CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

In this chapter the methods of spirometer recording, prediction and classification using neural networks are presented.

3.2 SPIROMETRIC RECORDING AND PROTOCOL

Spirometry is the most commonly performed non-invasive test of ventilatory function. It is used to detect the presence and degree of pulmonary function abnormalities. The spirometers are used to identify how lungs and chest wall interact under normal and pathological conditions before and after the surgical interventions. The spirometer recordings are carried out on adult volunteers \((N = 225)\) for the present study. The age, gender and race of the subject are recorded and height, weight are measured before recording. The portable Microlab spirometer with a volumetric transducer employing Knudson’s reference equations is used for the acquisition of the data from the subjects (Knudson, 1983). The spirometric recordings are obtained from Asthma and Allergy Care Centre, Chennai, India.

Spirometry is based on the measurement of volume \((V)\) and the flow rate of air \((Q)\) as a function of time during breath–in and breath–out actions. The dependence between the air volume and the flow rate is given by,

\[
V(t) = \int Q(t)dt \tag{3.1}
\]
The spirometric test requires maximal effort from the patient and it takes time to perform the pulmonary function test. It is essential that the procedure is carefully and clearly explained to motivate the patient to perform maximally. During the test, the nose is pinched off as the subject breathes through a mouthpiece attached to the spirometer. At the time of recording, the subject inhales to total lung capacity and then exhales as hard and as completely as possible. The tracings are called flow-volume curves and these tracings are very effective in identifying normal, obstructive and restrictive ventilatory states. The test should be repeated until three acceptable and reproducible results are obtained. The acceptability and reproducibility criteria for measurement are adopted as per the recommendation of American Thoracic Society (Clotet et al 2004).

3.2.1 Flow – volume curve and parameters recorded using spirometer

Forced vital capacity, forced expiratory volume at the first second, forced expiratory volume at the sixth second and peak expiratory flow are the parameters recorded using spirometer. FVC is the volume change of the lung between a full inspiration to total lung capacity and is performed during forceful exhalation. FEV$_1$ is the forced expiration volume in one second, which is a vital parameter in pulmonary function test.

A typical flow-volume curve is shown in Figure 3.1. At the beginning of the test, both flow and volume are equal to zero. After this starting point, the curve rises rapidly to a peak point known as Peak Expiratory Flow. PEF measures how fast and hard a subject can exhale while breathing out and is used to evaluate the condition of the airways in cases such as asthma and related breathing disorders. If the test is performed correctly, this PEF is attained within 150 milliseconds of the test. After the PEF the curve descends as more air is expired. 25% of Forced Expiratory Flow is reached when the patient has expired 25% of the expired volume.
Similarly FEF50%, 75% of Forced Expiratory Flow are reached when the patient has expired 50% and 75% of the FVC respectively. The mean flow between the points FEF25% and FEF75% is also an important parameter and is called FEF25-75%. This is actually the first parameter that will decline in many respiratory diseases. When the flow reaches zero, the FVC is reached which indicates that the patient has blown out as much air as they can. After expiration, it is recommended that the patient makes a complete and forced inspiration so as to obtain a closed flow-volume loop (Hansen et al 2001). The flow-volume curves provide more detail for the initial portion (first one second) of the FVC maneuver. The ability to overlay a series of flow-volume curves registered at the point of maximum inhalation may be helpful in evaluating repeatability and detecting sub-maximal efforts (Miller et al 2005).

The FVC, FEV$_1$ and FEV$_6$ are the most important measurements in spirometry and used as the standard index for respiratory system assessment. The spirometric parameters FEV$_1$, FEV$_6$, FEV$_1$/FVC ratio, FEV$_1$/FEV$_6$ ratio, FEF$_{25%}$ and FEF$_{50%}$ represent smaller airways function and correlate better with respiratory abnormalities. The diagnosis of airflow obstruction is based
on an abnormally low value of the ratio of forced expiratory volume in the first second to forced vital capacity (Bruce 2006). The FEV\textsubscript{1} and PEF are considered as indexes of airway caliber and therefore are useful in the assessment of the variable airways obstruction.

