

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; klgo@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

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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

account for 17.9 million deaths per year, are from cardiovascular disease. Hypertension is a crucial factor in the development of cardiovascular disease. Hypertension, often known as high blood pressure, is defined as a systolic blood pressure reading ≥ 140 mmHg and/or a diastolic blood pressure reading ≥ 90 mmHg. Systolic blood pressure measurements show the pressure in the blood vessels when the heart beats or contracts, whereas diastolic blood pressure measurements represent the pressure in the blood vessels when the heart rests in between beats. A recent study [2] reported that every 20 mmHg systolic and 10 mmHg diastolic pressure increase above a baseline blood pressure of 115/75 doubles the risk of cardiovascular death.

Hypertension is no longer only an “adult disease”; a growing number of teenagers and younger children are succumbing to it [3,4], as a result of today’s youth’s physically inactive lifestyle. There is growing evidence that childhood hypertension is a precursor to adult hypertension [5]. Unfortunately, children’s hypertension is not diagnosed until it has progressed to the point of being life-threatening or until they reach adulthood [6]. Given the long-term health consequences of uncontrolled hypertension, as well as the fact that pediatric (age 2 to 18) hypertension is a diagnostic signal for numerous important underlying medical illnesses, the need for early and correct diagnosis cannot be overstated. Furthermore, before any further clinical indications arise, childhood and adolescence are the important stages for effective treatment and prevention of hypertension-related cardiovascular problems. Malaysia, a middle-income country, has a prevalence of 24.5% for hypertension in adolescents [7].

Anthropometric measures are non-invasive quantitative measurements of the body that include height, weight, head circumference, Body Mass Index (BMI), body circumferences (waist, hip, and limbs) to determine adiposity, and skinfold thickness [8]. There is a growing body of evidence on the use of anthropometric measures to predict hypertension in children and adolescents, some of which may be found in [4,9–12]. The commonly used anthropometric measurements for hypertension prediction include body mass index (BMI), waist circumference (WC), weight-to-hip ratio (WHR), and weight-to-height ratio (WHtR). Nonetheless, data suggest that the predictive abilities of anthropometric measurements for hypertension vary by country and ethnicity [13].

As machine learning (ML) has gained traction in the medical field, new algorithms for predicting hypertension have emerged. When it comes to hypertension, machine learning technologies might be used as a supplementary tool or a second opinion to assist medical doctors in making timely decisions. The use of anthropometric measurements in ML models yielded varied results with different models. The following review focuses on some of the latest research that employed multiple ML models to predict hypertension using anthropometric measures as the input features. Zhao et al. [14] utilized a dataset of 29,700 participants aged 18 to 70 years old to deploy four ML models, namely Random Forest (RF), CatBoost, Multi-layer Perceptron (MLP) neural network, and Logistic Regression (LR) for hypertension risk prediction. Along with anthropometric measures, their work utilized demographic and lifestyle data as inputs to the machine learning algorithms. The ten selected input features were age, gender, BMI, WC, family history, occupation, smoke, drink, healthy diet, and physical activity. The data were randomly divided into training and validation in the ratio of 4:1. During the training stage, the training set was divided into 9:1 for training and verification sets. On the test set, the models’ performance was measured by the Area Under Curve (AUC), accuracy, sensitivity, and specificity. They concluded that RF performed the best with AUC = 0.92, accuracy = 0.82, sensitivity = 0.83, and specificity = 0.81.

In another research by Boutilier et al. [15], ML models were used to develop risk stratification algorithms for diabetes and hypertension. Five ML models, including Decision Tree, regularized Logistic Regression, k-Nearest Neighbor, RF, and AdaBoost, were developed and tested in their study. The input of the models included data from the questionnaire: weight, height, waist circumference, blood pressure, heart rate, and blood glucose, and the output for hypertension classification was based on the assessment of the medical doctor.

Using AUC to measure the performance of the models, they discovered that RF (0.792) performed slightly better than Logistic Regression (0.776), followed by AdaBoost (0.770), k-Nearest Neighbor (0.705), and Decision Tree (0.610). The sample size employed in the study was 2278, with an average age of 50.6.

A three hidden layers Artificial Neural Network (ANN) model was developed as a classification model for hypertension patients using gender, race, BMI, age, smoking, kidney disease, and diabetes in [16]. Using an imbalanced dataset of 24,434, with 69.71% non-hypertensive patients and 30.29% hypertensive patients, the model was compared with decision forest, Logistic Regression, Support Vector Machine, boosted Decision Tree and Bayes point machine. The ANN model developed managed to achieve a sensitivity of 40%, a specificity of 87%, precision of 57.8%, and a measured AUC of 0.77. They concluded that the accuracy of the approaches is relatively similar when compared to the other five ML models, but that the AUC and *F1*-score of the ANN method are somewhat higher and more competitive.

As evidenced by the review, various ML models were developed and employed for hypertension prediction. It was observed that different performance metrics were used to choose the best model. This makes it difficult for field researchers to select the most appropriate candidate for this. Furthermore, the researchers compared three to six ML models without concentrating on the three categories of supervised algorithms, namely neural network, ensemble model, and classical model. In terms of input features, it is obvious that anthropometric measurements were not the only data used in the models produced. Aside from demographic data, lifestyle data, such as smoking, a healthy diet, and physical data, as well as physiological data, such as blood pressure, heart rate, and blood glucose, were used. As self-reporting lifestyle parameters are subjective [17], and there is a need for specialized instruments to acquire physiological data, our study will investigate the use of anthropometric measures along with easily collected demographic data for hypertension prediction using ML models. Furthermore, studies on the associations between anthropometric measures and hypertension in adolescents are relatively limited when compared to adults [4,18]. Therefore, our study intends to fill this research gap. In our previous work [19], we used anthropometric measurements and simple demographic data to develop a one hidden layer of 50 neurons Multilayer Perceptron (MLP) neural network to predict hypertension in adolescents, yielding a sensitivity of 0.41, specificity of 0.91, precision of 0.65, *F1*-score of 0.50, accuracy of 0.76, and AUC of 0.75. In this study, we extend our previous work by investigating the efficacy of thirteen different ML models: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, k-Nearest Neighbor, Multilayer Perceptron, Gradient Boosting, XGBoost, LightGBM, CatBoost, AdaBoost, and LogitBoost, from the three supervised ML categories of neural network, ensemble model and classical model for hypertension prediction in adolescents using anthropometric measurements and simple demographic data. In order to tackle the imbalanced data problem, we implement and evaluate the Synthetic Minority Over-sampling Technique (SMOTE) and the combination of SMOTE with under sampling. To the best of our knowledge, this study is the first to look at the efficacy of ML models for hypertension prediction utilizing thirteen distinct algorithms of the three supervised ML categories in adolescents. In this study, we would be investigating if simple anthropometric measurements are viable for hypertension prediction using machine learning models, and what are the effects of the different models on the prediction results? The objectives of the study are two-fold: (a) investigate the feasibility of anthropometric measurements and simple demographic data for hypertension prediction, and (b) implement, evaluate, and analyze the performance of the thirteen different ML models for hypertension prediction in adolescents using easy-to-collect data.

