applied sciences

MDPI

Article

Hypertension Prediction in Adolescents Using Anthropometric Measurements: Do Machine Learning Models Perform Equally Well?

Soo See Chai [1,*], Kok Luong Goh [2], Whye Lian Cheah [3], Yee Hui Robin Chang [4] and Giap Weng Ng [5]

[1] Faculty of Computer Science and Information Technology, University of Malaysia Sarawak (UNIMAS), Kota Samarahan 94300, Sarawak, Malaysia
[2] Faculty of Science and Technology, i-CATS University College, Kuching 93350, Sarawak, Malaysia; klgoh@ppks.edu.my
[3] Department of Community Medicine and Public Health, Faculty of Medicine and Health Sciences, University of Malaysia Sarawak (UNIMAS), Kota Samarahan 94300, Sarawak, Malaysia; wlcheah@unimas.my
[4] Faculty of Applied Sciences, Universiti Teknologi MARA, Kota Samarahan 94300, Sarawak, Malaysia; robincyh@uitm.edu.my
[5] Faculty of Computing & Informatics, Universiti Malaysia Sabah, Kota Kinabalu 88400, Sabah, Malaysia; nggiapweng@ums.edu.my
* Correspondence: sschai@unimas.my

Abstract: The use of anthropometric measurements in machine learning algorithms for hypertension prediction enables the development of simple, non-invasive prediction models. However, different machine learning algorithms were utilized in conjunction with various anthropometric data, either alone or in combination with other biophysical and lifestyle variables. It is essential to assess the impacts of the chosen machine learning models using simple anthropometric measurements. We developed and tested 13 machine learning methods of neural network, ensemble, and classical categories to predict hypertension in adolescents using only simple anthropometric measurements. The imbalanced dataset of 2461 samples with 30.1% hypertension subjects was first partitioned into 90% for training and 10% for validation. The training dataset was reduced to eight simple anthropometric measurements: age, C index, ethnicity, gender, height, location, parental hypertension, and waist circumference using correlation coefficient. The Synthetic Minority Oversampling Technique (SMOTE) combined with random under-sampling was used to balance the dataset. The models with optimal hyperparameters were assessed using accuracy, precision, sensitivity, specificity, *F1*-score, misclassification rate, and AUC on the testing dataset. Across all seven performance measures, no model consistently outperformed the others. LightGBM was the best model for all six performance metrics, except sensitivity, whereas Decision Tree was the worst. We proposed using Bayes' Theorem to assess the models' applicability in the Sarawak adolescent population, resulting in the top four models being LightGBM, Random Forest, XGBoost, and CatBoost, and the bottom four models being Logistic Regression, LogitBoost, SVM, and Decision Tree. This study demonstrates that the choice of machine learning models has an effect on the prediction outcomes.

Keywords: adolescents; anthropometric; hypertension; imbalanced dataset; machine learning prediction; SMOTE

Citation: Chai, S.S.; Goh, K.L.; Cheah, W.L.; Chang, Y.H.R.; Ng, G.W. Hypertension Prediction in Adolescents Using Anthropometric Measurements: Do Machine Learning Models Perform Equally Well? *Appl. Sci.* **2022**, *12*, 1600. https://doi.org/10.3390/app12031600

Academic Editors: Keun Ho Ryu, Nipon Theera-Umpon and Kwang Sun Ryu

Received: 28 December 2021
Accepted: 31 January 2022
Published: 2 February 2022

1. Introduction

A chronic disease, also known as a non-communicable disease, is a health condition that is not contagious and can endure for a long time. According to a recent World Health Organization (WHO) report [1], chronic diseases claim the lives of 41 million people each year, accounting for 71% of all deaths worldwide. Low and middle-income countries make up for 77% of chronic disease mortality. The majority of chronic disease fatalities, which