ABSTRACT

Spirometry is the widely used diagnostic tool in the assessment of pulmonary function abnormality. It is a physiological test that measures individual exhalation and inhalation lung volumes as a function of time.

Spirometric indices of forced expiratory volume in one second, forced vital capacity, peak expiratory flow have long been the mainstay in diagnosis of lung disorders. The valuable information from the spirometry investigations is useful in monitoring patients with suspected or previously diagnosed respiratory pathology.

Spirometry, although a powerful tool that plays a significant role in the early diagnosis of lung damage and its associated structures, is effort-dependent. It sometimes becomes complex to ensure co-operation and attain completion of the test in geriatric, foreign-language subjects and young children.

The objective of the research is estimation of the significant lung volumes parameters and characterization of changes in pulmonary function dynamics, correlating various lung conditions in inconclusive pulmonary function test. The lung function data is acquired from volunteer participants (N =476) using flow volume spirometer as per standard recording protocol.

Flow volume estimations of the significant parameters are carried out using parametric and non-parametric learning models that includes ridge regression, modified ridge regression, multivariate adaptive regression splines, random forest and multivariate relevance vector machine. The performances of the models are evaluated using the prediction error statistics of root mean square, the coefficient of determination, information-based measure of disagreement. The correlation with true data for the normal, restrictive and obstructive subjects is also analyzed.
Further, characterization of the incomplete test is carried out using the spirometric features, the geometrical features derived from the flow-volume curve and the predicted significant flow volumes. Kernel based Extreme Learning machine, Probabilistic neural network, Random forest and Naïve Bayes classifiers are employed in the analysis of multiclass respiratory abnormalities.

The performance analysis reveals that the non-parametric, ensemble based Random forest exhibit highest model efficiency of $R^2 = 0.88$, 0.83 in FEV$_1$ and FVC prediction models. Experimental insights show that the model is able to rank candidate spirometric predictors through its inbuilt feature importance measures. PEF, FEF$_{25}$, FEF$_{50}$ are identified as dominant features in the FEV$_1$ prediction model. The bootstrap sampling technique and feature optimization at each split, while constructing the forest in the RF model, had outperformed the other regressor with minimal error between the measured and predicted significant flow volumes.

The assessment of the classifiers in the ability to classify subjects with the measured as well as predicted flow volumes is performed. Comparison of the results describes the probabilistic neural network classifier model allows adaptation to classify subjects with a consistent performance. The results exhibited sensitivity of 89.51% and 88.97%, specificity of 94.81% and 94.59% in both measured and predicted flow volumes feature sets.

This unified framework of randomized trees based regressor with artificial intelligence emerges out to be an efficient evaluation and monitoring system. It appears to support clinicians with enhanced diagnostic relevance in spirometric investigations with incomplete data and poor recording.