CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter presents the methods adopted in the prediction of significant flow volume spirometric indices. A detailed description on the use of geometric features and the multiclass classifiers employed in the classification of normal, restrictive and obstructive subjects is presented.

3.2 WORK FLOW

![Figure 3.1 Proposed pipeline of process](image)
A comprehensive pipeline of work flow illustrating the analysis of characterization of inherently dependent multiclass spirometric feature is shown in the Figure 3.1. Ridge regression, modified ridge regression, multivariate adaptive regression splines, random forest regressor and multivariate relevance vector machine are the models adopted in this analysis to predict the missing significant flow parameters namely FEV$_1$ and FVC. Further classification of multiclass subjects as normal, restrictive and obstructive based on the measured spirometric parameters and the derived features is also attempted.

3.3 FLOW-VOLUME SPIROMETRY

Spirometry is pivotal to the screening, diagnosis and monitoring of respiratory diseases. It is increasingly being advocated in primary clinical care practice. Spirometry assesses the integrated mechanical function of the lung, chest wall, and respiratory muscles measuring the total volume of air exhaled from a full inflated lung (Total Lung Capacity (TLC)) to maximal expiration (residual volume (RV)). The simple pulmonary function test is very effective, well validated in diagnosing and monitoring upper, lower airway abnormalities. This primary test is devised to distinguish airway obstruction disorders like asthma, chronic obstructive lung disease from respiratory restriction associated with a wide variety of illnesses such as pulmonary fibrosis, pleural effusions and fibrosis or muscle weakness.

3.3.1 Method of Recording

Spirometry, unlike other tests such as electrocardiogram and ultrasound scan where subjects remain passive, requires effort on the part of the subject. Hence it is essential that the procedure is carefully and clearly explained prior to the start of testing. The technicians enthusiastically coach
the subject throughout the manoeuvre that motivates them to perform maximally.

The American Thoracic Society has recommended guidelines for the standardization of spirometry equipment and acceptability criteria. As per these guidelines, subjects are made to perform three forced manoeuvre with a satisfactory start, continued exhalation and without the evidence of leakage or artifact. After expiration, it is recommended that the subject makes a complete forced inspiration so as to obtain a closed flow-volume loop. The criteria for acceptance of repeatability is achieved when the difference between the largest and second largest maneuver includes (a) FVC should be within 0.150L (b) FEV$_1$ within 0.150L (Miller et al. 2005). When the aforementioned two criteria are not met, then additional maneuver is performed. The best acceptable flow-volume curve is defined as the curve with the largest sum of FVC and FEV$_1$.

The spirometer recordings were carried out on adult volunteers (N=476) for the present study. The age, gender along with height and weight of the subject were recorded prior to the test. The test was performed in a sitting position. The portable Spirolab III spirometer with gold standard volumetric transducer was used in the acquisition of data.

At the time of recording, the subject inhales to total lung capacity and then exhales as hard as possible. Each participant completed up to three iterations with acceptable and reproducible results according to the ATS standards. The tracings obtained are called flow-volume curves and is informative in identifying normal, obstructive and restrictive ventilatory states. The ability to overlay a series of flow-volume curves registered at the point of maximum inhalation may be helpful in evaluating repeatability and detection of sub-maximal efforts (Miller et al. 2005).
3.3.2 Parameters obtained from Spirometry

The parameters FVC, FEV₁, PEF, FEV₁/FVC ratio, FEF₂₅%, FEF₅₀%, FEF₇₅%, FEF₂₅-₇₅%, are recorded from spirometer and utilized in this study. The FVC is the total amount of air that the patient can forcibly exhale after taking a maximal inhalation. FEV₁ is forced expiration volume in first one second. The FVC and FEV₁ are the two most important parameters in spirometry for they are primarily used to distinguish airway obstruction from respiratory restriction.