2. Materials and Methods

2.1. Data Source

The data for this study originate from a cross-sectional study conducted in Sarawak, Malaysia, to determine the blood pressure of secondary school students aged 13 to 17. The data were collected over seven months, beginning on 9 March 2016, and ending on 27 September 2016. Sarawak is located on the island of Borneo. It is the largest state in Malaysia. The total Sarawak population in 2021 was estimated to be 2.82 million [20]. There are more than 40 different sub-ethnic groups in Sarawak, and the six major sub-ethnic groups are Iban, Chinese, Malay, Bidayuh, Melanau, and Orang Ulu. Data were collected from 19 secondary schools in Sarawak. Among these schools, 14 schools were in rural areas and 5 were urban schools. Exclusion criteria during data collection include students with physical and mental disabilities, pre-diagnosed hypertension, and illnesses that might lead to secondary hypertension. Prior to data collection, the students were asked to relax for 5 min and confirmed that they had not consumed any coffee, medicine, or exercised. A total of 2461 data samples were collected.

2.1.1. Demographic Data

The participants were asked to fill in their age, gender, ethnicity, school, and parents' hypertension history (none, one, or both). Table 1 shows the details of the demographic data obtained. The distribution of gender in the dataset was quite fair, with a slightly higher percentage of females (58.0%) compared to male students (42.0%). The majority of the students were from rural areas (74.2%). The number of students with parents with no hypertension history (78.5%) was much higher than the number of students with one or both hypertensive parents (21.5%). Throughout the data collecting process, trained personnel double-checked to ensure that students supplied the essential demographic information.

Table 1. Demographic data description.

Age (years \pm std)	14.5 \pm 1.50
Gender	
Male	1033 (42.0%)
Female	1428 (58.0%)
Ethnicity	
Iban	737 (29.9%)
Malay	681 (27.7%)
Chinese	475 (19.3%)
Bidayuh	256 (10.4%)
Other	312 (12.7%)
Location	
Urban	634 (25.8%)
Rural	1827 (74.2%)
Parent Hypertension History	
One	448 (18.2)
Both	80 (3.3)
None	1933 (78.5)

2.1.2. Anthropometric Measurements

The anthropometric measurements were measured and recorded by a team of trained personnel using a SECA body meter and a portable weighing scale.

- Weights

During the weighting procedure, students were required to wear only their school uniforms and no shoes. The weights were recorded to a 0.1 kg precision.

- Heights

The students were asked to stand straight with no footwear on a level surface with the backs of their heels and occiput against the apparatus for height measurements. The heights were recorded to a 0.1 cm precision.

- Circumferences

Waist circumference was measured using a plastic non-elastic tape positioned between the midway of the last rib and the top of the hip bone (iliac crest).

- Anthropometric Indices

Body Mass Index (BMI), weight-to-height ratio (WHtR) and conicity index (CI) were calculated for each student based on the anthropometric measurements obtained.

1. Body Mass Index (BMI)

The Body Mass Index (BMI) is a weight-to-height ratio ($\text{weight}(\text{kg})/\text{height}(\text{m}^2)$). Although this is an imperfect measurement, the index is commonly used to assess obesity [21].

2. Weight-to-Height Ratio (WHtR)

WHtR is the ratio of the circumference of the waist to the height of the person. Compared to BMI or waist circumference alone, WHtR is more sensitive in detecting the clustering of coronary risk factors in non-obese men and women [22].

3. Conicity index (CI)

CI is calculated using body mass, height, and waist circumference as indicators of abdominal obesity [23]. Cardiovascular risk factors and CI have been shown to have a substantial relationship [24].

- Blood Pressure

Blood pressure was measured using a digital blood pressure monitor. Before the measurement, the participants were asked to rest for 5 min to make sure that they had not exercised. The participants were also screened to make sure they had not had any coffee or medicine prior to the test. Two measurements were collected for each subject. Between these two measurements, there was a one-minute gap. A third reading would be obtained if the discrepancy between these two values was greater than 5 mm Hg. A third reading would be obtained if an individual was determined to have prehypertension or hypertension. The final blood pressure value for each participant would be derived as the average of these readings. Following the 4th report on the diagnosis, assessment, and treatment of high blood pressure in children and adolescents [25], the participants were divided into three groups: pre-hypertension, hypertension, and normal, based on the cut-off point of age, sex, and height.

2.1.3. Ethics Statement

The research obtained ethical approval from the Medical and Ethical Committee of Universiti Malaysia Sarawak (UNIMAS/TNC (AA)-03.02/06-11 Jld.3(1)) and the Ministry of Education Malaysia. Prior to data collection, the students' parents and caregivers were provided an information leaflet related to the research and their consent was obtained.

2.2. Methodology Design and Implementation

This section will explain the methodology used to develop and evaluate the different ML models for hypertension prediction in adolescents using simple anthropometric measurements. The general processes involved in this study are depicted in Figure 1. A detailed explanation of each process is provided in the sections that follow.

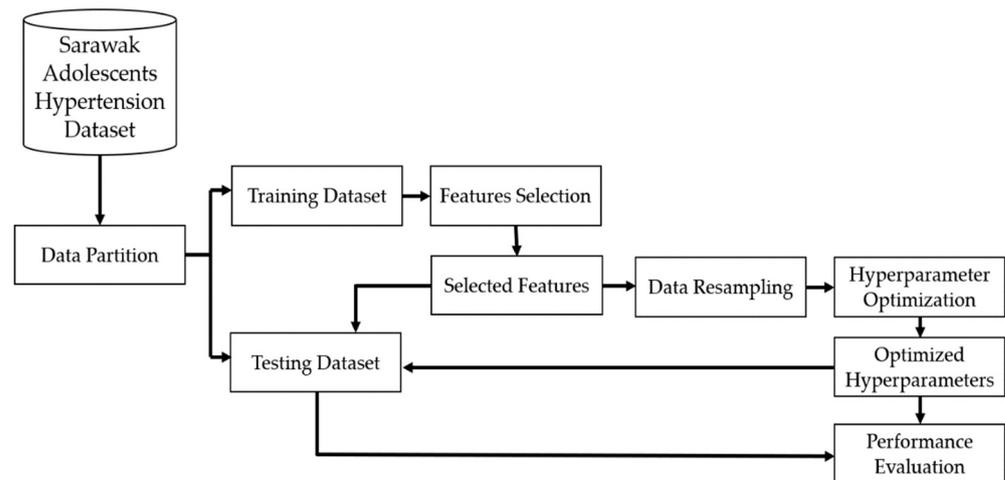


Figure 1. Overall processes involved in the study.

2.2.1. Data Partitioning

The ability of supervised classification algorithms to generalize new, previously unseen datasets is limited by the dataset's similarity to the available training data [26,27]. Prior to the data partitioning procedure, the pre-hypertension and hypertension groups were grouped together. As a consequence, a binary classification with predictions on whether a target is normal or hypertensive was developed. The dataset contains 69.9% to 30.1% normal to hypertensive participants, respectively. The original dataset was processed using SAS Visual Data Mining and Machine Learning software. Stratified sampling was employed to split the data into 90% for training (2215) and 10% (246) for testing based on hypertension and normal groups. Table 2 shows the distribution of hypertensive and normal groups in the training and testing data.

