|ICAS2025|

Non-Communicable Disease Prediction By using Machine Learning Approach

Abstract: Cardiovascular and other non-communicable diseases (NCDs), especially diabetes and cancers, are some of the main critical global health problems that are mainly caused by a high rate of morbidity and mortality. The paper proposes a novel machine learning (ML) scheme to model NCD risks based on anthropometric data on 300 participants in the Jaffna Teaching Hospital and Sabaragamuwa University of Sri Lanka. Important characteristics like age, gender, height, weight, body mass index (BMI), and visceral fat area will be derived in order to determine risk factors for the disease. The data analysis methodology is based on strong data preprocessing, removing noise and normalisation using min-max scaling, and the correction of outliers using Python libraries such as pandas. Classification makes use of supervised ML algorithms, namely, Random Forest, Extreme Gradient Boosting, Artificial Neural Network, Decision Tree, AdaBoost, Logistic Regression, CatBoost, and Support Vector Machine. The data is divided into 80 percent training set and a 20 percent testing set, which are optimised by grid search cross-validation to provide strong model parameters. The strategy is an effective way to improve the early identification of NCDs, which allows providers to have a flexible, data-intensive resource to provide high-quality and timely interventions, leading to the overall improvement of preventive care and population health in resource-limited contexts.

Keywords: Obesity, Data mining, Supervised learning models, prediction, Non- Non-communicable disease.