In subjects with obstructive lung disorder, during exhalation, the diameter of the intrathoracic airways decreases reflecting the increased airway resistance. This further compromises the ability of patients with obstructive diseases to empty the air from the lungs. Trapping of air within the lungs of a subject with obstruction, result in an increase in the Total lung capacity. Obstruction disorder is therefore reduction in the volume of air that can be forcibly expelled from the lungs, reflecting the airway resistance with decrease in volume of FEV₁ and in turn FVC too.

Restrictive disorders manifest as certain changes in the anatomic structures in the alveoli and surrounding lung parenchyma. Unlike obstructive lung diseases, that show a normal or increased TLC, restrictive disorders are associated with a decreased TLC. Measures of expiratory airflow are preserved, with the ratio of forced expiratory volume in 1 second (FEV₁) to forced vital capacity (FVC) increased. In clinical evaluation, a reduced total lung capacity (TLC) without evidence of expiratory flow reduction (increased or normal FEV₁/FVC ratio) is interpreted as restriction.

Forced expiratory flow (FEF) is the flow of air specified at discrete time intervals of 25%, 50% and 75% (FEF₂₅, FEF₅₀ and FEF₇₅). It defines the
fraction air that remains in the forced vital capacity (FVC). The spirometric parameters FEV$_1$, ratio FEV$_1$/FVC, FEF$_{25\%}$ and FEF$_{50\%}$ correlate well with respiratory abnormalities relating to the small and large airways.

### 3.3.3 Geometric Features

The shape of the maximum expiratory flow-volume curve conveys clinical information pertaining to the mechanical properties of the respiratory system. In 2005, the American Thoracic Society (ATS)/European Respiratory Society (ERS) task force noted that the earliest changes associated with airflow obstruction in small airways are thought to be slowing in the terminal portion of the spirogram, even when the initial part of the spirogram is hardly affected. Increased convexity of the curve with respect to the volume axis commonly referred to as ‘scooping’ has been considered a potential functional indicator of the initial stages of the chronic airflow obstruction. This scooping is indicative of non-homogeneous emptying of the lungs (Mead et al. 1967).

Although the clinical interpretation of maximal expiratory flow volume curves is widespread, typical interpretation is based on the absolute and relative (% predicted) flow volumes. In this analysis several methods that includes beta angle, flow ratio index, global and peripheral concavity indices to quantify the shape of the flow-volume curve are studied.

The Beta angle ($\beta^\circ$) method calculates an angle around the flow at 50% of the vital capacity. The $\beta^\circ$ is defined by connecting three points on the flow-volume curve (Figure 3.2). Point A is located at the intersection of co-ordinates for FVC on the x-axis and PEF flow on the y-axis. Point B is forced expiratory flow at 50% of vital capacity (FEF$_{50}$) and with point C at zero volume (Kapp et al. 1988; Dominelli et al. 2015).
Figure 3.2  Illustration of the procedure to determine beta angle from flow-volume curve

Figure 3.3  Illustration of the procedure to determine flow ratio index from flow-volume curve
\[
\beta^\circ = 180 - \tan^{-1}\frac{PEF - FEF_{50}}{0.5\ FVC} + \tan^{-1}\frac{FEF_{50}}{0.5\ FVC}
\]  

(3.1)

The flow ratio index is calculated at 75 percent (FR\textsubscript{75}) of the expired Vital Capacity (VC). This ratio is given as (O’Donnell et al. 1990)

\[
FR_{75} = \frac{(FEF_{50}/2) - FEF_{75}}{FEF_{75}}
\]  

(3.2)

This index is derived on the basis that if the maximal expiratory flow volume curve is linear between FEF\textsubscript{50} and FEF\textsubscript{75}, then flow at 75 percent of the expired vital capacity (FEF\textsubscript{75}) would be exactly one half of the FEF\textsubscript{50}. Subsequently the difference between one half the FEF\textsubscript{50} and the observed value of the FEF\textsubscript{75} reflect the degree of departure of the curve from the straight line at 75 percent of the expired vital capacity (Figure 3.3).

