Bisong *et al.*
BMC Medical Informatics and Decision Making (2024) 24:234
https://doi.org/10.1186/s12911-024-02634-9

BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Predicting high blood pressure using machine learning models in low- and middle-income countries

Ekaba Bisong[1*], Noor Jibril[2], Preethi Premnath[3], Elsy Buligwa[4], George Oboh[1] and Adanna Chukwuma[4,5*]

Abstract

Responding to the rising global prevalence of noncommunicable diseases (NCDs) requires improvements in the management of high blood pressure. Therefore, this study aims to develop an explainable machine learning model for predicting high blood pressure, a key NCD risk factor, using data from the STEPwise approach to NCD risk factor surveillance (STEPS) surveys. Nationally representative samples of adults aged 18-69 years were acquired from 57 countries spanning six World Health Organization (WHO) regions. Data harmonization and processing were performed to standardize the selected predictors and synchronize features across countries, yielding 41 variables, including demographic, behavioural, physical, and biochemical factors. Five machine learning models - logistic regression, k-nearest neighbours, random forest, XGBoost, and a fully connected neural network - were trained and evaluated at global, regional, and country-specific levels using an 80/20 train-test split. The models' performance was assessed using accuracy, precision, recall, and F1 score. Feature importance analysis identified age, weight, heart rate, waist circumference, and height as key predictors of blood pressure. Across the 57 countries studied, model performances varied considerably, with accuracy ranging from as low as 58.96% in some models for specific countries to as high as 81.41% in others, underscoring the need for region and country-specific adaptations in modelling approaches. The explainable model offers an opportunity for population-level screening and continuous risk assessment in resource-limited settings.

Keywords Noncommunicable diseases, Hypertension, Machine learning, Explainable models, Low- and middle-income countries, STEPwise approach, Risk factors, Blood pressure control, Global health, Public health, Decision support, Clinical management

Background

High blood pressure affects over 1.2 billion people globally, with two-thirds of them residing in low- and middle-income countries (LMICs) [1–3]. High blood pressure is a risk factor for premature death and disability due to cardiovascular diseases, stroke, and chronic kidney disease [4]. In low- and middle-income countries, the economic impact of high blood pressure and its related complications is significant, often exceeding per capita health expenditure multiple times. For example, the average cost of managing high blood pressure can range from $500 to $1500 per episode, starkly contrasting with the

*Correspondence:
Ekaba Bisong
ebisong@siliconblast.com
Adanna Chukwuma
adc785@mail.harvard.edu
[1] SiliconBlast Ltd., Calgary, AB, Canada
[2] Medtronic, Abu Dhabi, United Arab Emirates
[3] Department of Government Enablement, Abu Dhabi, United Arab Emirates
[4] World Bank, Washington, DC 20433, USA
[5] Health, Nutrition, and Population Global Practice, World Bank Group, Washington, DC 20433, USA

more modest health budgets in these regions [5]. This financial strain is compounded by the broader economic effects, including substantial productivity losses due to the disease.

While individual blood pressure measurements are relatively straightforward to obtain, conducting widespread screenings and maintaining long-term monitoring across large populations, especially in resource-limited settings, remains a significant challenge. This study aims to develop an explainable machine learning model for predicting blood pressure levels using demographic, lifestyle, and other data available across contexts.

An explainable machine learning model that predicts blood pressure across settings using data that can be collected virtually can serve multiple purposes. Firstly, it can function as an efficient initial screening tool to identify high-risk individuals to prioritise for direct intervention in resource-limited settings. Secondly, the model has the potential for early risk identification, potentially flagging individuals at high risk of developing hypertension in the future based on current risk factors. Furthermore, the integration of such a predictive model with existing health data systems could provide continuous risk assessment without requiring frequent direct measurements. Lastly, the model's ability to predict blood pressure trends over time could offer valuable insights into population health trajectories.

In light of these considerations, this study aims to develop an explainable machine learning model that predicts high blood pressure using clinical and demographic data from a large, diverse population across multiple LMICs.

Methods

The STEPS noncommunicable disease risk factor surveillance instrument

The STEPwise approach to noncommunicable disease (NCD) risk factor surveillance (STEPS) constitutes a standardized tool, designed for low- and middle- income countries (LMICs) to systematically collect, analyze, and disseminate data on key NCD risk factors [6–9]. This dataset encompasses behavioural risk factors, including tobacco and alcohol consumption, physical inactivity, and unhealthy dietary patterns, as well as biological risk factors such as overweight and obesity, high blood pressure, high blood glucose, and dyslipidemia.