### 3.3 PREDICTION OF FEV\textsubscript{1} AND FEV\textsubscript{6}

In this work, prediction of FEV\textsubscript{1} and FEV\textsubscript{6} is carried out using support vector regression and artificial neural networks.

#### 3.3.1 Support vector regression

The Support Vector Regression model maps the original data, \( x \) into a higher dimensional feature space nonlinearly. A given a set of data \( G = (x_i, a_i)_{i=1}^{N} \), where \( x_i \) is the input vector, \( a_i \) is the actual value, and \( N \) is the total number of data patterns, the SVM regression function is

\[
y = f(x) = w_i \phi_i(x) + b
\]

where \( \phi_i(x) \) is the feature space of inputs \( x \), and both \( w_i \) and \( b \) are coefficients which are estimated by minimizing the regularized risk function,

\[
R(C) = C \frac{1}{N} \sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}\|w\|^2
\]

(3.3)

where, \( L_\varepsilon(d_i, y_i) = \begin{cases} 
|d_i - y_i| - \varepsilon, & |d_i - y_i| \geq \varepsilon, \\
0, & \text{else}
\end{cases} 
\)

(3.4)

\( L_\varepsilon \) is called the loss function, \( d_i \) is the desired value and \( y_i \) is the predicted value (Pai et al 2006, Cristianini and Shawe-Taylor 2000, Smola and Scholkopf 2004). The term \( C \) is the regularization constant which specifies
the trade-off between the training errors and the generalization ability of the support vector model.

The equation (3.3) is transformed into the following constrained form;

\[
\text{Minimize : } R(w, \zeta, \zeta^*) = \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^{N} (\zeta_i + \zeta_i^*) \right)
\]

(3.5)

With the constraints,

\[
w_i \phi(x_i) + b - d_i \leq \varepsilon + \zeta_i^* \quad i = 1, 2, \ldots, N,
\]

\[
d_i - w_i \phi(x_i) - b \leq \varepsilon + \zeta_i \quad i = 1, 2, \ldots, N,
\]

\[
\zeta_i, \quad \zeta_i^* \geq 0 \quad i = 1, 2, \ldots, N,
\]

where \(\zeta\) and \(\zeta^*\) are positive slack variables.

This constrained optimization problem is solved using the following primal Lagrangian form:

\[
L(w, b, \zeta, \zeta^*, \alpha, \alpha^*, \beta, \beta^*) = \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^{N} (\zeta_i + \zeta_i^*) \right)
- \sum_{i=1}^{N} \alpha_i \left[ w_i \phi(x_i) + b - d_i + \varepsilon + \zeta_i \right]
- \sum_{i=1}^{N} \alpha_i^* \left[ d_i - w_i \phi(x_i) - b + \varepsilon + \zeta_i^* \right]
- \sum_{i=1}^{N} \left( \beta_i \zeta_i + \beta_i^* \zeta_i^* \right)
\]

(3.6)

Equation (3.6) is minimized with respect to primal variables \(w, b, \zeta\) and \(\zeta^*\), and maximized with respect to nonnegative Lagrangian multipliers \(\alpha, \alpha^*, \beta, \beta^*\).
The dual Lagrangian is obtained as,

\[
J(\alpha_i, \alpha_i^*) = \sum_{i=1}^{N} d_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j)
\]

(3.7)

subject to the constraints:

\[
\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, 2, \ldots, N.
\]

where, \( \alpha_i \) and \( \alpha_i^* \) are Lagrangian multipliers that satisfy the equalities, \( \alpha_i \cdot \alpha_i^* = 0 \). The optimal desired weights vector of the regression hyperplane is represented as

\[
w^* = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x_j)
\]

(3.8)

And the regression function is

\[
f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b
\]

(3.9)

where, \( K(x_i, x_j) \) is the kernel function. The value of the kernel equals the inner product of two vectors \( x_i \) and \( x_j \) in the feature space \( \phi(x_i) \) and \( \phi(x_j) \), i.e.,

\[
K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j).
\]

An inner product in feature space has an equivalent kernel in input space,

\[
K(x_i, x_j) = \left\langle \phi(x_i), \phi(x_j) \right\rangle
\]

(3.10)

The following kernels are used in the regression process.

