5.1 CONCLUSION

Spirometry is an important investigation in assessment of lung function and presents overall information relating to the large and small airways. It measures the quantity of air that can be moved into and out of the lungs. The flow volume parameters estimated through the measurement are useful for clinical inferences and treatment modalities (Gillette et al. 2011). The forced breathing maneuver also generates flow-volume curve that represent tidal, inspiratory and expiratory phases of breathing. Changes in the contour of the loop can aid in the diagnosis and localization of airway disorder along with prior patient history and physical examination. Spirometry, consequently allows the identification of obstructive and restrictive defects, the two broad categories of respiratory diseases.

Forced expiratory volume in one second is a well established index to quantify the severity of respiratory disease. A decrease in FEV$_1$ may reflect reduction in the maximum inflation of the lungs assessing the elastic properties, obstruction of the airways, loss of lung elastic recoil or respiratory muscle weakness (Husain et al. 2008). For these reasons, the role of FEV$_1$ is found to be vital in the discrimination of lung disorders. FVC reflect the lung size as well as the rate of lung emptying. The American Thoracic Society standards for diagnosis and management of patients with pulmonary impairment have graded disease severity based on the ratio of FEV$_1$/FVC.
Spirometry is therefore a widely accepted diagnostic test to assess ventilatory function. Performing a reliable spirometric test involves patient co-operation satisfying the current guidelines for acceptability and repeatability criteria of the American Thoracic Society. However, impairment in pulmonary physiology or apraxia may sometimes lead to difficulty in performing spirometry, with low performance expectations compounded by suboptimal test quality or inconclusive test. Thus the uncertainty associated with diagnostic testing can be aided with a decision analysis support system.

In this analysis, attempt has been made to predict most significant flow parameters of Forced expiratory volume in one second and Forced vital capacity in inconclusive tests. Regularization based parametric ridge regression model was employed to predict the significant flow volume parameters with the incomplete, inherently dependent spirometric measurements. The optimal value of the control parameter was selected at $\lambda = 2100$. The correlation coefficient for obstructive and restrictive subjects represented a stronger relationship than the normal subjects. However, the agreement plots between the measured and predicted values showed a presence of homoscedasticity. Additionally, the modified weighted ridge regression model showed an improved performance in comparison to the traditional ridge regression model. It was inferred that the modified ridge model is able to predict both FEV\textsubscript{1} and FVC with higher positive correlation, in particular the restrictive group of subjects more accurately.

Non-parametric, non-linear regression models involving basis expansions were applied to analyze the data with complex interdependent spirometric measurements. The Multivariate Adaptive regression splines, based on recursive partitioning and multi-stage regression automatically captured the nonlinearities and interactions present in the measurement. The generalized cross validation enabled in the construction of less complex
MARS based FEV₁, FVC prediction models with fewer features. Non-linear spline functions associating PEF, FEF₂₅, age and height were obtained in MARS FEV₁ regression model. Similarly features FEV₁, FEF₂₅₋₇₅, FEF₇₅ and height were identified as significant features in the FVC model. The R-squared model assessment statistic of the final models was in good agreement with the measured dataset.

The ensemble of decision trees provided a unique combination of prediction accuracy and model interpretability. Even though the RF is so-called black-box model, the implicit feature selection measure enabled the study of significant features and ranked them. The spirometric indices in the FEV₁ regression model were ranked in the order PEF, FEF₂₅, FEF₅₀ and FEF₂₅₋₇₅. The FEV₁ was identified as the single most important predictor in the FVC model.

Evaluation of Multivariate relevance vector machine performance was based on the selection of optimal kernel width. Several trials were performed for obtaining the optimal values. The model with kernel width = 100 achieved a maximum accuracy. The relevance vectors in the model carries significant information that is used for estimation of missing flow values and is also proportional to the complexity of the model. The final model was generated with 29 relevance vectors. This conveys the ability of the Bayesian learning embodied in the MVRVM in producing very sparse models yielding results with high efficiency.

The clear distinctions of estimated accuracies, RMSE values between parametric and nonparametric methods when underlying feature pattern is highly dependent suggests that data analysis of a combination of parametric and nonparametric methods could be used as a diagnostic model to predict the missing values of inconclusive tests. The results also suggest that
accurate prediction of the significant flow parameters is attained best with the nonparametric Random forest regressor.

Respiratory disease requires a complex series of decisions due to high interdependency in the flow volume parameters. Following estimation of significant flow volumes parameters, further study comprised the automated multiclass characterization of subjects according to disease risk. The performances of the classifiers were evaluated both with measured and Random forest regressor predicted volumes.

The estimates of prior probability of the classes and the associated class-conditional probability of each of the input feature in the Naïve Bayes classifier showed that the model had an extremely close proximity in the classification performance metrics in the measured and predicted flow volume data sets.

The extreme learning machine classifier with linear and Gaussian kernel functions show improvement in accuracy than the Naïve Bayes classifier. High values of accuracies were also observed with the measured spirometric feature set. A wide variation is observed in output performances when the classifiers were presented with significant flow volumes predicted by the RF regressor model.

A random forest classifier with an ensemble of 700 trees is applied in the analysis of identification of normal, obstructive and restrictive subjects. The spirometric features that include Height, beta angle, FEF<sub>50</sub>, FEV<sub>1</sub>, FEF<sub>25-75</sub>, FVC and the ratio of FEV<sub>1</sub>/FVC were considered as the more significant features contributing to the classification accuracy. The results show the evidence of highest values in the output performance metrics.
However, the classifier yielded lower accuracies when the RF model predicted values were presented.

PNN, the multi layered network based on probability density estimation and competitive learning, yielded consistent results when presented with a spirometric measurement and RF model predicted volume in inconclusive test. The results support the potential of creating an accurate and widely accessible method for differentiating pulmonary function disorders.

Considering the high incidence of chronic respiratory diseases coupled with the similarities in symptomatology between disorders, there is a clear need for decision support system that can quickly and accurately estimate risk for disease conditions. The unified framework of ensemble trees based regressor with the probabilistic neural network model can be used to differentiate obstructive, restrictive pulmonary disorders from normal subjects. This aids understanding of the pulmonary functions with incomplete recordings for therapeutic choices and prognostic purposes.

The significance and the clinical relevance of this study are perceived as follows:

- Research show that 14% population had difficulty in performing spirometry. Particularly, in elderly subjects with cognitive impairment the possibility of failing spirometry is high (Vanjare et al. 2015)

- Analysis conducted in children show only 54% of them were able to perform acceptable and repeatable spirometry based on the ATS criteria (Gaffin et al. 2012)
- The spirometric results are thus unacceptable due to the failure in completing the test which makes the diagnosis difficult.

- This work enhances the diagnostic relevance of spirometry as it helps in prediction of the most significant parameters namely FEV$_1$ and FVC and characterizing the diverse categories of respiratory abnormalities.

- The analysis included the quantification of derived features specifically the beta angle, concavity indices and their applicability in disorder identification.

- The study demonstrated that even with small training samples, non-parametric methods provide better result than the parametric ones.

5.2 SCOPE FOR FUTURE WORK

In this work, comprehensive analysis of the human respiratory functions is carried out using flow-volume spirometry. The analysis based on the identification of an efficient classifier model capable to differentiate subjects as normal, obstructive and restrictive. The classifier preceded by the regression model predicts the significant flow parameters missing in the incoming feature set. The effective methods identified to be successful in this analysis with inconclusive, inherently dependent spirometric features could be extended to analyze the mechanics of lung functions under varied experimental conditions and disease stages incorporated with feature selection techniques.