STEPS employs a multistage cluster sampling methodology to generate a nationally representative sample of adults between the ages of 18 and 69 years. Data collection is conducted via in-person interviews with selected respondents at their residences. The survey comprises three distinct levels or "steps", as detailed in Table 1.

For the study, STEPS data from 57 countries spanning 6 WHO regions were acquired, as delineated in Fig. 1. 57 of the 71 countries comprising the WHO STEPS dataset that were obtained, are those who questionnaire are in English or could be easily converted to English. The country-level sample size ranged from 275 in Liberia to 9,183 in Ethiopia.

Data harmonization

For the harmonization of features of interest in predicting high blood pressure, we utilized author-led expert surveys and analyzed the STEPS questionnaire to identify relevant variables for predicting blood pressure. These variables were then adjusted for consistency across all participating countries based on the harmonization strategy outlined in our feature-engineering plan.

The harmonization process was critical in ensuring that the data collected from different countries could be accurately compared and analyzed. This involved aligning variable definitions, categories, and measurement techniques across different sections of the STEPS questionnaire:

1. Demographic information: We condensed earnings-related questions into a single "Earnings per year" variable.

Table 1 The STEPS survey encompasses distinct levels of risk-factor assessment [10]

Level	Collection Method	Risk factors
Step 1	Collected via self-report	**Demographics and behavioural risk factors:** The scope of inquiry includes tobacco and alcohol consumption, dietary habits such as fruit and vegetable intake, and salt/sodium ingestion, physical inactivity, medical history of NCDs and related conditions, such as hypertension, diabetes, dyslipidemia, and cardiovascular diseases. Additional factors encompass cervical cancer screening coverage in women and the provision of comprehensive lifestyle guidance to mitigate NCD risks..
Step 2	Collected via self-report	**Physical measurements:** Assessment includes height and weight measurements for body mass index (BMI) determination, calculated as weight in kilograms divided by the square of height in meters. Additionally, waist circumference and blood pressure evaluations form part of the physical measurements.
Step 3	Conducted at local clinics or health centers	**Biochemical measurements:** The analysis encompasses fasting blood glucose, total cholesterol levels, and urinary sodium concentrations, providing essential insights into the biochemical aspects of NCD risk factors.