Table 2. Distribution of hypertensive and normal participants in the training and testing dataset.

Category	Training	Testing
Hypertensive	667 (30.1%)	74 (30.1%)
Normal	1548 (69.9%)	172 (69.9%)

2.2.2. Correlation Feature Selection

Irrelevant or redundant features confuse the machine learning algorithm, resulting in poor learning and mining outcomes [28,29]. Feature selection is beneficial in improving learning efficiency and predictability by deleting unnecessary or duplicated information [30]. In our study, the filter method, which is independent of any learning algorithm, is used for feature selection. It is less computationally demanding than the wrapper technique. Using the correlation coefficient, the dependency of the features in this study is determined in order to remove the redundant features. High correlation features (either positive or negative) are more linearly dependent and so have approximately the same influence on the dependent variable. When two features have a strong correlation, one of them can be dropped. Values closer to zero indicate that there is no linear relationship between the two features, while values closer to one suggest strong linear relationships, with +1 indicating that the two features are positively correlated and −1 indicating that they are negatively correlated.

2.2.3. Data Resampling

Machine learning classifiers designed to optimize total accuracy throughout the whole dataset have resulted in a stronger emphasis on the majority class, leading to poor prediction for the minority class [31]. A dataset with nearly identical numbers of instances

in each class, i.e., similar prior probabilities of the target classes, could aid in overcoming this problem. In the real world, however, class imbalance, with one of the target values having a significantly smaller number of instances, is common, particularly in the field of health. The machine learning classification model's performance is heavily influenced by the underlying imbalance dataset [31,32]. In order to deal with the imbalanced dataset problem, the training data are balanced using the Synthetic Minority Oversampling Technique (SMOTE). SMOTE will generate new instances of the minority class using interpolation [33]. With solely SMOTE, although the number of instances in the minority class is increased to the same as the majority class, the distribution of the dataset is still skewed [34]. The combination of oversampling and undersampling approaches had shown to be an effective solution to this [35]. In our study, we examine the usage of SMOTE alone and the combination of SMOTE with random undersampling on the prediction results before deciding on the resampling approach. To assess the impact of the resampling approaches, the Logistic Regression model was utilized.

2.2.4. Machine Learning Models

We developed and tuned thirteen ML models of three categories: neural network (Multilayer Perceptron), classical model (Logistic Regression, Decision Tree, Naïve Bayes, k-Nearest Neighbor), and ensemble model (Random Forest, Support Vector Machine, Gradient Boosting, XGBoost, LightGBM, CatBoost, AdaBoost and LogitBoost) to predict hypertension. Python was used throughout the implementation. The hyperparameters of the models were established during training using grid search and 10-fold cross-validation on resampled data.

- Logistic Regression

This is the most basic and widely used machine learning model for binary classification, which may easily be extended to multi-label classification problems. The Logistic Regression technique utilizes the sigmoid function to construct a regression model that predicts the chance that an input belongs to a particular category.

- Decision Tree

This technique repeatedly separates the dataset according to an optimal data separation criterion, resulting in a tree-like structure [36]. The most popular splitting criteria used include the Information Gain, Gini index and Gain ratio. Splitting aims to create pure nodes, i.e., to reduce the impurity of a node.

- Random Forest

Random Forest is a Decision Tree-based ensemble learning model composed of multiple Decision Trees. Each Decision Tree in the Random Forest will eventually yield a leaf node. The Random Forest makes predictions based on the output chosen by the majority of Decision Tree leaf nodes.

- Support Vector Machine (SVM)

The SVM method classifies data by building a multidimensional hyperplane that best separates two classes by finding the maximum margin between two data clusters. This approach achieves a high level of discrimination by changing the input space into a multidimensional space through the use of unique nonlinear functions referred to as kernels.

- Naïve Bayes

Naïve Bayes is a Bayes' Theorem-based probabilistic algorithm. It asserts that the presence of one feature in a class does not influence the presence of any other features.

- k-Nearest Neighbor (kNN)

A kNN algorithm assumes that similar objects are close to each other. The similarity is expressed in kNN by determining the distance between two points on a graph. The

classification of a data point is based on the majority vote of its k-Nearest Neighbor in a distance function.

- Multilayer Perceptron (MLP)

An MLP is a feedforward artificial neural network (ANN) that has input, hidden, and output layers. The input layer accepts signals, whereas the output layer classifies or predicts them. There is an arbitrary number of hidden layers between the input and output layers. These hidden layers are the MLP's true computational engine. The non-linear activation function is used in the hidden layers and the output layer. The backpropagation procedure is used to train MLP models.

- Gradient Boosting

This is an ensemble learning model that makes predictions by combining many weak learning models, often Decision Trees. Gradient Boosting works by iteratively reducing the loss function by selecting a function that points to a negative gradient, i.e., a weak hypothesis.

- XGBoost

XGBoost is the abbreviation for eXtreme Gradient Boosting developed by Chen et al. [37]. It is a Decision Tree-based ensemble machine learning model that leverages the Gradient Boosting framework. Using parallel boosting, a new Decision Tree model is added to compensate for the weaknesses of the previous model. It is designed for speed and performance.

- LightGBM

LightGBM, or Light Gradient Boosting Machine, is a Gradient Boosting framework comparable to XGBoost that employs Decision Tree learning approaches. Microsoft designed it in 2017 [38] for increased speed. The attributes are sorted and categorized into bins using a histogram-based Decision Tree learning method, and the leaves are grown leaf-wise, yielding improved efficiency and memory usage advantages over XGBoost.

- CatBoost

CatBoost, short for Category Boosting, as with XGBoost and LightGBM, is based on a Gradient Boosting framework that uses the Decision Tree learning technique. Developed in 2017 [39], CatBoost attempts to solve categorical features using permutation techniques.

- AdaBoost

AdaBoost or Adaptive Boosting is the first practical boosting algorithm by Freund et al. [40]. This is an ensemble machine learning algorithm in which weights are allocated adaptively, with greater weights assigned to incorrectly classified instances. When the subsequent learners are grown from the previously grown learners, the weak learners are transformed into strong learners.

- LogitBoost

LogitBoost is a variation of AdaBoost. The LogitBoost method was developed as an alternative to AdaBoost in order to solve the shortcomings of AdaBoost in dealing with noise and outliers.