Linear kernel: \( K(x_i, x_j) = x_i^T x_j \)
Radial basis function kernel:  
\[ K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \]  
(3.12)

and polynomial kernel:  
\[ K(x, x') = (x \cdot x')^d \]  
(3.13)

\( \sigma \) and \( d \) values are chosen on a trial and error basis and the optimal value is identified for the minimum error in prediction or classification.

![Figure 3.2 Representation of maximum margin hyperplane](image)

**Figure 3.2 Representation of maximum margin hyperplane**

Figure 3.2 explains the nature of the optimal hyperplane which has maximum margin between both the classes. Filled and open circles represent outputs of different classes. The dashed lines represent the maximum margin to be obtained in classification. Maximizing the margin seems good because points near the decision surface represent very uncertain classification decisions. The width and the bias estimated are represented diagrammatically.

### 3.3.1.1 Prediction of FEV\(_1\) using support vector regression

Support Vector Regression has been employed in the prediction of FEV\(_1\) for normal and abnormal data. The dataset containing 175 subjects with their corresponding pulmonary function recordings is divided into training set (N=105) and testing set (N=70). Each training input is represented by a vector of length 17 where the first 3 parameters are demographic data and the next 14 are from the spirometer. FEV\(_1\) values for all the 105 training data are fed
as target to the support vector regression model. The regularization constant which determines the upper bound for error in the training of the support vector model is varied from 1 to 10. Linear, quadratic, RBF and polynomial kernels are adopted in the prediction of FEV₁. The width of the RBF kernel is varied from 1 to 5 and the order of the polynomial kernel is varied from 2 to 5 to optimize the values of width and order for the two kernels respectively. The average accuracy in prediction for normal, restrictive and obstructive subjects is estimated based on the difference between the measured and predicted values of FEV₁. The correlation between the measured and the optimal predicted values of FEV₁ for all the subjects are estimated for all the four kernels. The average prediction error and the number of support vectors employed by each kernel are analyzed.

3.3.1.2 Prediction of FEV₆ using support vector regression

Support Vector Regression, has been employed in the prediction of FEV₆ for normal and abnormal data. The dataset containing 50 subjects with their corresponding pulmonary function recordings is divided into training and testing set. Each training input is represented by vector of length 18 where the first 3 parameters are demographic data and the next 15 are from the spirometer.

Linear, quadratic, RBF and polynomial kernels are adopted for this study and the performance of the models with the adopted kernels are evaluated by estimating the average prediction accuracies. The regularization constant which determines the upper bound for error in the training of the support vector model is varied from 1 to 10 and the corresponding accuracies are estimated. The support vector regression model with optimum kernel parameters which include width, order and regularization parameter is selected based on the highest accuracy in prediction of FEV₆. The correlation
between the measured and the optimal predicted values of FEV$_6$ for different kernels are estimated.

### 3.3.2 Artificial neural networks

ANN is able to learn key information patterns with multi-information domain. The advantage of the neural networks is that they can be used to predict one or more output types through a flexible network of weights, transfer functions and input variables. The network is trained on a data set to find the relationship between the input and output by optimizing the weights through minimization of the differences between measured and predicted output variables. Once when the weights of the neural networks have been determined, it can be used for prediction of output from input variables other than the training set.

#### 3.3.2.1 Prediction of FEV$_6$ using back propagation neural networks

The network is trained on the same data set as that of the SVR model. The feed forward neural network with the architecture having one hidden layer, and operating on tan sigmoid transfer function, is employed for the prediction FEV$_6$ of normal and abnormal data. The parameters of a network are adjusted by training the network on a Training of the network is performed under back propagation of the error using conjugate gradient algorithm. The trained network is used to predict values of FEV$_1$.