The global concavity index is based on FEF\textsubscript{50\%}. This measure quantifies concavity that involves the entire descending limb. The peripheral index is based on the FEF\textsubscript{75\%} and quantifies concavity present near the terminal portion of the curve (Figure 3.4). The degree of concavity is obtained by calculating the percentage decrease of the measured flows from the corresponding idealised reference flows as follows (Johns et al. 2014)

\[
Reference\ FEF_{50\%} = \frac{PEF \times (FVC/2)}{FVC - y}
\]  

(3.3)

\[
Reference\ FEF_{75\%} = \frac{PEF \times (FVC/4)}{FVC - y}
\]  

(3.4)
Figure 3.4 Illustration of the procedure to determine concavity indices from flow-volume curve

The reference flows are calculated assuming that the descending limb is a straight line from PEF to end-expiration and therefore has no curvature. The variable, y, is the volume of air expired to PEF and it is assumed to be fixed at a value of 0.6 litres as suggested by the authors. The degrees of concavity are calculated as follows:

\[
Global\ concavity = 100 \times \left( \frac{reference\ FEF_{50} - measured\ FEF_{50}}{reference\ FEF_{50}} \right) \tag{3.5}
\]

\[
Reference\ concavity = 100 \times \left( \frac{reference\ FEF_{75} - measured\ FEF_{75}}{reference\ FEF_{75}} \right) \tag{3.6}
\]

(Source: Johns, et al. 2014)
3.4 PARAMETRIC AND NON PARAMETRIC REGRESSION

3.4.1 Ridge Regression

Ridge regression is a flexible, penalized regression approach performing $L_2$ regularization. The technique minimizes a loss function that includes the sum of squared regression residuals and a positive penalty parameter $\lambda$. This penalty prevents overfitting by shrinking the regression coefficients ($\beta_{ridge}$) towards zero.

Let $\{X = (x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(i)}, y^{(i)})\}$ represent the spirometric samples in the training set with $N$ number of instances. The ridge estimator $\hat{\beta}_{ridge}$ for a model of the form $\hat{y} = \beta^T x^{(i)} + \beta_0 + \epsilon$ is given as (Erika Cule et al. 2011)

$$
\min_{\beta_0, \beta} \left\{ \sum_{i=1}^{N} (\hat{y} - y)^2 + \lambda \sum_{j=1}^{M} \beta_j^2 \right\} 
$$

(3.7)

where $\hat{y}$ is the model output target of $i^{th}$ sample, $x^{(i)}$ is the $i^{th}$ input sample with $M$ predictors, $N$ the number of samples, $\beta_0$ and $\beta=(\beta_1, \beta_2, \ldots, \beta_M)$ are the regression coefficients to be estimated.

The amount of shrinkage of the regression coefficients is determined by the control parameter ($\lambda$). A commonly used technique to determine the optimal control parameter is through visual examination of the ridge trace-plot. The value that stabilises the estimates is considered. Alternatively cross validation can also be used for model selection. The ridge regression estimates $\beta$ that have smaller mean square error (MSE) can be used in the prediction of significant flow volumes. In this analysis, the optimal $\lambda$ is empirically chosen by selecting the value that gives minimum leave-one-out cross-validation error.
A further modification in the ridge regression is the feature weighted model. A weighted penalty $W_j$ for each feature is employed such that each regression coefficient is penalized individually depending on the nonnegative generic weight $W_j$. The modified ridge regression is given as (Yan et al. 2015)

$$
\min_{\beta_0, \beta} \left\{ \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 + \lambda \sum_{j=1}^{M} W_j \beta_j^2 \right\}
$$

(3.8)

with $\beta_j$ is the coefficient of spirometric feature j and $W_j$ a weighted term based on Tikhonov regularization.

### 3.4.2 Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines are non-parametric, adaptive regression procedure with a greater flexibility to explore the non linear relationships of the underlying data. MARS makes no assumptions of the functional relationships between dependent and independent variables.