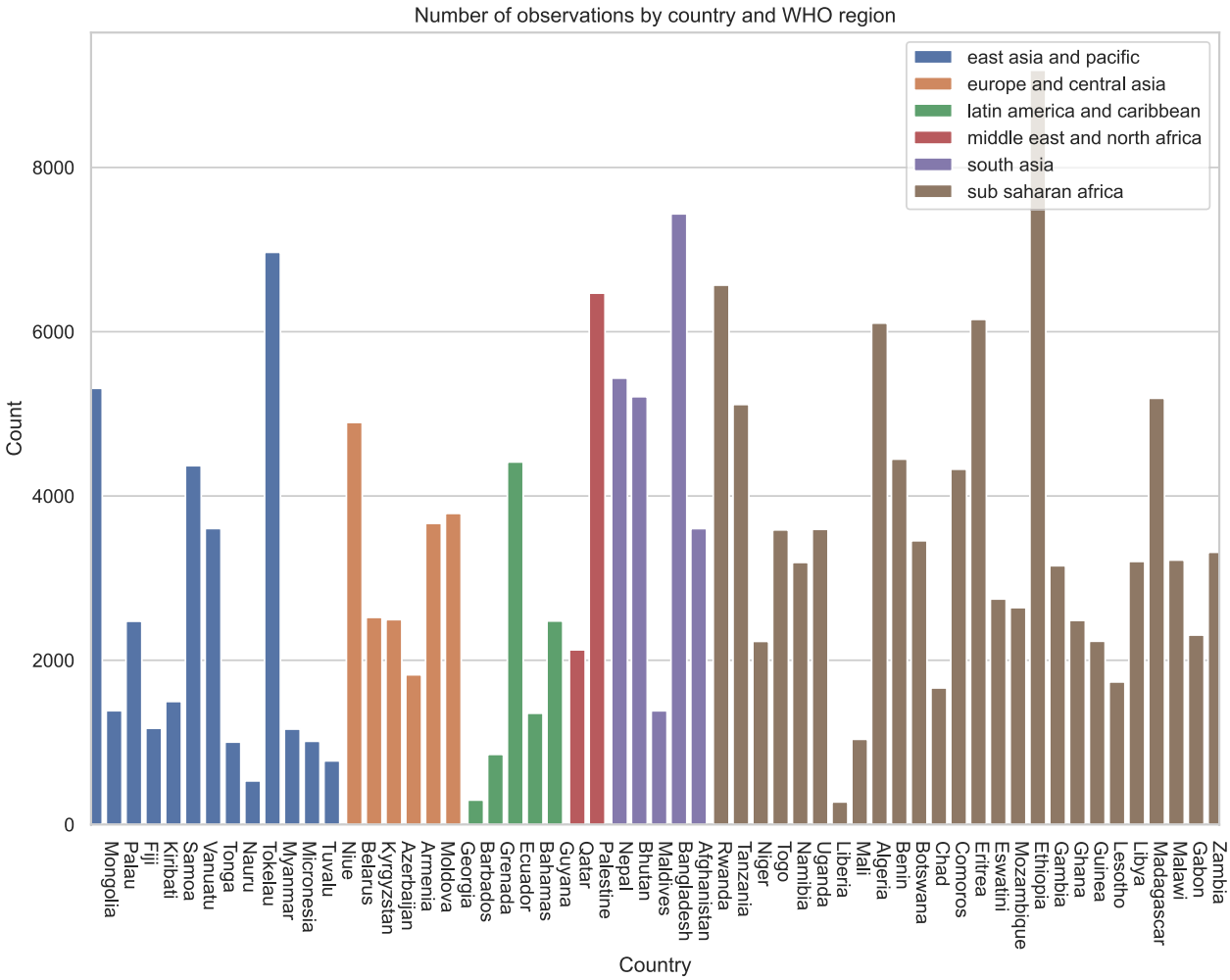

Fig. 1 LMIC countries and their sample sizes in the STEPS dataset

2. Tobacco Use: We merged several questions to create more concise variables, such as "Length of time smoking" and "Number of tobacco products per day." We also consolidated different types of tobacco products, recognizing their similar health risks.
3. Alcohol Consumption: We condensed frequency and quantity questions into more manageable variables like "How often do you drink alcohol?" and "How many alcoholic drinks do you consume per day?"
4. Diet: We consolidated fruit and vegetable consumption questions into a single "How many fruit/vegetables do you eat per day?" variable.
5. Physical Activity: We simplified work intensity and physical activity questions to capture key information more efficiently.
6. Medical History: We focused on key variables related to blood pressure, diabetes, cholesterol, and cardiovascular diseases, removing redundant or less relevant questions.
7. Physical Measurements: We retained key measurements like blood pressure readings, height, weight, waist circumference, and relevant biochemical measurements.

This harmonization strategy allowed us to create a more streamlined and consistent dataset across all countries, focusing on the most relevant predictors of blood pressure while reducing redundancy and potential inconsistencies in data collection across different settings.

The study dataset included 48 variables, as illustrated in Table 2. Among these variables, 11 represent demographic factors, including sex, age, years of schooling, educational level, marital status, and employment status. There are 24 variables associated with behavioural measurements, including factors like smoking habits, alcohol consumption, fruit and vegetable intake, work

