact negatively on accuracy.

3.1 Principal Component Analysis

Principle Component Analysis (PCA) was first introduced by Pearson in 1901[12] and later independently developed by Hotelling in 1933, where the name principle component first appears. In various fields, it is also known as the singular value decomposition (SVD), the Karhunen-Loeve transform, the Hotelling transform, and the empirical orthogonal function (EOF) method.

PCA is a well known scheme for dimension reduction. PCA finds a set of most representative projection vectors, and the most relevant information about original dataset. PCA is a useful statistical technique that has found application in many fields such as video/audio classification, face recognition, image compression etc., It is a simple method of extracting relevant information from high dimensional data sets. With minimal effort, PCA provides a road map for reducing the complex data set to a lower dimensional dataset.

PCA is a useful statistical procedure that has found importance in many fields, and is a well-known technique for finding patterns in high dimensional data. It is a way of identifying patterns in data, and expressing the data in such a way to highlight their similarities and differences. The other main advantage of PCA is that once these patterns are found, the data can be compressed by reducing the number of dimensions, without much loss of information. Thus, PCA ‘combines’ the essence of attributes by creating an alternative, smaller set of variables.

This section explains the steps to perform principle components analysis on a set of data.

Approach:

- (i) Dataset

Table 1: PCA approach example data.

Items	X	Y
1	25	21
2	50	71
3	24	29
4	45	64
5	32	30
6	27	48
7	4	17
8	2	12
9	12	18
10	55	59

(ii) Subtract the mean from the data items

The next step is to subtract the mean from each of the data item. This is the average across each dimension. Thus, all the data items in each dimension will have its mean subtracted. This new dataset will thus have a mean whose value is 'zero'. The mean subtracted data of the dataset shown in Table 2.

Table 2 : PCA mean subtracted data

Items	X	Y
1	-2.6	-15.9
2	22.4	34.1
3	-3.6	-7.9
4	17.4	27.1
5	4.4	-6.9
6	-0.6	11.1
7	-23.6	-19.9
8	-25.6	-24.9
9	-15.6	-18.9
10	27.4	27.4

(iii) Calculate covariance matrix

Covariance matrix is computed for the data in Table 2. The data set is 2-dimensional and hence the covariance matrix will be (2x2). For the dataset under consideration, the covariance matrix is:

$$\begin{pmatrix} 305.04 & 327.56 \\ 327.56 & 424.49 \end{pmatrix}$$

(iv) Calculate eigenvectors and eigenvalues of the covariance matrix:

Eigenvectors can be calculated only for square matrices. Since the covariance matrix is a square matrix, eigenvectors can be calculated. The eigenvalues and eigenvectors are:

$$\text{eigenvalues} = \begin{pmatrix} 31.8046 & 0 \\ 0 & 697.7254 \end{pmatrix}$$

$$\text{eigenvectors} = \begin{pmatrix} -0.7679 & 0.6406 \\ 0.6406 & 0.7679 \end{pmatrix}$$

(v) Sort the eigen values

In this step, the idea of dimension reduction comes in effect. Generally, once the eigen vectors are computed, the eigen values are sorted in descending order, which gives the principle components in the order of significance. Then, the less significant components can be ignored. Because of this, some

information loss can be found, but, if the eigen values are small, the information loss will not be much significant. Thus, the final data will have fewer dimensions than the original data set.

Obviously, if the original dataset is of 'n' dimensions, this given 'n' eigenvalues and eigenvectors. If only k eigenvalues are chosen in the procedure, the final data has only k dimensions, where, $k \leq n$.