Mathematically, the functionality of a hidden neuron is described by following equation

$$
\sigma \left( \sum_{j=1}^{n} w_j x_j + b_j \right)
$$

(3.14)
where \( w_j \) and \( b_j \) represent the weights and bias respectively. The number of output neurons equals the number of outputs in the approximation problem. The output of this network is given by

\[
\hat{Y}(\theta) = q(\theta; x) \sum_{i=1}^{n} w_i^2 \sigma \left( \sum_{j=1}^{n} w_{ij} x_j + b_{ij} \right) + j^2
\] (3.15)

where \( n \) is the number of inputs and \( nh \) is the number of neurons in the hidden layer. The variables \( \{w_{ij}^1, b_{ij}^1, w_{ij}^2, b_{ij}^2\} \) are the parameters of the network model and are represented collectively by the parameter vector. In training the network, its parameters are adjusted incrementally until the training data satisfy the desired mapping. In the present study, a nonlinear tan sigmoid function is used and is represented as

\[
Sigmoid [x] = \frac{1}{1+e^{-x}}
\] (3.16)

The network had a configuration of 18 input units, 5 hidden units and 1 output unit. The network is trained using conjugate gradient backpropagation learning algorithm. Each training input is represented by vector of length 18 where the first 3 are demographic data and 15 next parameters are from spirometer data. After the network is trained with 5 hidden units, it is used for the prediction of FEV\(_6\).

3.3.2.2 Prediction of FEV\(_6\) using radial basis function neural networks

Radial basis function networks have proven their practical applicability in prediction and classification tasks. The network consists of three layers: an input layer, a hidden layer, and an output layer. The hidden layer of the RBF network can be viewed as a function that maps the input patterns from a nonlinear separable space to a linear separable space. In the
new space, the responses of the hidden layer neurons form new feature vectors for pattern representation.

An activation function for a hidden layer node is a locally radial symmetric function typically Gaussian function, whose output decays to zero as the distance between the input vector and its center increases. The response characteristics of the $j^{th}$ hidden unit is a gaussian function given by

$$ \psi_h = \exp \left( -\frac{\| X - C_h \|}{2\sigma_h^2} \right)^2 $$

(3.17)

where $\psi_h$ is the output value of the hidden node $h$, $C_h$ is the center of the gaussian, $\sigma_h$ is the standard deviation (width) of the function, $X$ is the network input vector. The response of an output node $m$ of the network can be expressed as

$$ Y_m = \sum_{h=1}^{H} W_{hm} \psi_h $$

(3.18)

where $W_{hm}$ is the weight connecting hidden node $h$ with output node $m$.

3.3.2.3 Radial basis function neural networks with k-means clustering algorithm

The choice of the number of centers and widths of Gaussian functions of nodes of the RBF networks are selected through k-means clustering algorithm. In this algorithm, k initial cluster centers are selected and each pattern $X_j$ is associated with the closest cluster center $C_h$. Each stage of the algorithm moves every center point to the centroid of the patterns in the cluster as given by,
\[ C_h = \frac{1}{n_h} \sum_{j=1}^{n_h} X_j^{(h)} \quad (3.19) \]

The center point is updated by recomputing the distance from each pattern to its nearest center. These steps are repeated until the following criterion function is minimized

\[ E = \sum_{h=1}^{H} \sum_{j=1}^{n_h} \left\| X_j^{(h)} - C_h \right\|^2 \quad (3.20) \]

where \( X_j^{(h)} \) is the training pattern \( j \) associated with the cluster \( h \), \( C_h \) is its vector center and \( n_h \) is the number of patterns belonging to that cluster.

Then the \( p \) nearest neighbour heuristic is applied to determine the width \( \sigma_h \) of the basis functions. According to this rule, the standard deviation of a Gaussian is computed as the root mean square of the distances from its center to the nearest centers

\[ \sigma_h = \left( \frac{1}{p} \sum_{j=1}^{p} \left\| C_h - C_j \right\|^2 \right)^{1/2} \quad (3.21) \]

where \( C_j \) are the \( p \) nearest neighbours of \( C_h \). A \( p \) value of 3 is assumed based on the prediction carried out for first few iterations called the observed trials.

Each training input of the network is represented by vector of length 18 where the first 3 are demographic data and the next 15 parameters are from spirometer data. The network is trained initially with 3 centers whose widths are obtained from k-means clustering algorithm. Then the network is used for the prediction of FEV\(_6\). The training of the network is repeated by varying the number of centers sequentially from 3 to 10. The performance of all NN models is evaluated by comparing their accuracy in
prediction. The network configuration with the best performance is selected for prediction of FEV$_6$.

