MACHINE LEARNING AND STATISTICAL APPROACHES FOR ASTHMA DISEASE PREDICTION AMONG ADULTS IN SRI LANKA

Jayakodi Ralalage Nishani Amalka Gunawardana

A dissertation submitted in partial fulfilment of the requirement for the Master of Science degree in Business Analytics

Department of Business

Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, Aberdeen, Scotland, UK

2021

Abstract

Background: Asthma is an airway-induced inflammatory lung disease that causes breathlessness, wheezing and regular potentially fatal attacks. Diagnosis of asthma at affordable costs is often challenging due to its variability. A machine learning approach and an appropriate application to provide asthma status prediction would be valuable in clinical practice and self-detection of the disease.

Methods: This study utilized data of 596 asthmatics and 5898 non-asthmatics who participated in the Sri Lanka Health and Ageing Survey (SLHAS) during the 2018-2019 period in Sri Lanka. Both doctor-diagnosed asthma and patient-reported asthma were considered when deciding the asthma status of a patient. In this research, thirteen machine learning classification algorithms were built on under-sampled data, and ten algorithms were created using imbalanced data. These include machine learning models such as; Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbors (KNN), Gradient Boost, XGBoost, AdaBoost, CatBoost, LightGBM, Multi-Layer Perceptron (MLP), and Probabilistic Neural Network (PNN). The performances of these algorithms were evaluated by employing various measures, including Area Under Curve of Receiver Operating Characteristics (AUC ROC) and confusion matrix related indices.

Results: The model comparison showed that a Hybrid version of Logistic Regression and LightGBM obtained the highest model performance with AUC and sensitivity of 0.9062 and 79.85%. The developed Hybrid models take wheeze related parameters, Shortness of Breath (SOB) attacks, coughing attacks, tightness in the chest, nasal allergies, physical activeness, exposure to passive smoking, ethnicity and sector, as input parameters and predicts the asthma status. The web application developed eases the burden of users by allowing them to get their own estimates upon entering data.

Conclusion: A combination of Logistic Regression and LightGBM models can be utilized to predict the presence of adult asthma successfully. The proposed expert system helps patients in their diagnosis of asthma in both self-diagnosis and clinical diagnosis aspects.