The basic principle in MARS that it approximate separate regression slopes over distinct intervals. The training data sets are partitioned into separate piecewise linear segments called splines that are of differing gradients. The spline functions are then connected smoothly together to form Basis Functions (BFs). The MARS regression model takes the form

$$
\hat{y} = a_0 + \sum_{m=1}^{M} a_mB_m(x)
$$

(3.9)

where $\hat{y}$ is the predicted response, $x$ is the spirometric feature, $a_0$ and $a_m$ are estimated coefficients to yield the best fit of data, $M$ is the number of basis
functions included into the model and $B_m(x)$ is the $m^{th}$ basis function. MARS employs two-sided truncated power functions as spline functions described by the following equations

\[
(-(x - t))^q_+ = \begin{cases} (t - x)^q & \text{if } x < t \\ 0 & \text{otherwise} \end{cases}
\]

\[
(+)(x - t))^q_+ = \begin{cases} (t - x)^q & \text{if } x \geq t \\ 0 & \text{otherwise} \end{cases}
\] (3.10)

where $t$ is the joining points of the polynomials called knots and $q$ is the power to which the splines are raised. This parameter determines the degree of smoothness of the resultant function estimate.

An optimal MARS model is constructed through a two-stage, forward and backward procedure. The forward phase places candidate knots at random positions within the range of each spirometric feature to define a pair of BFs. At each step, the model adapts the knot and its corresponding pair of BFs to give the maximum decrease in sum-of-squares residual error. This process of adding basis function continues until a predefined maximum number is reached. But, this often results in a very complicated and over fitted model contributing to poor performance.

Model pruning is performed in the backward phase. The redundant basis functions that contribute least to the accuracy of fit are removed in the backward phase one at a time. To determine the basis function that should be included in the final model, MARS employs Generalized Cross-Validation (GCV) criteria. The GCV is the mean-squared residual error divided by a penalty term dependent on the model complexity. It is defined as

\[
GCV(M) = \frac{1}{n} \sum_{m=1}^{n} \left( \hat{y}_m - \hat{f}(x_m) \right)^2 \left( 1 - \frac{C(M)}{n} \right)^2
\] (3.11)
where \( n \) is the number of observations, \( \left( 1 - \frac{C(M)}{n} \right)^2 \) is the penalty measure of model complexity. The complexity cost function \( C(M) \) is expressed as

\[
C(M) = M + d \cdot M
\]  

(3.12)

where \( M \) is the number of basis, \( d \) represents the cost function for each basis function included in the developed sub-model and is also the smoothing parameter.

### 3.4.3 Random Forest Regressor

A Random forests is a non parametric algorithm based on model aggregation ideas. RF applies bagging sampling approach (Breiman 2001) and random feature selection to construct a collection of decision trees. The strategies employed to build the forest, enable it to achieve better generalization with low variance.

In RF each decision tree in the ensemble is constructed using a bootstrap sample \( \{X_b\} \) with replacement from the training data. Statistically, the subsets have about 64% of train input appearing at least once in the samples that are used in constructing the tree. It is referred to as in-bag instances. The remaining of about 36% of input train data are referred to as Out-Of-Bag (OOB) instances.

Further each tree is grown, evaluating the best split at every interior node with a subset of spirometric features \( (m_{try}) \) chosen at random. The subset feature that produces the least sum of squares is regarded to be the optimal split. Prediction of a new data is the average of the response values predicted by each single tree. Both bagging and random feature selection assures the low correlation amongst the trees.
The out-of-bag mean square error is given as

\[
\text{OOB} - \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_{\text{OOB}})^2
\]  

(3.13)

where \( \hat{y}_{\text{OOB}} \) is the average prediction for the \( i \)th observation from all trees for which this observation has been out-of-bag instance. This internal validation is typically considered as a good estimator of the error expected for future input data.

An important attribute of RF is its feature ranking measure. The permutation based feature importance measure is obtained by randomly permuting the feature variables and computing the increase in the accuracy loss in the out-of-bag instances. Specifically, to measure a feature ‘k’s importance in RF trees, the values of this feature is randomly shuffled in the out-of-bag samples. The permutation importance is then evaluated comparing the prediction accuracy of a tree before and after random permutation of the predictor variable of interest. Thus for a feature \( X_k \),

\[
\text{OOB} - \text{MSE}_t(X_k \text{ permuted}) = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \hat{y}_{\text{OOB}}(X_k \text{ permuted}) \right)^2 
\]  

(3.14)

When the feature is of relevance, then accuracy decreases as the original association to the response is disturbed by permutation. The average of differences over all trees provides the concluding importance score of the corresponding feature.