3.3.2.4 Radial basis function neural networks with fuzzy c-means clustering algorithm

The fuzzy c-means clustering algorithm is used to determine the Gaussian centers of radial basis function neural network from training data. Fuzzy c-means is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. The initial number of clusters and cluster centers are obtained using subtractive clustering. The data point with the highest potential is selected as the first cluster center and all data points within some radii of the first cluster center are removed, in order to determine the next data cluster and its center location. This process is iterated until all of the data is within radii of a cluster center. The cluster estimates thus obtained are used as the initial cluster centers for iterative fuzzy clustering method.

The membership matrix $U$ is defined for the partitioned clusters where each element takes a value in $[0, 1]$ to represent the grade of membership of a training pattern $X_n$ to a cluster $C_h$. The membership function is initialized such that the elements of membership function $U (u_{hn})$ must fulfill the condition

$$\sum_{h=1}^{H} u_{hn} = 1$$  \hspace{1cm} (3.22)

where $H$ is the number of clusters and $n$ is the number of patterns. The centers of the basis functions are computed according to the expression
where $f$ is the fuzziness index and its value equal to 2 is assumed based on observed trials (Marcos et al 2008). The $U$ elements are updated according to

\[ u_{hn} = \frac{1}{2} \left( \frac{\sum_{j=1}^{N} \| x_n - c_j \|^2}{\sum_{j=1}^{N} \| x_n - c_j \|^2} \right)^{j-1} \]  \hspace{1cm} (3.24)

These steps are repeated until the following criterion function is minimized

\[ E_{FCM} (H, f) = \sum_{h=1}^{H} \sum_{n=1}^{N} U(h,n)^f \| x_n - c_h \|^2 \]  \hspace{1cm} (3.25)

Then the $p$ nearest neighbour heuristic is applied to determine the width $\sigma_h$ of the basis functions.

The spirometer data (15 parameters) and 3 demographic data are given as the inputs to the network. The network is trained with 4 centers whose widths are obtained from fuzzy c-means clustering algorithm. Then the network is used for the prediction of $FEV_6$. The prediction performance is evaluated by values of prediction accuracy obtained for the different widths.

### 3.3.2.5 Radial basis function neural networks with self organizing map

RBF networks are good at modeling nonlinear data and can be trained in one stage rather than using an iterative process as in multilayer perceptron and also learn the given application quickly. The training of RBF
network can be divided into two stages. In the first stage, the parameters of radial functions, i.e., Gaussian centers and widths are determined by unsupervised training. This training stage can be adapted from the learning algorithm of self organizing map. This stage creates a topology preserving mapping from input space to a two dimensional map space. In the second stage, the output weights are determined by supervised learning method.

SOM consists of two layers; the input layer and the output layer. The output layer is composed of a set of nodes arranged in a geometric pattern, typically two dimensional lattices. Each neuron from the output layer has a double representation, one is its position in the grid and another is its weight vector. The dimension of the weight vector equals the dimension of the input data vectors. In the training process, the weights are gradually changed in order to span the weight vectors across the input data set. The training is based on competitive and cooperative learning (Kohonen 2000).

In the SOM network, the input pattern $x$ represented by vector of length 18 where the first 15 parameters are from spirometer data and 3 demographic data are presented to a two dimensional map of 5x5 nodes. During each training cycle $t_k$, every input pattern is considered in turn and the best matching weight vector $w$, also called winner node is determined such that

$$|| x - w_i || = min || x - w_i ||, \quad (i = 0, 1, \ldots, N) \quad (3.26)$$

The weights of the neuron, which resembles the most to that of the input data, are updated. The weight vectors are updated using the following adaptation function

$$w_i(t_k + 1) = w_i(t_k) + \left[a(t_k)(w_i(t_k) - x) \right] \quad \text{for } i \in N_i(t_k)$$

$$= w_i(t_k), \quad \text{otherwise} \quad (3.27)$$
where $a(t_k)$ is the learning coefficient that decreases over time and $N_i(t_k)$ is the set of nodes considered to be in the topological neighborhood of node $i$, the winner node. Node $i$ represents the neuron that maximally responds to the input signal, i.e., its weight vector matches most closely, among all the Kohonen layer nodes, to that of the input vector. $N_i$ contains all nodes that are within a certain radius from node $i$. The weight vectors of all the nodes within the set $N_i$ are updated at the same rate (Aguado et al 2008).