2.2.5. Performance Metrics

The predictions of the models developed in this study can generate four possible outcomes: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Positive individuals in our study had hypertension, whereas negative patients had normal blood pressure. TP and TN are correct predictions. FP outcomes are positive predictions when they are actually negative. On the other hand, FN outcomes are predictions that are negative when they are actually positive. We evaluate the prediction models with the following performance metrics:

- Sensitivity or recall or TP rate—This metric indicates the proportion of true positives predicted out of all positives in a dataset:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

- Specificity or False Negative rate—This is the number of negative cases that are mistakenly identified as positive:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2)$$

- Accuracy—This is a metric used to determine how many correct predictions a model produced through the whole test dataset:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

- Precision—This indicates the correctness of the correct prediction:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

- *F1*-score—This metric measures the accuracy of the model based on sensitivity and precision. A higher *F1*-score value indicates the model is more accurate:

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

- Misclassification rate or classification error—a performance indicator that indicates the proportion of incorrect predictions without differentiating between positive and negative predictions:

$$\text{Misclassification rate} = 1 - \text{Accuracy} \quad (6)$$

- Area Under the Receiver Operating Characteristic Curve (AUC) is used to assess the model's prediction accuracy. This statistic assesses the algorithm's ability to distinguish between hypertensive and normal individuals.

3. Results

3.1. Features Selection

The original dataset consists of 11 features: age, gender, location (urban or rural), BMI, C index, ethnicity, height, weight (Wt), waist circumference (WC), parental hypertension, and weight-to-height ratio (WHtR). Figure 2 illustrates the feature dependency through the use of a correlation heatmap. It is observed that BMI, WC, WHtR and Wt have very high features dependency. Table 3 contains the correlation coefficients for these features.

When BMI was removed from the dataset, there was a strong correlation between Wt, WHtR, and WC, as seen in Table 4. This resulted in the subsequent removal of WHtR and Wt, and the resulting correlation coefficients heatmap is presented in Figure 3. The following features were chosen by the feature selection process: age, C index, ethnicity, gender, height, location, parental hypertension, and waist circumference.

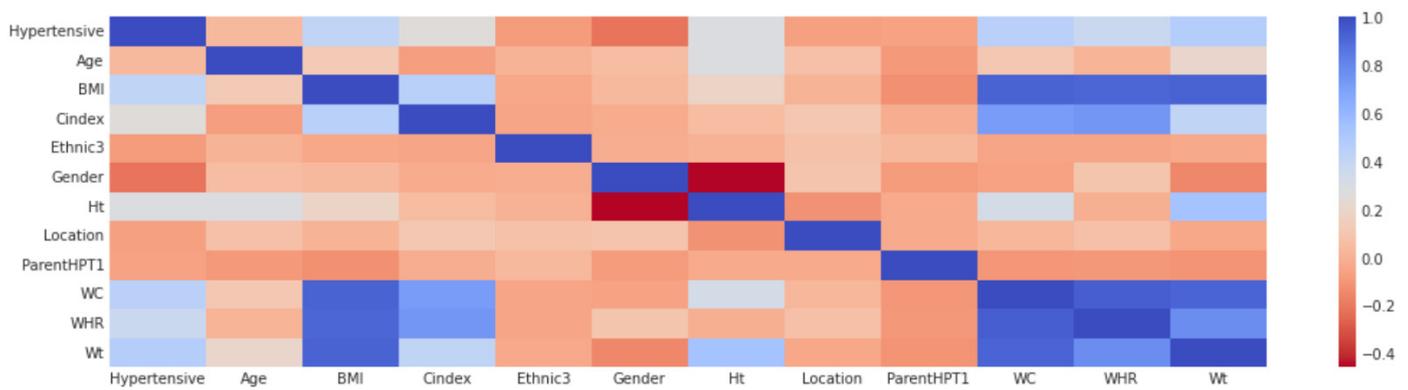


Figure 2. Heatmap showing the dependency of the 11 features in the study: hypertensive, age, BMI, C Index (Cindex), ethnicity (Ethnic3), gender, height (Ht), location, parental hypertension (ParentHPT1), waist circumference (WC), weight-to-height ratio (WHR) and weight (Wt).

Table 3. Correlation coefficients of BMI with WC, WHtR and weight (Wt).

	WC	WHtR	Wt
BMI	0.9248	0.9121	0.9219

Table 4. Correlation coefficients of WC with WHtR and weight (Wt).

	WHtR	Wt
WC	0.9416	0.9163

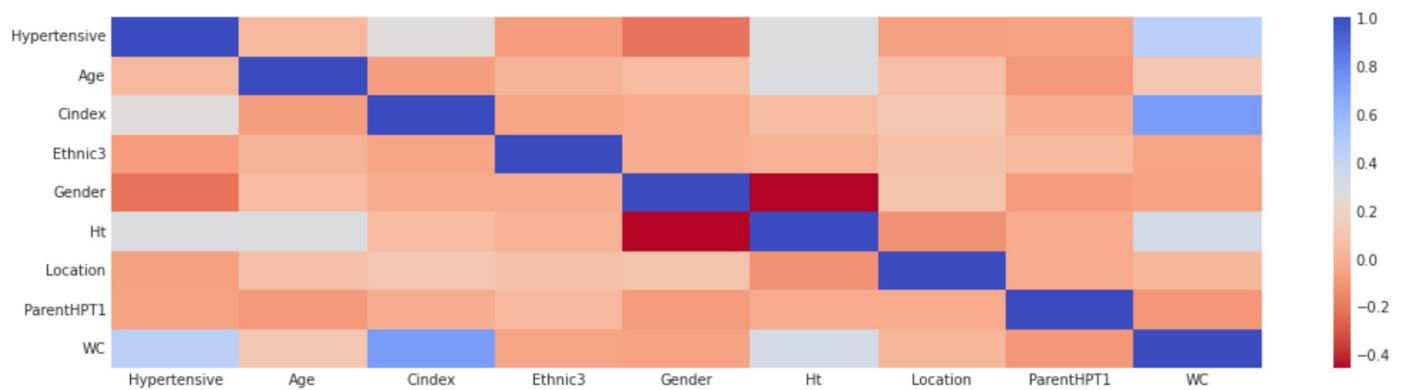


Figure 3. Heatmap showing the correlation coefficient of the 8 selected features: hypertensive, age, C Index (Cindex), ethnicity (Ethnic3), gender, height (Ht), location, parental hypertension (ParentHPT1), and waist circumference (WC).

3.2. Data Resampling

The original dataset was resampled using SMOTE and SMOTE with random under-sampling to address the data imbalance problem. The original training dataset included 1548 healthy adolescents and 667 hypertensive teenagers. This dataset is enlarged to an equal number of normal and hypertensive adolescents. SMOTE and SMOTE with random undersampling both resulted in a balanced dataset of 1548 samples for the healthy and hypertensive categories respectively. We compared the effects of SMOTE and SMOTE with random undersampling using Logistic Regression. To assess the impact of data resampling, we also conducted training and testing using the original imbalanced dataset with the Logistic Regression model.

When the model was trained using the original imbalanced dataset, the disparity between sensitivity and specificity was significant, as seen in Table 5. This is due to the

dataset's imbalance, which causes the model to focus on the majority group, i.e., normal samples. The SMOTE resampling process produced a more balanced dataset, which resulted in more balanced sensitivity and specificity values. It should be noted that the SMOTE resampled dataset's specificity was lower than the original imbalanced dataset, while the sensitivity values improved. When comparing the combination of SMOTE with the random undersampling process to the SMOTE alone approach, the sensitivity levels declined throughout training. The testing dataset, on the other hand, preserved the same value for this metric as the SMOTE resampled data, with greater specificity. With this observation, the study would train and test the 13 ML models employing data resampled with the combination of SMOTE and random undersampling procedure.

