Pulmonary function test plays an important role for the diagnosis, prognosis and mass screening of respiratory disorders. These tests are often performed using flow-volume spirometers. The parameters estimated through such measurements are useful for clinical inferences and treatment modalities. Spirometry is an effort dependent pulmonary function test that requires cooperation of the subject with the examiner (Enright et al 2005).

It has been shown that 50% of the results obtained from spirometric measurements are unacceptable due to inability of the patients to complete the test. As full inspiration or expiration is difficult to achieve for subjects with breathing difficulties, many times spirometric measurements result in incomplete dataset (Ulmer 2003). Also, with large number of parameters to assess, the analysis of spirometric data becomes difficult. The conventional predicted values used in spirometers employ only regression equations involving age, height, sex and race. A few correction factors are included if the prediction equations for the appropriate race is not available in the literature in such cases. This includes lot of assumptions and approximations in the predicted values. Hence, there is a need to predict significant parameters in case of missing data set and to analyze the spirometric recordings using selected features.

In this work, investigations are carried out on enhancing the diagnostic relevance of spirometric measurements using support vector
regression and classification. Four different kernels which include linear, quadratic, RBF and polynomial kernels are employed in the regression and classification models. The pulmonary function data are recorded from volunteers using flow volume spirometer and the prediction of most significant parameters such as forced expiratory volume in 1 second and 6 seconds are carried out. The maximum prediction accuracy is obtained by appropriate choice of kernel parameters of SVR model.

Back propagation and radial basis function neural networks are employed for prediction of FEV$_6$ and the prediction performance is compared with that of SVR. The optimum number of hidden nodes of back propagation neural networks is selected based on the maximum prediction accuracy. The centers and widths of the radial basis function networks are selected using three different clustering algorithms such as k-means, fuzzy c-means and self organizing map.

The prediction performance of SVR is evaluated using Principal component analysis. The interdependency among the spirometric parameters is assessed using the variation in magnitude and direction of the principal components. The influences of spirometric transducer resistance on the measured and predicted parameters are analyzed using error factor.

Principal component analysis is used to identify the most significant parameters from the entire dataset. The features selected using PCA are utilized for classification of normal and abnormal subjects using support vector machines. The kernel for classification is chosen based on the performance estimates and the number of support vectors used by the model.

Results demonstrate that SVR is capable of predicting both FEV$_1$ and FEV$_6$ values. The accuracy in prediction of FEV$_1$ is found to be high for normal subjects when compared to abnormal subjects. A regularization
constant of C=10 and polynomial kernel of order 2 are found to predict the FEV₁ values with higher accuracy when compared to SVR with other kernels and neural network algorithms. Similarly, Radial Basis Function kernel of width five and second order polynomial kernel with a regularization constant C=10 predict FEV₆ values with high accuracies in both normal and abnormal cases.

The value of C and number of support vectors used by the prediction models is found to have a significant influence on prediction accuracy. High prediction accuracies are obtained for the model with higher value of C and lower number of support vectors. The polynomial kernel employs lesser number of support vectors than RBF kernel as the data becomes comparatively linearly separable.

The number of nodes used by back propagation neural networks for the prediction of FEV₆ with higher accuracy is found to be five. In RBF neural networks, the optimum width of the radial basis function is determined using self organizing map. The prediction error of the neural networks is less for obstructive subjects when compared to normal and restrictive subjects. SVR is found to perform better than the neural network model as it employs structural risk minimization principle which minimizes the upper bound on the expected risk, whereas neural networks works on minimizing the error.

Analysis on the prediction of FEV₁ and FEV₆ using PCA shows that the magnitude and direction of the principal components corresponding to measured and predicted values of the two parameters remained the same. Also, the variations in error factors due to influence of transducer resistance are identical. These results demonstrate high similarity between measured and predicted values and thus validate the efficiency of prediction of support vector regression.
The prediction of most significant parameter, $\text{FEV}_1$, is useful in assisting the diagnosis for elderly patients and children who cannot perform the test completely. Similarly, prediction of $\text{FEV}_6$ is significant since $\text{FEV}_6$ values are not available in many recordings of patients with severe restriction and obstruction. This method of prediction of $\text{FEV}_1$ and $\text{FEV}_6$ using SVR shows that there is a possibility to improve the analysis of spirometric investigations with little intervention from clinician.

Principal component analysis is able to reduce the number of features required for further analysis and select the most significant features from the dataset. The parameters selected by the principal component analysis include FVC, $\text{FEV}_1$, PEF, $\text{FEV}_1$/FVC and $\text{FEV}_6$ which represent the forced expiratory capabilities of the subject. It is found that, the volume of the air that can be forcibly blown out during the first and sixth second of expiration and the maximal flow achieved during forced expiration are the most important parameters that characterize the respiratory state of the subject. Principal component analysis is able to extract these relevant parameters as the most significant features of spirometric recordings.

In the classification of abnormalities based on the selected features, SVC with fourth order polynomial kernel is found to perform better with a sensitivity of 60%, specificity of 80%, accuracy of 70% and the number of support vectors equal to 13. Thus, SVM classification performed using these selected parameters are found to show better results in distinguishing respiratory abnormalities. Thus it appears that this method of prediction of most significant parameters and classification of abnormalities using SVM enhances the diagnostic relevance of spirometry.
5.1 SCOPE FOR FUTURE WORK

In this work, comprehensive analysis of the human respiratory functions is carried out using flow-volume spirometry and SVM. The normal and major abnormalities such as obstructive and restrictive conditions are considered to validate the approaches. This study could be extended to analyze the mechanics of lung functions under varied experimental conditions and disease states. The prediction accuracy of SVM could further be improved by optimizing the kernel parameters using particle swarm optimization and genetic algorithm. Also, multiclass support vector machines could be employed in further classifying the abnormal data into restrictive and obstructive subjects. In addition, the prediction using SVM could also be extended for other spirometric parameters like PEF, FEF etc, as these parameters are either missing or improperly recorded in most of the investigations.