When the SOM has been trained, the centroids of the input patterns that create winning neuron are identified as cluster centers of hidden units of radial basis function network. Using nearest neighbour rule, the widths $\sigma_h$ of the basis functions is obtained from these centers. The radial basis function network is trained initially with widths obtained from 5×5 SOM units. Then the network is used for prediction of FEV$_6$. The performance of prediction is evaluated by average prediction accuracy.

3.4 ANALYSIS ON THE EFFECT OF TRANSDUCER RESISTANCE

During forced expiration the resistance of the spirometric flow transducer cooperates with the patient’s respiratory resistance and changes the values of all expiratory parameters, depending on the relation which exists between both resistances. If the flow resistance of transducer $R_s$ is a fraction of the respiratory system resistance $R$, then the time constant is given by

$$\tau_f = (R + R_s)C$$  \hspace{1cm} (3.28)

The error in FEV$_1$ parameter is described by equation 3.29, proposed by Juroszek (2006).
\[ \delta_{FEV_1} = (1 - \eta) - (1 - \eta) \left(1 - \frac{R_1}{R} \right) \]  

(3.29)

where, \( \eta = \frac{FEV_1}{FVC} \) is the Tiffeneau index, the ratio of forced expiratory volume in one second to the forced volume capacity, indicating disease of the respiratory system. The effect of prediction of the most significant parameters is analyzed by estimating the error factor \( \delta_{FEV_1} \) for different \( R/R \) ratios ranging from 0.02 to 0.1. The measured values of \( FEV_1 \) are replaced by optimal values of predicted \( FEV_1 \). Also, the FVC values are replaced by optimal predicted values of \( FEV_6 \) and the corresponding error factors are estimated for different ratios of \( R/R \).

3.5 FEATURE SELECTION USING PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is based on the statistical representation of a random variable. For a random vector population

\[ x = (x_1, ..., x_n)^T \]

The mean \( (\mu_x) \) and the covariance matrix \( (C_x) \) are given by

\[ \mu_x = E(x) \]  

(3.30)

\[ C_x = E \left\{ (x - \mu_x)(x - \mu_x)^T \right\} \]  

(3.31)

The components of \( C_x \), denoted by \( C_{ij} \), represent the covariance between the random variable components and \( x_i \). The component \( C_{ii} \) is the variance of the \( x_i \) component. The variance of a component indicates the spread of the component values around its mean value. From the covariance
matrix, the eigenvectors $e_i$ and the corresponding Eigenvalues $\lambda_i$ are the solutions of the equation

$$C_x e_i = \lambda_i e_i, i = 1,...,n$$ (3.32)

Eigenvalues are ordered in descending Eigenvalues with the first eigenvector having the direction of largest variance of the data to find directions in which the data set has the most significant amounts of energy. For a matrix $(A)$ consisting of eigenvectors of the covariance matrix as the row vectors, by transforming a data vector $(x)$,

$$y = A(x - \mu_x)$$ (3.33)

which is a point in the orthogonal coordinate system defined by the Eigenvectors. Components of $y$ are the coordinates in the orthogonal base. The original data vector $x$ can be reconstructed from $y$ using

$$x = A^T y + \mu_x$$ (3.34)

The original vector $x$ is projected on the coordinate axes defined by the orthogonal basis. The original vector is then reconstructed by a linear combination of the orthogonal basis vectors.