Table 5. Effects of using SMOTE and SMOTE with random sampling with a Logistic Regression model.

	Original Dataset		SMOTE		SMOTE with Random Undersampling	
	Training	Testing	Training	Testing	Training	Testing
Accuracy	0.7035	0.6626	0.7230	0.6626	0.7302	0.6748
Precision	0.5075	0.4545	0.7246	0.4563	0.7581	0.4700
Sensitivity	0.5667	0.6081	0.7194	0.6351	0.6763	0.6351
Specificity	0.7626	0.6860	0.7266	0.6744	0.7842	0.6919
F1-score	0.5354	0.5202	0.7220	0.5311	0.7148	0.5402
Misclassification Rate	0.2965	0.3374	0.2770	0.3374	0.2698	0.3252
AUC	0.6646	0.6471	0.7158	0.6548	0.7302	0.6635

3.3. Machine Learning Models

The hyperparameters of the machine learning models should be configured to tailor the machine learning models to the dataset. In most cases, the effects of hyperparameters on a model are well understood. The task of selecting the appropriate collection of hyperparameters and combinations of interacting hyperparameters for the dataset, on the other hand, is difficult. In this work, the grid search procedure is used to objectively explore multiple values for the machine learning model hyperparameters and choose a subset that results in a model that achieves the greatest performance on the dataset used. Grid search works by doing a full search across a specified subset of the training algorithm's hyperparameter space. Table 6 presents the optimal hyperparameters discovered using grid search for each model.

Table 6. Optimal hyperparameters obtained for each model using grid search.

Model	Hyperparameters
Logistic Regression	C: 2.0, penalty: l2, solver: liblinear
LogitBoost	C: 1.0
MLP	activation: tanh, alpha: 0.0001, hidden_layer_sizes: (50,), solver: adam
AdaBoost	learning_rate: 1, n_estimator: 200
CatBoost	depth: 13, iterations: 200, learning_rate: 0.1
LightGBM	learning_rate: 0.05, n_estimators: 350, num_leaves: 7
XGBoost	learning_rate: 0.1, max_depth: 15, n_estimators: 250
Gradient Boosting	learning_rate: 0.15, n_estimators: 100
kNN	n_neighbors: 1
Naïve Bayes	var_smoothing: $2.84803586 \times 10^{-6}$
SVM	C: 0.1, gamma: scale, kernel: linear
Decision Tree	max_depth: 13, min_samples_leaf: 3, min_samples_split: 2
Random Forest	max_depth: 25, min_samples_leaf: 1, n_estimators: 55

The trained model was validated against the testing dataset using the best hyperparameters discovered via grid search. To make the comparison clearer, Table 7 shows the results obtained during training, whereas Table 8 shows the results of the trained model

using the testing dataset. According to these two tables, the models perform better during the training stage with 10-fold cross-validation than during the model testing stage.

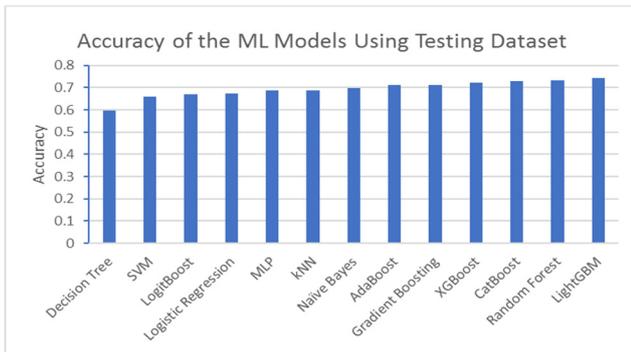
Table 7. Performances of the model during training.

	Accuracy	Precision	Sensitivity	Specificity	<i>F1</i> -Score	Misclassification Rate	AUC
Logistic Regression	0.7302	0.7581	0.6763	0.7842	0.7148	0.2698	0.7302
LogitBoost	0.7518	0.7869	0.6906	0.8129	0.7356	0.2482	0.7518
MLP	0.7446	0.7361	0.7626	0.7266	0.7491	0.2554	0.7446
AdaBoost	0.7986	0.8374	0.7410	0.8561	0.7863	0.2014	0.7986
CatBoost	0.8381	0.8357	0.8417	0.8345	0.8387	0.1619	0.8381
LightGBM	0.8165	0.8548	0.7626	0.8705	0.8061	0.1835	0.8165
XGBoost	0.8165	0.8235	0.8058	0.8273	0.8145	0.1835	0.8165
Gradient Boosting	0.8273	0.8760	0.7626	0.8921	0.8154	0.1727	0.8273
kNN	0.8201	0.7764	0.8993	0.7410	0.8333	0.1799	0.8201
Naïve Bayes	0.6942	0.7647	0.5612	0.8273	0.6473	0.3058	0.6942
SVM	0.7518	0.7917	0.6835	0.8201	0.7336	0.2482	0.7518
Decision Tree	0.7266	0.7203	0.7410	0.7122	0.7305	0.2734	0.7266
Random Forest	0.8273	0.8421	0.8058	0.8489	0.8235	0.1727	0.8273

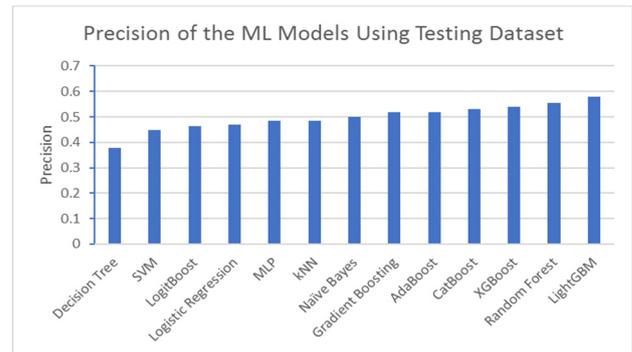
Table 8. Performances of the models using the optimum hyperparameters obtained.