Considering the eigen values obtained for the example data shown in the table 1, they are different values. Of course, the eigen vector of the largest eigenvalue, 697.7254 turns out to be the first principle component of the dataset under consideration.

$$\text{eigenvalues} = \begin{pmatrix} 31.8046 & 0 \\ 0 & 697.7254 \end{pmatrix}$$

Thus, for the largest eigen value 697.7254 of the example dataset, the principle component (eigenvector) is

$$\text{eigenvector} = \begin{pmatrix} 0.6406 \\ 0.7679 \end{pmatrix}$$

(vi) Compute the new projected data set

After choosing the set of principle components (eigenvectors), the final data set of reduced dimension is obtained as follows:

$$\text{Final data} = \text{Eigenvector}^T \times \text{Mean subtracted data} \quad (1)$$

Where, Eigenvector^T is the matrix with the eigenvectors in the columns transposed, Final data is the final data set with data items in columns and dimension along rows.

Table 3: PCA projected data with 2 eigenvectors

Items	X	Y
1	-13.8752	-8.1883
2	40.5342	4.6417
3	-8.3725	-2.2959
4	31.9561	3.9974
5	-2.4801	-7.7986
6	8.1395	7.5709
7	-30.3986	5.3756
8	-35.5192	3.7087
9	-24.5062	-0.1271
10	34.5221	-6.8845

Suppose, if it is decided to keep both the principle components (eigenvectors), then the dimension of the Final data will be the same as that of the dimension of the original data set. In the example data set, which is of two-dimension, if both the principle components are kept for obtaining the Final data, then this Final data will also be of same two-dimension. The PCA projected data set, keeping both the dimensions is shown in the table 3. But, if it is decided to leave one component, then the

obtained Final data will be of reduced dimension, i.e., one-dimension.

Final data actually represent the original data purely in terms of the given vectors chosen. The original data set is of two dimension, and hence had two axes x and y. And, any two dimensional data can be represented in terms of any two axes. If these new axes are perpendicular, it will be efficient to represent. The eigen vectors are perpendicular to each other. And hence, the original data set can be represented in terms of the eigen vectors.

If it is decided to keep both the eigenvectors, then the final transformed data is the original data, represented in terms of both the eigenvectors.

4. SUPPORT VECTOR MACHINE (SVM)

SVM is a learning machine technique based on statistical learning theory. It was developed by Cortes and Vapnik in 1995[16] and modified by Vapnik in 1999. Fig 3. gives the basic principle of a SVM. When used in classification SVM maps the input space to higher dimensional feature space and constructs a hyperplane, which separates class members from non-members. The training set consists of skeletal abnormality data belonging to two classes namely normal and abnormal. One class of skeletal abnormality is labeled positive and those which are not in that class are labeled negative. From this data SVM can learn to distinguish between the classes in the training set. They can then classify new test data as being normal or abnormal using the information already learnt in the training phase.