Representing only a few basis vectors of the orthogonal basis instead of using all the eigenvectors of the covariance matrix which are the principal components, the matrix having the first $K$ eigenvectors as rows is denoted by $A_k$ and a similar transformation can be created for which

$$y' = A_k (x - \mu_x)$$ (3.35)

$$x' = A_k^T y' + \mu_x$$ (3.36)

The original data vector is projected on the coordinate axes having the dimension $K$ and transforming the vector back by a linear combination of
the basis vectors. This minimizes the mean-square error between the data and the representation with the given number of eigenvectors (Gabriel 1981, Samanwoy et al 2008 and Aguado et al 2008).

The original feature space consists of 17 spirometric parameters which included the measured FEV$_1$ and FEV$_6$ obtained from normal and abnormal subjects. PCA is employed in transforming this feature space into a new space and the components that account for most of the variability are retained whereas the remaining components are ignored. The principal components obtained from PCA are analyzed for ranking the most significant features. The percentage variances between the various parameters are estimated for the normal and abnormal subjects. The principal components that explain the maximum percentage variance are chosen and the corresponding component magnitudes are analyzed. The parameters with highest magnitudes in the loadings of the principal components are chosen for further classification using SVM.

3.6 SUPPORT VECTOR CLASSIFICATION USING SELECTED FEATURE SET

The SVM is a linear classifier in the parameter space, but it becomes a nonlinear classifier as a result of the nonlinear mapping of the space of the input patterns into the high dimensional feature space. Figure 3.3 depicts the basic architecture of support vector machine classification. The number of support vectors is estimated by $(l-1)E(p)$ where $E(p)$ is the expectation of the probability of error on a test vector and $l$ is the number of training samples (Burges, 1998).

For the set of training vectors belonging to two separate classes,

$$D = \{ (x^i, y^i), \ldots, (x'^i, y'^i) \}, x \in \mathbb{R}^n, y \in \{-1, 1\}.$$  (3.37)
with a hyperplane,

$$\langle w, x \rangle + b = 0$$  \hspace{1cm} (3.38)

where, \( w \) is the vector of hyperplane coefficients, \( b \) is a bias term so that the periphery between the hyperplane and the nearest point is maximized (Hsu and Lin 2002).

![Figure 3.3 Basic architecture of support vector classification](image)

**Figure 3.3 Basic architecture of support vector classification**

Without loss of generality it is appropriate to consider a canonical hyperplane, where the parameters \( w, b \) are constrained by, \( \min_i |\langle w, x^i \rangle + b| = 1 \).

A separating hyperplane in canonical form must satisfy the following constraints,

$$y^i[<w, x^i > + b] \geq 1, \quad i = 1, \ldots, l$$  \hspace{1cm} (3.39)

The distance \( d(w, b; x) \) of a point \( x \) from the hyperplane \((w, b)\) is,
\[ d(w, b; x) = \frac{|\langle w, x' \rangle + b|}{\|w\|} \]  

(3.40)

The optimal hyperplane is obtained by maximising the margin, \( \rho \), which is given by,

\[ \rho(w, b) = \min_{x', y' = -1} d(w, b; x') + \min_{x', y' = 1} d(w, b; x') = \frac{2}{\|w\|} \]  

(3.41)

Hence the hyperplane that optimally separates the data is the one that minimizes

\[ \phi(w) = \frac{1}{2}\|w\|^2 \]  

(3.42)

The solution to the optimization problem of Equation 3.39 under the constraints of Equation 3.36 is given by the saddle point of the Lagrange functional

\[ \phi(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i (y_i [\langle w, x' \rangle + b] - 1) \]  

(3.43)

where \( \alpha \) and \( \alpha_i \) are the Lagrange multipliers. The Lagrange functional has to be minimized with respect to \( w, b \) and maximized with respect to \( \alpha \geq 0 \). Classical Lagrangian duality enables the primal problem to be transformed to its dual problem, which is easier to solve. The dual problem is given by,

\[ \max_{\alpha} W(\alpha) = \max_{\alpha} (\min_{w, b} \phi(w, b, \alpha)) \]  

(3.44)

The minimum with respect to \( w \) and \( b \) of the Lagrangian \( \phi \), are given by,
\[
\frac{\partial \phi}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \quad \text{and} \\
\frac{\partial \phi}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^{l} \alpha_i y_i x_i
\]  