	Accuracy	Precision	Sensitivity	Specificity	<i>F1</i> -Score	Misclassification Rate	AUC
Logistic Regression	0.6748	0.4700	0.6351	0.6919	0.5402	0.3252	0.6635
LogitBoost	0.6707	0.4646	0.6216	0.6919	0.5318	0.3293	0.6567
MLP	0.6870	0.4835	0.5946	0.7267	0.5333	0.3130	0.6607
AdaBoost	0.7114	0.5190	0.5541	0.7791	0.5359	0.2886	0.6666
CatBoost	0.7295	0.5316	0.5676	0.7849	0.5490	0.2805	0.6762
LightGBM	0.7439	0.5797	0.5405	0.8314	0.5594	0.2561	0.6860
XGBoost	0.7236	0.5395	0.5541	0.7965	0.5467	0.2764	0.6753
Gradient Boosting	0.7114	0.5185	0.5676	0.7733	0.5419	0.2886	0.6704
kNN	0.6870	0.4842	0.6216	0.7151	0.5444	0.3130	0.6684
Naïve Bayes	0.6992	0.5000	0.5811	0.7500	0.5375	0.3008	0.6655
SVM	0.6585	0.4468	0.5676	0.6977	0.5000	0.3415	0.6326
Decision Tree	0.5976	0.3786	0.5270	0.6279	0.4407	0.4024	0.5775
Random Forest	0.7317	0.5541	0.5541	0.8081	0.5541	0.2683	0.6811

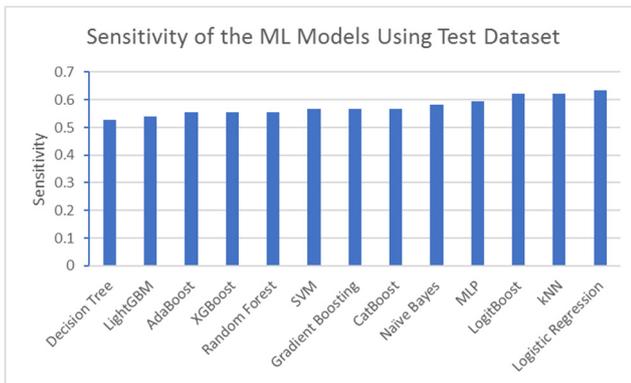
Because this work used a hold-out technique, the applicability of the model should be established by the findings obtained when the models were tested using the testing dataset. Figure 4 illustrates the performance of each machine learning model for each performance metric when the trained models were assessed using the testing dataset. The accuracy, precision, specificity, *F1*-score, misclassification rate, and AUC of LightGBM are the greatest. LightGBM, on the other hand, performs badly in the sensitivity test. Except for the sensitivity measure, Random Forest was the second-best model across all six performance metrics. On the other side, Decision Tree was the least performant machine learning model across all seven performance metrics. XGBoost exceeds CatBoost in terms of precision, specificity, and misclassification rate, whereas CatBoost outperforms XGBoost in terms of accuracy, *F1*-score, and AUC. The other machine learning models do not exhibit a consistent pattern in performance ranking. When AUC was used to compare LightGBM to Random Forest, CatBoost, and XGBoost, the differences were 0.005, 0.010, and 0.011, respectively.



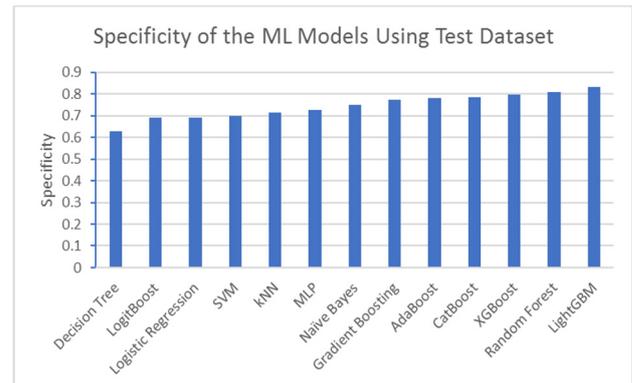
(a)



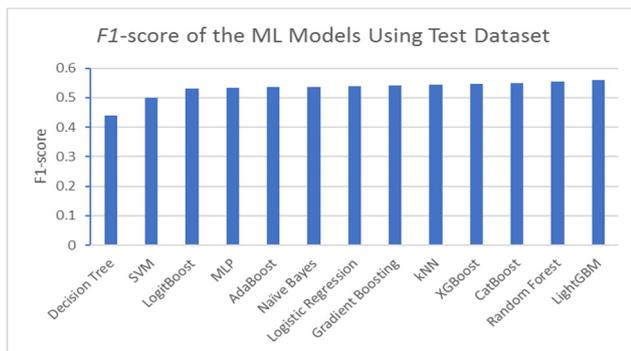
(b)



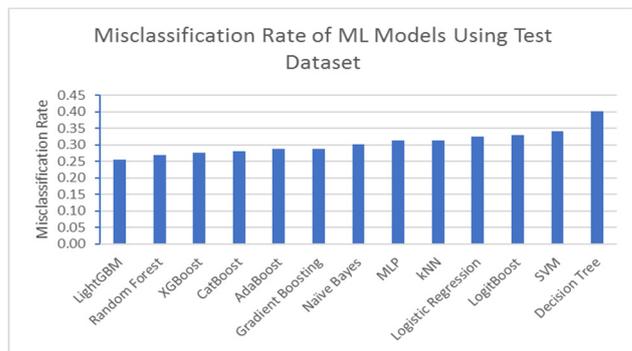
(c)



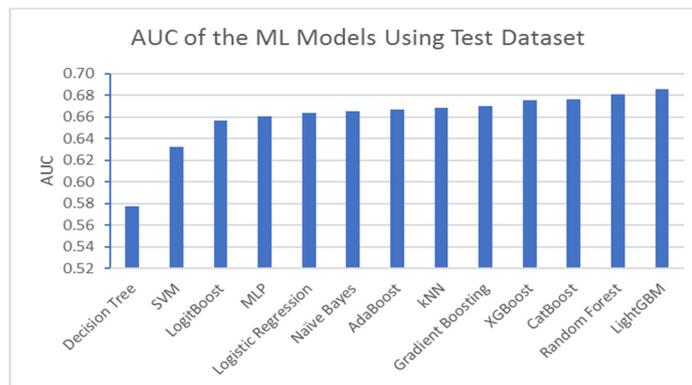
(d)



(e)



(f)



(g)

Figure 4. Bar graphs showing the performances of the trained models for the different performance measures: (a) accuracy, (b) precision, (c) sensitivity, (d) specificity, (e) *F1*-score, (f) misclassification rate, and (g) AUC.

4. Discussion

Machine learning techniques are increasingly being used to predict hypertension. However, the models’ comparability and effectiveness in real-world applications have been hampered by the incorporation of diverse features and learning techniques. In this work, we examined the possibility of employing simple anthropometric factors to predict hypertension. Additionally, we examined the effects of various machine learning approaches on this prediction using only basic anthropometric data. The use of simple anthropometric measures promises a simple, straightforward, affordable, and practical technique of predicting hypertension, particularly when a blood pressure monitor is not accessible. We defer the use of additional physiological data, such as blood pressure and heart rate, as well as self-reported lifestyle parameters, for the sake of simplicity and objective input features.

In terms of machine learning models, none of the 13 ML models scored well across all seven performance metrics: accuracy, precision, sensitivity, specificity, *F1*-score, misclassification rate, and AUC. Except for sensitivity, LightGBM, Random Forest, CatBoost, and XGBoost are the four leading models in the six performance criteria. The performance rankings of these four models, on the other hand, are inconsistent across these six performance measures. On the other hand, the Decision Tree has the lowest performance across all performance measures. In this study, we investigated three different types of supervised learning algorithms: neural networks, ensemble models, and classical models. The models in each of these categories are listed in Table 9. The kNN model beat the other classical models in terms of AUC and *F1*-score but lagged behind the Naïve Bayes model in terms of accuracy, precision, specificity, and misclassification rate. The Logistic Regression model performed the best in terms of sensitivity. While ensemble models are known to produce more accurate results than classical models, their performances for certain types of these models are not as good as classical models. AdaBoost, for example, falls short of kNN in terms of AUC. On the other hand, the MLP model performs modestly across the board, although it is an interesting model for sensitivity.