Two main parameters, the number of trees in the forest \( N_{\text{tree}} \) and the number of input features randomly chosen at each split are optimized in the RF algorithm. The number of input features \( (m/3) \) is the default choice. To get an appropriate value of \( N_{\text{tree}} \), several iterations increasing number of
trees are performed and the value when the prediction error stabilizes is chosen to be the ideal.

### 3.4.4 Multivariate Relevance Vector Machine

Thayananthan (2005) proposed the Multivariate Relevance Vector Machine to provide a regression tool capable of generating multivariate outputs. This model is the extension of the sparse Bayesian model developed by Tipping & Faul (200).

Given a training data set of input-target vector pairs \( \{x^{(n)}, t^{(n)}\}_{n=1}^{N} \) where \( N \) is the number of observations \( x \in \mathbb{R}^D \) is a D-dimensional input vector, \( t \in \mathbb{R}^M \) is the M-dimensional output target vector. The model learns the dependency between input and output target with the aim of making accurate predictions of \( t \) for previously unseen input \( x \)

\[
t = W\Phi(x) + \varepsilon
\]  

(3.15)

where \( W \) is a \( M \times P \) weight matrix and \( P = N+1 \). A fixed kernel function \( K(x, x^{(n)}) \) is used to create a vector of basis functions of the form \( \Phi(x) = (1, K(x, x^{(1)}), \ldots, K(x, x^{(N)})). \) The error is assumed to be zero-mean Gaussian.

Let \( t = [\tau_1, \ldots, \tau_r, \ldots, \tau_M]^T \) and \( W = [w_1, \ldots, w_r, \ldots, w_M]^T \). A likelihood distribution of the weight matrix can be written as the product of Gaussians of weight vectors \( (w_r) \) corresponding to each target output \( (\tau_r) \) (Thayananthan et al. 2005).

\[
p \left( \{t^{(n)}\}_{n=1}^{N} \mid W, S \right) = \prod_{r=1}^{M} \mathcal{N} (\tau_r \mid w_r\Phi, \sigma_r^2)
\]  

(3.16)
where $\Phi = (1, \Phi_{x_1}, \Phi_{x_1}, \ldots, \Phi_{x_1})$ is the design matrix. The above equation involves parameters of weight $(W, \sigma_r^2)$ whose maximum likelihood estimation may sometimes suffer from severe over-fitting. To overcome this, Tipping (2001) proposed a constraint on selection of parameters in Bayesian perspective by defining an explicit zero-mean Gaussian prior probability distribution over them which is

$$p(W|A) = \prod_{r=1}^{M} \prod_{j=1}^{P} \mathcal{N}(w_{rj}|0, \alpha_j^{-2}) = \prod_{r=1}^{M} \mathcal{N}(w_r|0, A) \quad (3.17)$$

where $A = diag(\alpha_1^{-2}, \ldots, \alpha_p^{-2})$ is a diagonal matrix of hyperparameters $\alpha_j$ and $w_{rj}$ is the $(r,j)^{th}$ element of the weight matrix $W$.