(3.45)

Hence the dual problem is formulated as,

\[
\max_W(\alpha) = \max_{\alpha} \left( \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle + \sum_{k=1}^{l} \alpha_k \right) 
\]

(3.46)

and the solution to the problem is given by,

\[
\alpha^* = \arg \min_{\alpha} \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j < x_i, x_j > - \sum_{k=1}^{l} \alpha_k
\]

(3.47)

with constraints,

\[
\alpha_i \geq 0, i = 1, \ldots, l \quad \text{and} \\
\sum_{j=1}^{l} \alpha_j y_j = 0
\]

(3.48)

Solving Equation 3.44 with constraints Equation 3.45 determines the Lagrange multipliers, and the optimal separating hyperplane is given by,

\[
w^* = \sum_{i=1}^{l} \alpha_i y_i x_i
\]

(3.49)

\[
b^* = -\frac{1}{2} \langle w^*, x_r + x_s \rangle
\]

(3.50)

where \( x_r \) and \( x_s \) are support vectors from each class satisfying,

\[
\alpha_r, \alpha_s > 0, \quad y_r = -1, y_s = 1
\]

(3.51)
The hard classifier is then,

\[ f(x) = \text{sgn}(\langle w^*, x \rangle + b) \]  \hspace{1cm} (3.52)

Alternatively, a soft classifier may be used which linearly interpolates the margin,

\[
\begin{cases}
  -1, & z < -1 \\
  z, & -1 \leq z \leq 1 \\
  +1, & z > 1
\end{cases}
\]

\[ f(x) = h(\langle w^*, x \rangle + b) = \text{where } h(z) \]  \hspace{1cm} (3.53)

This may be more appropriate than the hard classifier of Equation 3.49, because it produces a real valued output between −1 and 1 when the classifier is queried within the margin, where no training data resides. From the Kuhn-Tucker conditions,

\[
\alpha_i (y_i' [\langle w, x_i' \rangle + b] - 1) = 0,= \quad i = 1, \ldots, l
\]  \hspace{1cm} (3.54)

and hence only the points \( x_i' \) which satisfy,

\[
y_i' [\langle w, x_i' \rangle + b] = 1
\]  \hspace{1cm} (3.55)

will have non-zero Lagrange multipliers. These points, termed Support Vectors, are the training errors in the dataset. If the data is linearly separable all the support vectors will lie on the margin and hence the number of support vectors can be very small (Steve 1998). Hence SVM can be used to summarize the information contained in a data set by the support vectors produced.
In the case where a linear boundary is inappropriate the SVM can map the input vector, \( x \) into a high dimensional feature space, \( z \). By choosing a non-linear mapping a priori, the SVM constructs an optimal separating hyperplane in this higher dimensional space. The spirometric dataset with parameters selected using PCA are classified using linear, quadratic, RBF and polynomial kernels. Their performance is compared using measures like sensitivity, specificity, accuracy and the total number of support vectors.

There are four outcomes from the classifier. If the outcome is normal for the input corresponding to a normal subject then it is called a true negative (TN). However if the outcome refers to abnormal subject then it is called false positive (FP). True Positive (TP) and False Negative (FN) are the cases where the abnormal is classified as abnormal and normal respectively. The accuracy, sensitivity and specificity are estimated using the following relation.

\[
\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+TN+FN)}
\]

\[
\text{Sensitivity} = \frac{TP}{(TP+FN)}
\]

\[
\text{Specificity} = \frac{TN}{(TN+FP)}
\]

Accuracy is the representation of classifier performance in global sense. Sensitivity and specificity are the proportions of abnormal data classified as abnormal, normal data classified as normal respectively. Further, a sub analysis on the classification of spirometric data is performed by choosing two significant parameters at a time. FVC and the optimal predicted values of FEV\(_1\) are chosen in the first study for classifying spirometric data with all the four kernels. Similar analysis is performed with FEV\(_1\) and predicted values of FEV\(_6\) for classification. The inter relationship between the chosen parameters are analyzed based on the performance estimates and the number of support vectors chosen for the classification. MATLAB 7.5 software is employed for performing the computations in all the studies.