Table 9. Different learning models in the three different supervised machine learning categories.

Category	Model
Neural Network	Multi-layer Perceptron (MLP)
Ensemble Model	LogitBoost, AdaBoost, Gradient Boosting, Random Forest, XGBoost, CatBoost, LightGBM
Classical Model	SVM, Decision Tree, Logistic Regression, Naïve Bayes, kNN

The results revealed that each model outperformed the others in terms of the numerous performance metrics used. Because of this, selecting the most appropriate model for practical application might be challenging. For the purpose of selecting a realistic model, we propose that Bayes’ Theorem be used to verify the model’s applicability before selecting it. Using Bayes’ Theorem, we can determine how well a model performs in a particular population when the prevalence of a specific condition is taken into consideration [19]. Bayes’ Theorem is a calculation of the posterior probability based on the mathematical formula shown in (7):

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)} \tag{7}$$

P(A): prevalence of adolescent hypertension in Sarawak population = 0.301

P(B): probability of the model returning positive

P(B | A): probability of event B given event A occurring

Table 10 presents a summary of the findings acquired through the use of Bayes’ Theorem. From this table, the top three performing ML models are LightGBM (0.5799), Random Forest (0.5542), and XGBoost (0.5397), whereas the bottom three are Decision Tree (0.3788), SVM (0.4471), and LogitBoost (0.4649). With the highest performing model, LightGBM predicts an adolescent in the Sarawak adolescent population of 200,130, with a

hypertension prevalence of 30.1%, will have a 57.99% chance of being hypertensive if he or she is predicted as hypertensive using this model. If Decision Tree was chosen as the prediction model, an adolescent who is predicted to be hypertensive has a 37.88% chance of being hypertensive.

Table 10. Bayes' Theorem values in ascending order for the 13 ML models.

Model	Bayes' Theorem Value
Decision Tree	0.3788
SVM	0.4471
LogitBoost	0.4649
Logistic Regression	0.4702
MLP	0.4837
kNN	0.4844
Naïve Bayes	0.5002
Gradient Boosting	0.5188
AdaBoost	0.5193
CatBoost	0.5319
XGBoost	0.5397
Random Forest	0.5542
LightGBM	0.5799

Although the study discovered that applying Bayes' Theorem to the prediction of hypertension in adolescents using simple anthropometric data results in only a moderate level of reliability for the best performing machine learning model, this value is only applicable to the Sarawak adolescent population, which has a hypertension prevalence of 30.1%. Additionally, the predictive ability of anthropometric measurements for hypertension varies by nation and ethnicity [13], hence, concluding that the use of basic anthropometric data is usually inapplicable for ML hypertension prediction would be biased. Additionally, we would like to emphasize that, in contrast to other research in [14–16], our study utilized only simple anthropometric measures. A noteworthy finding from this study is that the disparities across models are discernible for the various models employed. As a result, when selecting a prediction model, it is critical to evaluate the appropriate model.

5. Conclusions

In this study, we managed to use simple anthropometric measurements for hypertension prediction in adolescents of the Sarawak population using 13 machine learning models. We had developed machine learning algorithms from three different supervised machine learning categories; namely, neural network, ensemble models and classical models. The feature dependency was evaluated using the correlation coefficient to eliminate redundant features. The original imbalanced dataset was resampled using SMOTE with random undersampling. While developing the ML models, grid search was used in order to find the optimal hyperparameters. The models were trained using 10-fold cross-validation using the resampled training dataset and the trained models were tested using the testing dataset. Seven performance metrics were used to evaluate the trained model.

According to the results of the study, the best-performing model was LightGBM, while the lowest-performing model was Decision Tree. Although the majority of the ensemble models outperformed the classical models, several ensemble models underperformed the classical models. We determined that the use of basic anthropometric measures for adolescents with hypertension in the Sarawak community is minor when using Bayes' Theorem. In other words, the model could not be utilized as a clinical decision-making tool to diagnose hypertension in adolescents in this population. The model, on the other hand, might serve as an early warning system for individuals who may be hypertensive, particularly when a blood pressure monitor is not available. We also showed that there is a considerable difference between the results obtained from the different prediction models used. Our study is valuable as it will pave the way for future researchers to provide a

better technique for generating a simple, inexpensive, straightforward, and reliable way to predict hypertension based on anthropometric measurement.

Author Contributions: The following is a breakdown of each author's contributions: data collection and analysis were completed by W.L.C.; results conceptualization and methodology; writing, reviewing, and editing were conducted by S.S.C., K.L.G., Y.H.R.C. and G.W.N. All authors have read and agreed to the published version of the manuscript.

Funding: The project is funded under the University of Malaysia Sarawak (UNIMAS) Cross Disciplinary Grant (F08/CDRG/1832/2019) and IMPACT Research Grant Scheme (F08/PARTNERS/2119/2021).

Institutional Review Board Statement: Ethics approval was obtained from the Medical Ethics committee of the University of Malaysia Sarawak (UNIMAS) [UNIMAS/NC-21.02/03-02(71)].

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data underlying the study's findings are obtainable upon reasonable request to the paper's co-author, Cheah Whye Lian, through email at wlcheah@unimas.my.

Acknowledgments: The authors would like to thank Faculty of Computer Science and Information Technology, University of Malaysia Sarawak for providing the facilities for this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