The posterior parameter distribution conditioned on the spirometric data can then be written as the product of Gaussians for the weight vectors of each target output and is given as

$$p(W|\{t\}_{n=1}^{N}, S, A) \propto p(\{t\}_{n=1}^{N} | W, S) p(W|A) \prod_{r=1}^{M} \mathcal{N}(w_r|\mu_r, \Sigma_r) \quad (3.18)$$

The posterior distribution of the weights is Gaussian represented as $\mathcal{N}(\mu_r, \Sigma_r)$ with mean of $\alpha_r^{-2} \Sigma_r \Phi^T \tau_r$ and covariance of $\Sigma_r = (A + \alpha_r^{-2} \Phi^T \Phi)^{-1}$. The marginal likelihood is then

$$p(\{t\}_{n=1}^{N}|A, S) = \int p(\{t\}_{n=1}^{N} | W, S) p(W|A) dW \quad (3.19)$$

$$= \prod_{r=1}^{M} \int \mathcal{N}(\tau_r | w_r \Phi, \sigma_r^2) \mathcal{N}(w_r|0, A) \quad (3.20)$$

$$= \prod_{r=1}^{M} |H_r|^\frac{1}{2} \exp\left(-\frac{1}{2} \mathbf{\tau}_r H_r^{-1} \mathbf{\tau}_r \right)$$
Thus, it is possible to obtain set of hyperparameters by maximizing the marginal likelihood using the marginal likelihood maximization. During optimization process, many elements of $\alpha$ go to infinity and their corresponding posterior probability of the weight becomes zero. The relatively few non-zero weights correspond to the input vectors that form the sparse core of the Relevance vector model and are called Relevance Vectors (RVs). Inducing sparsity can be an effective method to control model complexity, avoid over-fitting and control computational characteristics of model performance. The optimal hyper parameters are then used to obtain the optimal weight matrix with optimal covariance $\Sigma^{opt} = \left\{ \Sigma_r^{opt} \right\}_{r=1}^M$ and mean $\mu^{opt} = \left\{ \mu_r^{opt} \right\}_{r=1}^M$. The predictive distribution for any new input $x^*$, corresponding to a target $t^*$

$$p(t^* | t, \alpha^{opt}, (\sigma^{opt})^2) = \int p(t^* | W, (\sigma^{opt})^2). p(W | t, \alpha^{opt}, (\sigma^{opt})^2) dW$$  \hspace{1cm} (3.22)

### 3.5 PARAMETRIC AND NON-PARAMETRIC CLASSIFICATION

#### 3.5.1 Random Forest

The Random forest classifier is a meta-learner consisting of a collection of tree-structured classifiers. To achieve diversity among decision trees, these base learners are constructed upon two stage randomization procedure. Each tree is based upon a bootstrap sample of the input introducing the first level of randomization. Thus each tree in the forest is constructed with different subset of spirometric data resulting in a collection of diverse trees. The next layer of randomization is introduced at the node level where random subsets of the spirometric features are used at each tree node to find the best split for the node. Once the forest is built, to classify a new instance $(X_t)$, it is run across all the trees grown in the forest. In a
classifier, majority voting of the ensemble of trees determine the class of this new instance.

Given an ensemble of classifiers $H = \{h_1(x), h_2(x), \ldots, h_N(x)\}$, input vector $x$ and its corresponding output $y$, the margin function measures the average number of votes. The margin of the new instance computed over all the other classes in all trees of the ensemble $H$ is,

$$mg(x,y,H) = \frac{1}{|H|} \left( \sum_{i=1}^{|H|} I(h_i(x) = y) - \sum_{i=1}^{|H|} \max_{j \neq y} I(h_i(x) = j) \right) \quad (3.23)$$

### 3.5.2 Probabilistic Neural Network

Probabilistic Neural Network (PNN) (Specht 1990) is a nonlinear classifier and uses minimum Bayesian risk criterion to classify the multidimensional spirometric data into $C$ classes. It consists of four layers namely Input layer, Pattern layer, Summation layer and Output layer (Figure 3.5). The Bayesian rule in the pattern layer estimates the conditional probability of each class of the given input vector. Likelihood of the test data to each training data is computed in the pattern layer using the simple Bayesian rule. PNN assigns the test data to the class with maximum likelihood compared with other classes. The kernel is a Gaussian probability distribution function. A spread parameter is used as a global parameter which determines the width of the kernel.