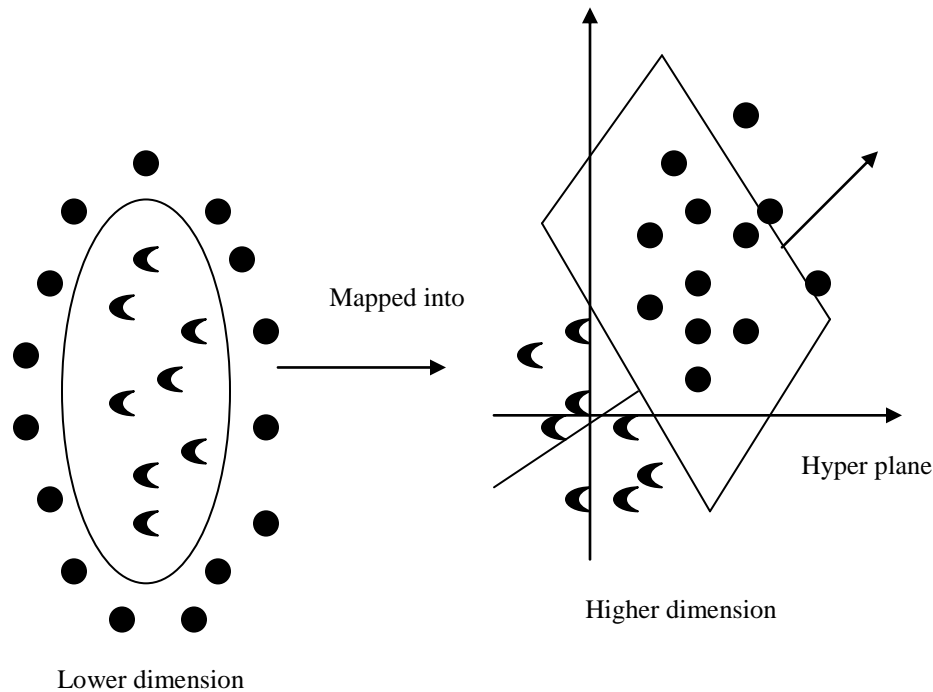


Fig 3: The principle of SVM

When the data is not linearly separable [17], SVM maps the data into a higher dimensional space called a feature space, where a maximal separating hyperplane is constructed. Many hyperplanes separate the data. However, only one achieves maximum separation. Thus a maximum margin hyperplane is the hyperplane that has maximal distance from members to non-members. Two parallel hyperplanes are constructed on each side of the hyperplane that separates the data. The points that lie on these hyperplanes are called support vectors and are shown in

Fig 4. The hyperplane found by an SVM in the feature space corresponds to a decision boundary in the input space.

A classification task usually involves training and testing data which consist of some data instances. Each instance in the training set contains one target value (class labels) and several attributes (features). The goal of SVM is to produce a model which predicts target value of data instances in the testing set which contains only feature attributes but not the class attribute.

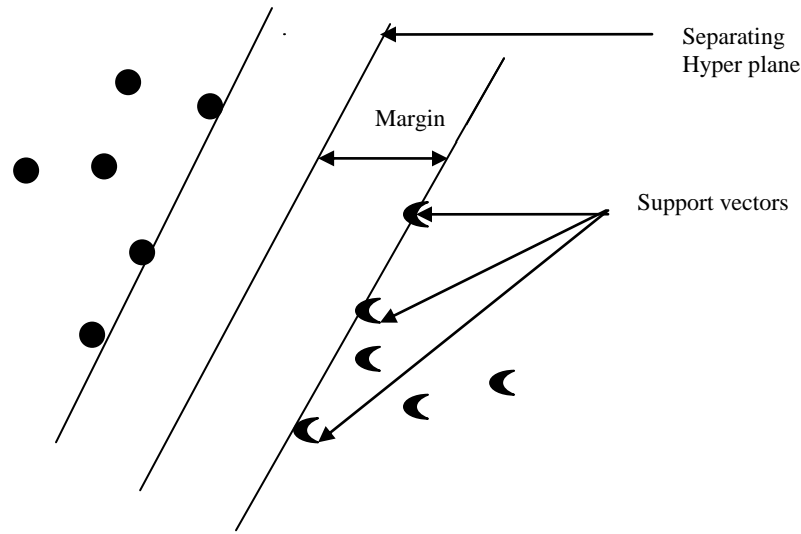


Fig 4: Maximum margin hyperplane.

Given a training set of instance-label pairs. $\{(X_1, C_1), (X_2, C_2) \dots (X_n, C_n)\}$ where C_i is either 1 or -1, a constant denoting the class to which the points X_i belongs. Each X_i is a p -dimensional real vector, of normalized $[-1, 1]$ values. The scaling is important to guard against variables (attributes) with larger variance that might otherwise dominant the classification. This is the training data, which denotes the correct classification that the SVM will eventually distinguish, by means of the dividing (or separating) hyperplane, which takes the form

$$XW - b = 0 \quad (2)$$

The vector W points perpendicular to the separating hyperplane. Adding the offset parameter b helps to increase the margin. In its absence, the hyperplane is forced to pass through the origin, restricting the solution.