- World Health Organization. Noncommunicable Diseases. 13 April 2021. Available online: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases> (accessed on 13 April 2021).
- Tackling, G.; Borhade, M.B. Hypertensive Heart Disease. StatPearls, 2021. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK539800/> (accessed on 20 October 2021).
- Feber, J.; Ahmed, M. Hypertension in children: New trends and challenges. *Clin. Sci.* **2010**, *119*, 151–161. [CrossRef]
- Katamba, G.; Agaba, D.C.; Migisha, R.; Namaganda, A.; Namayanja, R.; Turyakira, E. Prevalence of hypertension in relation to anthropometric indices among secondary adolescents in Mbarara, Southwestern Uganda. *Ital. J. Pediatr.* **2020**, *46*, 76. [CrossRef] [PubMed]
- Mattoo, T.K. *Definition and Diagnosis of Hypertension in Children and Adolescents*; UpToDate: Waltham, MA, USA, 2009.
- Ewald, D.R.; Haldeman, L.A. Risk Factors in Adolescent Hypertension. *Glob. Pediatr. Health* **2016**, *3*, 2333794X15625159. [CrossRef] [PubMed]
- Liew, J.K.; Cheong, X.P.; Law, L.; Teo, W.H.; Eng, S.S.; Ngim, C.F.; Ramadas, A. Prevalence and factors associated with hypertension among adolescents in Malaysia. *IIUM Med. J. Malays.* **2019**, *18*, 55–64.
- Casadei, K.; Kiel, J. Anthropometric measurement. StatPearls, 2020. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK537315/> (accessed on 29 November 2021).
- Vaquero-Álvarez, M.; Molina-Luque, R.; Fonseca-Pozo, F.J.; Molina-Recio, G.; López-Miranda, J.; Romero-Saldaña, M. Diagnostic Precision of Anthropometric Variables for the Detection of Hypertension in Children and Adolescents. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4415. [CrossRef]
- Erdal, I.; Yalçın, S.S.; Aksan, A.; Gençal, D.; Kanbur, N. How useful are anthropometric measurements as predictive markers for elevated blood pressure in adolescents in different gender? *J. Pediatr. Endocrinol. Metab.* **2020**, *33*, 1203–1211. [CrossRef]
- Li, Y.; Zou, Z.; Luo, J.; Ma, J.; Ma, Y.; Jing, J.; Zhang, X.; Luo, C.; Wang, H.; Zhao, H.; et al. The predictive value of anthropometric indices for cardiometabolic risk factors in Chinese children and adolescents: A national multicenter school-based study. *PLoS ONE* **2020**, *15*, e0227954. [CrossRef]
- Prastowo, N.A.; Haryono, I.R. Elevated blood pressure and its relationship with bodyweight and anthropometric measurements among 8–11-year-old Indonesian school children. *J. Public Health Res.* **2020**, *9*, 1723. [CrossRef]
- Khader, Y.; Batieha, A.; Jaddou, H.; El-Khateeb, M.; Ajlouni, K. The performance of anthropometric measures to predict diabetes mellitus and hypertension among adults in Jordan. *BMC Public Health* **2019**, *19*, 1416. [CrossRef]
- Zhao, H.; Zhang, X.; Xu, Y.; Gao, L.; Ma, Z.; Sun, Y.; Wang, W. Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method. *Front. Public Health* **2021**, *9*, 1395. [CrossRef]
- Boutilier, J.J.; Chan, T.C.Y.; Ranjan, M.; Deo, S. Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis. *J. Med. Internet Res.* **2021**, *23*, e20123. [CrossRef]
- López-Martínez, F.; Núñez-Valdez, E.R.; Crespo, R.G.; García-Díaz, V. An artificial neural network approach for predicting hypertension using NHANES data. *Sci. Rep.* **2020**, *10*, 10620. [CrossRef]
- Amoah, E.M.; Okai, D.E.; Manu, A.; Laar, A.; Akamah, J.; Torpey, K. The Role of Lifestyle Factors in Controlling Blood Pressure among Hypertensive Patients in Two Health Facilities in Urban Ghana: A Cross-Sectional Study. *Int. J. Hypertens.* **2020**, *2020*, 9379128. [CrossRef]

18. Rimárová, K.; Dorko, E.; Diabelková, J.; Sulínová, Z.; Frank, K.; Baková, J.; Uhrin, T.; Makovický, P.; Pelechová, N.; Konrádyová, N. Anthropometric predictors of systolic and diastolic blood pressure considering intersexual differences in a group of selected schoolchildren. *Central Eur. J. Public Health* **2018**, *26*, S04–S11. [CrossRef]
19. Chai, S.S.; Cheah, W.L.; Goh, K.L.; Chang, Y.H.R.; Sim, K.Y.; Chin, K.O. A Multilayer Perceptron Neural Network Model to Classify Hypertension in Adolescents Using Anthropometric Measurements: A Cross-Sectional Study in Sarawak, Malaysia. *Comput. Math. Methods Med.* **2021**, *2021*, 2794888. [CrossRef]
20. Department of Statistics Malaysia. Available online: https://www.dosm.gov.my/v1/index.php?r=column/cone&menu_id=clJnWTITbWFHdmUwbmtSTE1EQStFZz09 (accessed on 10 December 2021).
21. Ghosh-Dastidar, M.B.; Haas, A.C.; Nicosia, N.; Datar, A. Accuracy of BMI correction using multiple reports in children. *BMC Obes.* **2016**, *3*, 37. [CrossRef]
22. Hsieh, S.D.; Muto, T. The superiority of waist-to-height ratio as an anthropometric index to evaluate clustering of coronary risk factors among non-obese men and women. *Prev. Med.* **2005**, *40*, 216–220. [CrossRef]
23. Passos, M.A.Z.; Vellozo, E.P.; Enes, C.C.; Hall, P.R.; Andrade, A.L.M.; da Silva, A.M.B.; Vitale, M.S.D.S.; Arcanjo, C.C.; Arcanjo, F.P.N. The Conicity Index Compared to Other Anthropometric Indicators as a Predictor of Excess Weight and Obesity in Adolescents. *Int. J. Health Sci. (IJHS)* **2021**, *9*, 38–49. [CrossRef]
24. Andrade, M.D.; De Freitas, M.C.P.; Sakumoto, A.M.; Pappiani, C.; De Andrade, S.C.; Vieira, V.L.; Damasceno, N.R.T. Association of the conicity index with diabetes and hypertension in Brazilian women. *Arch. Endocrinol. Metab.* **2016**, *60*, 436–442. [CrossRef]
25. National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics* **2004**, *114* (Suppl. S2), 555–576. [CrossRef]
26. Schat, E.; Van De Schoot, R.; Kouw, W.M.; Veen, D.; Mendrik, A.M. The data representativeness criterion: Predicting the performance of supervised classification based on data set similarity. *PLoS ONE* **2020**, *15*, e0237009. [CrossRef]
27. Chai, S.S.; Goh, K.L.; Chang, Y.H.R.; Sim, K.Y. Coupling Normalization with Moving Window in Backpropagation Neural Network (BNN) for Passive Microwave Soil Moisture Retrieval. *Int. J. Comput. Intell. Syst.* **2021**, *14*, 179. [CrossRef]
28. Wang, S.; Tang, J.; Liu, H. Feature Selection. In *Encyclopedia of Machine Learning and Data Mining*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2017; pp. 503–511.
29. Jiang, S.-Y.; Wang, L.-X. Efficient feature selection based on correlation measure between continuous and discrete features. *Inf. Process. Lett.* **2016**, *116*, 203–215. [CrossRef]
30. Hsu, H.-H.; Hsieh, C.-W. Feature Selection via Correlation Coefficient Clustering. *J. Softw.* **2010**, *5*, 1371–1377. [CrossRef]
31. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [CrossRef]
32. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef]
33. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
34. Junsomboon, N.; Phienthrakul, T. Combining Over-Sampling and Under-Sampling Techniques for Imbalance Dataset. In Proceedings of the 9th International Conference on Machine Learning and Computing, Singapore, 24–26 February 2017; pp. 243–247.
35. Estabrooks, A.; Jo, T.; Japkowicz, N. A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Comput. Intell.* **2004**, *20*, 18–36. [CrossRef]
36. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: London, UK, 2017.
37. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
38. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Processing Syst.* **2017**, *30*, 3146–3154.
39. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *arXiv* **2017**, arXiv:1706.09516.
40. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]