World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:9, No:04, 2015

Comparison of Multivariate Adaptive Regression Splines and Random Forest Regression in Predicting Forced Expiratory Volume in One Second

Authors : P. V. Pramila , V. Mahesh

Abstract : Pulmonary Function Tests are important non-invasive diagnostic tests to assess respiratory impairments and provides quantifiable measures of lung function. Spirometry is the most frequently used measure of lung function and plays an essential role in the diagnosis and management of pulmonary diseases. However, the test requires considerable patient effort and cooperation, markedly related to the age of patients esulting in incomplete data sets. This paper presents, a nonlinear model built using Multivariate adaptive regression splines and Random forest regression model to predict the missing spirometric features. Random forest based feature selection is used to enhance both the generalization capability and the model interpretability. In the present study, flow-volume data are recorded for N= 198 subjects. The ranked order of feature importance index calculated by the random forests model shows that the spirometric features FVC, FEF 25, PEF,FEF 25-75, FEF50, and the demographic parameter height are the important descriptors. A comparison of performance assessment of both models prove that, the prediction ability of MARS with the `top two ranked features namely the FVC and FEF 25 is higher, yielding a model fit of R2= 0.96 and R2= 0.99 for normal and abnormal subjects. The Root Mean Square Error analysis of the RF model and the MARS model also shows that the latter is capable of predicting the missing values of FEV1 with a notably lower error value of 0.0191 (normal subjects) and 0.0106 (abnormal subjects). It is concluded that combining feature selection with a prediction model provides a minimum subset of predominant features to train the model, yielding better prediction performance. This analysis can assist clinicians with a intelligence support system in the medical diagnosis and improvement of clinical care.

Keywords : FEV, multivariate adaptive regression splines pulmonary function test, random forest
Conference Title : ICBCBBE 2015 : International Conference on Bioinformatics, Computational Biology and Biomedical Engineering
Conference Location : Los Angeles, United States
Conference Dates : April 03-04, 2015