To achieve maximum margin, support vectors and the parallel hyperplanes (to the optimal hyperplane) closest to these support vectors in either class has to be found out as shown in Figure 4. It can be shown that these parallel hyperplanes can be described by the following equations

$$XW - b = 1 \quad (3)$$

$$XW - b = -1 \quad (4)$$

If the training data are linearly separable, these hyperplanes can be selected so that there are no points between them and then their distance is maximized to the possible extent. By using geometry, the distance between the hyperplanes is $2/|W|$, which implies that $|W|$ has to be minimized. To exclude data points, care must be taken to ensure that for all i either

$$XW - b \geq 1 \quad (5)$$

or

$$XW - b \leq -1 \quad (6)$$

This can be written as

$$C_i (XW - b) \geq 1, \quad 1 \leq i \leq n \quad (7)$$

The original optimal hyper plane algorithm proposed by Vapnik was a linear classifier. However non-linear classifiers applying the kernel trick was suggested to classify linearly inseparable problems. The algorithm is formally similar; expect that every dot product is replaced by a non-linear kernel function. This allows the algorithm to fit the maximum margin hyperplane in the transformed feature space. This transformation may be non-linear and the transformed space high-dimensional, thus though the classifier is a hyperplane in the high-dimensional feature space it may be non-linear in the original input space.

5. PROPOSED METHOD

The skeletal abnormalities are assessed by using the following steps and shown in the Fig.5:

- In cephalometric digital images, landmark regions such as S, N, Ar, Go, Me, A, B which is explained in Section 2 needed for detecting skeletal abnormalities are selected for this work.
- Reduce the dimension of the data using PCA which is explained in Section 3.1 to extract the features.
- Normalize the feature vectors by using the equation 9 which is explained in the Section 5.1.
- Data are trained and tested by using the SVM model to detect the skeletal abnormalities between normal and abnormal.
- Various performance measure such as Accuracy, Specificity, and Sensitivity are calculated which is explained in section 5.2.

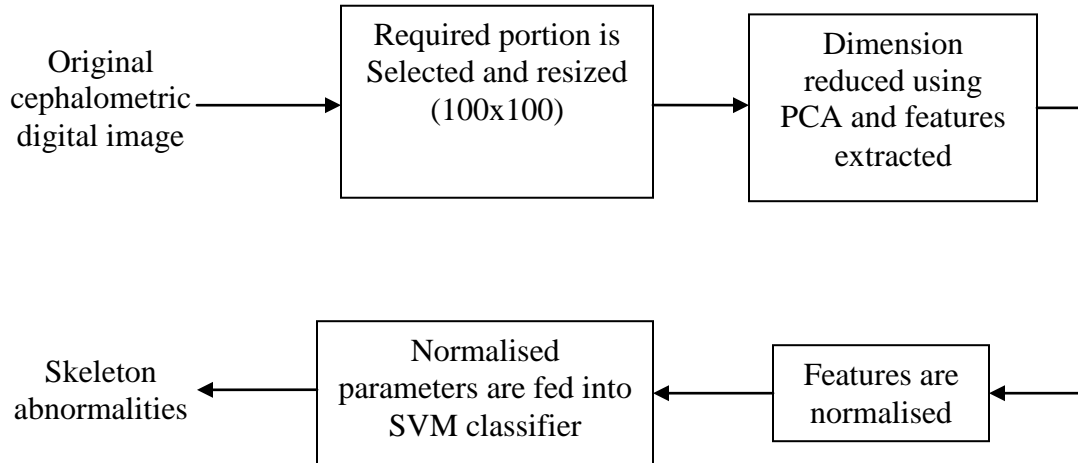


Fig 5: Block diagram of the proposed method

5.1 Normalization of Input parameters

The development of the training network starts with the selection of a number of different combinations of input variables to evaluate the most reliable neural network model. All input and output data are normalized to values between 0 and 1. It is assumed that 'x' has only finite real values, and that the elements of each row are not all equal. The method is described by the following equations for converting any 'Xi' value with normalized 'Yi' value.

$$Y_i = \frac{(Y_{max} - Y_{min}) * (X_i - X_{min})}{(X_{max} - X_{min})} + Y_{min} \quad (8)$$

This equation converts any 'Xi' value into corresponding 'Yi' value in the range of Ymin to Ymax.