The input layer neurons are applied with the training samples. Let $\mathbf{X}_k^j$ for $j = 1, 2, \ldots, N_i$ be the spirometric samples with $k$ features vectors from a class $C_i$. These inputs are distributed to the pattern layer where the neurons are divided into the number of classes $C$. 
Figure 3.5 Architecture of Probabilistic neural network

The likelihood function for the $j^{th}$ pattern neuron in class ‘$c$’ is calculated using

$$f(X_i) = \frac{1}{(2\pi\sigma)^{k/2}} \exp\left(-\frac{||X - \bar{X}_c||^2}{2\sigma^2}\right)$$

(3.24)

where $\bar{X}_c$ is the center of the kernel and $\sigma$ is the spread parameter.

The summation layer of the network computes the class conditional probability functions through summation of the previously computed densities

$$S_k(X_i) = \sum_{t=1}^{N_c} \exp\left[\frac{\bar{X}_t w_{ik} - 1}{\sigma^2}\right]$$

(3.25)
where $N_c$ is the number of pattern neurons of class $c$ and $w_{cj}$ is the weight between the pattern and summation layer and is expressed as

$$
w_{kj} = \begin{cases} 
1 & \text{if } T_c = 1 \\
0 & \text{else}
\end{cases} \quad (3.26)$$

The value of $T_c$ is 1 only when the sample is associated with class $c$ and is 0 elsewhere. The output layer classifies the input sample into a specific class if that class had a maximum output value from the corresponding node at the summation layer

$$y = \arg\max_{1 \leq c \leq C} (S_k) \quad (3.27)$$

### 3.5.3 Kernel based Extreme Learning Machine

Extreme learning machine (ELM) is a much wider type of generalized single layer feed forward network (Figure 3.6). The network aims to reach the lowest training error with the smallest norm of output weight. Given $N$ training samples $\{(x_i, t_i)\}^N$, the ELM maps the input data into $L$-dimensional hidden layer feature space $h(x_i)$. The output function of the generalized ELM is given as

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \quad (3.28)$$

where $\beta = [\beta_1, \ldots, \beta_L]^T$ is the vector of the output weights between the hidden layer of $L$ nodes and the output node; $h(x) = [h_1(x) \ldots h_L(x)]$ is the output vector of the hidden layer with respect to the input $x_i$. The multiclass constrained optimization is formulated as

Minimize: $L_{PELM} = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^{N} \|\varepsilon_i\|^2 \quad (3.29)$

Subject to: $h(x_i)\beta = t_i^T - \varepsilon_i^T$. 


Where \( \varepsilon_i = [ \varepsilon_1 \ldots \varepsilon_{i,m}]^T \) are the training error vector of the \( m \) output nodes with respect to the input spirometric sample \( x_i \) and \( t_i = [ t_1 \ldots t_{i,m}]^T \) the target vector. Using the Karush-Kuhn-Tucker theorem and its optimality conditions the target vector is given as

\[
T = \left( \frac{I}{C} + HH^T \right) \alpha \tag{3.30}
\]

Followed by, the \( \beta \) is given as

\[
\beta = H^T \left( \frac{I}{C} + HH^T \right)^{-1} T \tag{3.31}
\]

The output of ELM classifier is

\[
f(x) = h(x)\beta = h(x)H^T \left( \frac{I}{C} + HH^T \right)^{-1} T \tag{3.32}
\]

where \( I \) is the identity matrix and \( C \) is the cost parameter. The Gaussian kernel function used in the hidden layer is given as

\[
G(a, b, x) = \exp \left( -b \|x - a\|^2 \right) \tag{3.33}
\]
The output of the ELM classifier is,

\[ f(x) = h(x)H^T \left( \left( \frac{I}{C} + HH^T \right)^{-1} \right)^T \]  

(3.34)

The generalization performance of the ELM classifier depends on the cost parameter C and kernel parameter Y.