For SVM model the values are normalized between -1 and +1. If Ymin = -1 and Ymax = 1 and substituted in equation 8 as,

$$Y_i = \frac{((X_i - X_{min}) / (X_{max} - X_{min})) * (Y_{max} - Y_{min})}{(Y_{max} - Y_{min}) - 1} \quad (9)$$

5.2 Performance Measure

Supervised Machine Learning (ML) has several ways of evaluating the performance of learning algorithms and the classifiers they produce. Measures of the quality of qualification are built from a confusion matrix which records correctly and incorrectly recognized examples for each class. Table presents a confusion matrix for binary classification where TP is true positive, FP is false positive, FN is false negative and TN is true negative values.

Table 4 A confusion matrix for binary classification

Class/recognized	As positive	As negative
Positive	TP	FN
Negative	FP	TN

5.2.1 Accuracy

Accuracy assesses the overall effectiveness of the algorithm and approximates how effective the algorithm is by showing the probability of the true values of the class label. It is given by

$$Accuracy = \frac{TP + TN}{(TP + FP + FN + TN)} \quad (10)$$

5.2.2 Sensitivity/Specificity

Two measures that separately estimate a classifier's performance on different classes are sensitivity and specificity. Sensitivity and Specificity approximates the probability of the positive/negative label being true (assesses the effectiveness of the algorithm on a single class). It is often employed in biomedical and medical applications, and in studies which involve image and visual data and given by

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (11)$$

$$Specificity = \frac{TN}{(FP + TN)} \quad (12)$$

6. EXPERIMENTAL RESULTS

Eight features are used to train and test the SVM model to classify the skeletal abnormalities between normal and abnormal.

6.1 Data preparation:

More than 80 cephalometric images are collected from the various Dental colleges in Tamilnadu. From the images the required portion (where the landmarks are presented) are selected. The digital image is converted into .pgm file format, which is shown in Fig 6. Then the landmarks needed for detecting skeletal abnormalities are selected from the image, which is shown in Fig 7. The image is resized into 100x100, which is shown in Fig 8. The dimension is reduced using principal component analysis and the features are extracted. Then the features are normalized. The normalized features are fed into SVM classifier for training and testing to detect the abnormalities in both facial skeleton analysis, and mandibular and maxillary base analysis.



Fig 6: Cephalometric digital image (Original)



Fig 7 : Selected portion of the image



Fig 8: Resized Image 100 x 100

6.2 Training and testing the SVM

SVMTORCH a free available C++ based object oriented machine learning library was used for training and testing the model. The training data included the eight feature attribute and a class attribute of 60 samples while the test data had only eight feature attribute of 20 samples excluding the class attribute. SVM was created by specifying number of input values, number of output values and type of kernel function. SVM model was trained by varying the types of kernel function. The radial basis function gives the maximum performance.

6.3 Performance evaluation

The classifier output for the test data was compared with the original class attribute for identifying the true positives, true negatives, false positives and false negative values. Table 5 gives these values. Using these values, the overall effectiveness of the algorithm was calculated as 70.00%, the probability of the positive label being true was 76.92% and the probability of the negative label being true was 56.14%. Table 6 gives the performance measure values as discussed in Section 5.2 and shown in Fig 9.