### 3.5.4 Naïve Bayes Classifier

Naïves Bayes classifier is a probabilistic classifier and uses Bayes rule with strong independence assumptions. The probability that an instance X belongs to class k is

\[ Pr(Y = k|X) = \frac{Pr(Y = k) \prod_{j=1}^{D} Pr(X_j|Y = k)}{\prod_{j=1}^{D} Pr(X_j|Y = i)} \]  

(3.35)

where \( X = (X_1, X_2, X_3 \ldots X_D)^T \). The probability distributions for each class i with each feature j, i.e. (\( Pr(X_j|Y = i) \)) is assumed to be Gaussian with mean (\( \mu \)) and the standard deviation (\( \sigma \)).

### 3.6 PERFORMANCE MEASURES

The assessment of model fitness is statistically evaluated with the measures of correlation coefficient, coefficient of determination and the root mean square. Pearson's product moment correlation coefficient is used to assess the linear association between the measured and predicted flow volumes.
The goodness of fit of a model is assessed using the $R^2$ statistics. The measure shows the proportion of the variance of the predicted flow volumes from the measured volumes. It is given as

$$R^2 = 1 - \frac{1}{n} \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$  \hspace{1cm} (3.37)$$

The root mean square of residuals defines the expected magnitude of the prediction error and is given by

$$\text{RMSE} = \left[ \frac{1}{n} \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]^{1/2}$$  \hspace{1cm} (3.38)$$

The Bland-Altman plot is used to study the bias and derive the 95% confidence interval. The Information Based Measure of Disagreement (IBMD) score compares two values on the basis of Shanon’s notion of entropy. IBMD uses logarithms to measure the amount of information contained in the differences between two observations. Let $N$ be the number of samples and $x_i, x_j$ represent the pair of observations for $x$. The disagreement score is

$$\text{IBMD} = \frac{1}{N} \sum_{k=1}^{N} \log_2 \left( \frac{|x_{k1} - x_{k2}|}{\max(x_{k1}, x_{k2})} + 1 \right)$$  \hspace{1cm} (3.39)$$

To develop models that guarantee not only high accuracy, but also good generalization, robustness of model parameter with respect to future changes in the nature of the input data is required. A change in the training data used to build a model is one way to achieve this. Different sets of training data produce models with very different generalization accuracies.
The bootstrap aggregating method is used to explore the implication of the change in the nature of input data. Bootstrap aggregating, also called bagging, is a machine learning ensemble technique designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid over fitting.

Empirical evaluation is used for assessment of classifiers to determine their feasibility on a specific domain. The classification measures of sensitivity, specificity and precision are presented below (Marina Sokolova et al. 2009)

\[
\text{Accuracy} = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + fn_i + tn_i + fp_i}}{l} \tag{3.40}
\]

\[
\text{Specificity} = \frac{\sum_{i=1}^{l} \frac{tn_i}{tn_i + fp_i}}{l} \tag{3.41}
\]

\[
\text{Precision} = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} \tag{3.42}
\]

\[
\text{Sensitivity} = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \tag{3.43}
\]

In a multiclass classification, for every individual class \( l \), the correctness of the classification model is evaluated using false-positive (\( fp_i \)), false-negative (\( fn_i \)), true-positive (\( tp_i \)), and true-negative (\( tn_i \)) values. The classification of normal data as abnormal is measured as \( fp \), while the classification of abnormal data as normal is considered as \( fn \). True positive and True negative refer to the state where abnormal is classified as abnormal and the normal is classified as normal, respectively.

In learning imbalanced data, the Geometric Mean has been suggested as a suitable alternative for measuring per class accuracy over regular accuracy. G_mean is defined as [Sun Y et al.2006]
$$G_{\text{mean}} = \left( \prod_{i=1}^{l} \frac{tr_i}{n_i} \right)^{\frac{1}{m}}$$

(3.44)

where $l$ is the number of classes, $n_i$ is the number of examples in class $l$ and $tr_i$ is the number of correctly classified examples in class $l$.

Cobweb representation (Diri & Albyrak 2008) of performance assessment for multiclass classification is used to analyze the misclassification error. It is a graphical representation in which each of the separate misclassifications represents one corner on a polygon. The likelihood is generally used to compare the performance of any classifier with the chance classifier in terms of misclassification rates. A chance classification with three classes is represented with $1/3$ (0.33) likelihood.