Table 5 Confusion matrix for the SVM model

Class/recognized	As positive	As negative
Positive	10	3
Negative	3	4

Table 6 Performance measures

Measures	Values(%)
Accuracy	70.00
Sensitivity	76.92
Specificity	57.14

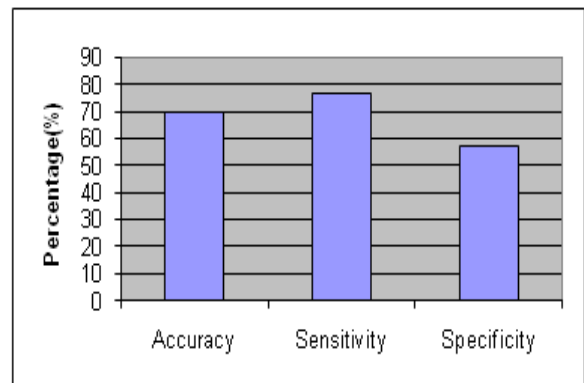


Fig 9: Performance measures of the skeletal abnormality

7. CONCLUSION

In this work, 80 training and 20 testing images was taken. The images are partitioned and the required portion is selected. The selected portions are resized. Principle component analysis (PCA) was used to reduce the dimension of the feature vector and the reduced features were normalized. SVM classifier was used to detect the skeletal abnormalities and the performance measures such as Accuracy, Sensitivity, and Specificity were calculated. The overall performance of the classifier was 70%, hence 70% of skeletal abnormalities were detected correctly.

8. REFERENCES

- [1] Broadbent B H, “A new X-ray technique and its application to orthodontia”, *Angle orthod*, 1: 45- 66, 1931.
- [2] S. Rueda, M. Alcaniz, “ An approach for the automatic cephalometric landmark detection using mathematical morphology and active appearance models”, pp. 159–166, 2006.
- [3] R.Leonard, D.Giordano, F. Maiorana and C.Spampinato, “Automatic Cephalometric Analysis”, the *Angle Orthodontist*: vol.78, no.1, pp. 145-151, 2007.
- [4] OrthoTP, “Orthodontic Folder and Computerized Cephalometry”, AXA Srl via Pierino Colombo, 2007.
- [5] D. Forsyth, D. Davis, “Assessment of an automated cephalometric analysis system”, *Eur J Orthod* 18, 471–478, 1996.
- [6] T. Hutton, S. Cunningham, P. H. P, “ An evaluation of active shape models for the automatic identification of cephalometric landmarks”, *Eur J Orthod* 22, 499–508, 2000.
- [7] V. Ciesielski, A. Innes, J. Sabu, J. Mamutil, “Genetic programming for landmark detection in cephalometric radiology images”, *Int. Journal Knowl Based Intell Eng System*.7, 164-171, 2003.
- [8] J. Yang, X. Ling, Y. Lu, M. Wei, G. Ding, “Cephalometric image analysis and measurement for orthognathic surgery”, *Med Biol Eng Comput* 39, 279–284, 2001.
- [9] M. West. “Bayesian factor regression models in the \large p, small n paradigm”. *Bayesian Statistics*, 7:723-732, 2003.
- [10] R. Bellman. “Adaptive Control Processes: A Guided Tour”. Princeton University Press, 1961.
- [11] Padraig Cunningham “Dimension Reduction” University College Dublin Technical Report UCD-CSI-2007-7, August 8th, 2007.
- [12] K.Pearson, “On lines and planes of closest fit to systems of points in space,” *Phil. Mag*, vol., pp.559-572, 1901.
- [13] H.Hotelling, “Analysis of a complex of statistical variables into principle components”, *Phil. Mag*, vol.24, pp.417-441, 1933.
- [14] Imola K. Fodor, “A survey of dimension reduction techniques”, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, P.O. Box 808, L-560, Livermor, fodor1@llnl.gov, June 2002.
- [15] <http://en.Wikipedia.org/Wiki/Support-Vector>, Machine for SVM.
- [16] Cortes.C & Vapnik.V), “Support-Vector network”, *Machine learning*, 20(2), 273- 297, 1995.
- [17] Duda, P.E.H.R.O & Stock D.G., “Pattern classification”, New York, John Wiley & sons, 2001.