Table 2 Variable descriptions

Variable name	Question	Variable type
Demographic Information - Step 1		
sex	Sex (Record Male / Female as observed)	Category
age	How old are you?	Numeric
number of years at school	In total, how many years have you spent at school and in full- time study (excluding pre-school)?	Numeric
level of education	What is the highest level of education you have completed?	Category
marital status	What is your marital status?	Category
work status	Which of the following best describes your main work status over the past 12 months?	Category
number of people in household	How many people older than 18 years, including yourself, live in your household?	Numeric
earnings per year	Taking the past year, can you tell me what the average earnings of the household have been?	Numeric
currently smoking tobacco	Do you currently smoke tobacco products?	Category
age started smoking	How old were you when you first started smoking?	Numeric
length of time smoking	Length of time smoking	Numeric
Behavioural Measurements - Step 1		
number of tobacco	On average, how many of the following products do you smoke each day?	Numeric
type of tobacco	What type of tobacco do you use?	Category
age stopped smoking	How old were you when you stopped smoking?	Numeric
smoking in the home or workplace	During the past 30 days, did someone smoke in your home or workplace?	Category
consume alcohol	Have you ever consumed any alcohol such as beer, wine, or spirits?	Category
quit drinking for health	Have you stopped drinking due to health reasons, such as a negative impact on your health or on the advice of your doctor or other health worker?	Category
number of alcoholic drinks	How many alcoholic drinks do you consume per day	Numeric
number daily fruit vegetables	How many fruit/ vegetables do you eat per day?	Numeric
salt consumption	How much salt or salty sauce do you think you consume?	Category
work intensity	How intense is your work?	Category
days vigorous exercise	In a typical week, on how many days do you do vigorous intensity activities as part of your work?	Numeric
days moderate exercise	In a typical week, on how many days do you do moderate intensity activities as part of your work?	Numeric
time walking bicycling minutes	How much time do you spend walking or bicycling for travel on a typical day?	Numeric
time sedentary	How much time do you usually spend sitting or reclining on a typical day?	Numeric
had blood pressure measurement	Have you ever had your blood pressure measured by a doctor or other health worker?	Category
taken drugs raised bp	Have you taken any drugs (medication) for raised blood pressure?	Category
blood sugar measurement	Have you ever had your blood sugar measured by a doctor or other health worker?	Category
taken diabetes drugs	Have you taken any drugs (medication) for diabetes?	Category
had cholesterol measurement	Have you ever had your cholesterol (fat levels in your blood) measured by a doctor or other health worker?	Category
taken cholesterol oral treatment	Have you taken any oral treatment (medication) for raised total cholesterol?	Category
had heart attack	Have you ever had a heart attack or chest pain from heart disease (angina) or a stroke (cerebrovascular accident or incident)?	Category
taking heart disease medication	Are you currently taking medication to prevent or treat heart disease?	Category
treated for raised bp	Have you been treated for raised blood pressure?	Category
are you pregnant	Are you pregnant?	Category
Physical measurements - Step 2		
height	Height in cm	Numeric
weight	Weight in kg	Numeric
waist circumference	Waist circumference in cm	Numeric
hip circumference	Hip circumference in cm	Numeric
reading bpm	Average readings beats per minute	Numeric

Table 2 (continued)

Variable name	Question	Variable type
Biochemical measurements - Step 3		
fasting blood glucose	Fasting blood glucose (mg/dl)	Numeric
total cholesterol	Total Cholesterol (mg/dl)	Numeric
urinary sodium	Urinary sodium (mg/dl)	Numeric
urinary creatinine	Urinary creatinine (mg/dl)	Numeric
triglycerides	Triglycerides (mg/dl)	Numeric
hdl cholesterol	HDL Cholesterol (mg/dl)	Numeric
systolic	Average systolic (mmHG)	Numeric
diastolic	Average diastolic (mmHG)	Numeric
Target variable		
blood pressure	Blood pressure readings	Category

intensity, and treatment for hypertension and heart disease. Five variables represent the physical measurements of the respondents, including the height, weight, waist, and hip circumference. Additionally, eight variables are related to the biochemical measurements of the respondents, including fasting blood glucose, cholesterol, urinary sodium, urinary creatinine, and the average systolic and diastolic measurements from three readings. The final study dataset comprises 27 numeric and 21 categorical variables in total.

Data processing

For numeric variables, the Z-score method was utilized to identify and remove outliers by setting a threshold value. Observations with a Z-score exceeding this threshold were considered outliers and were removed from the dataset [11, 12]. It must, however, be noted that, our approach only replaced extreme outliers, not all values below Q1 and Q3. This ensures that extreme values do not unduly influence the model training.

To handle categorical variables, a dictionary was created, containing mappings of categorical encodings for each variable. This dictionary facilitated the transformation of the categorical data into numerical values, aiding further analysis. Missing values within categorical columns were replaced with a designated "no response" category to ensure these instances were still accounted for in the dataset.

The creation of the target variable for the analysis involved first determining each person's blood pressure status by averaging their systolic and diastolic readings. We then used the average of the three blood pressure measurements taken during the STEPS survey to minimize variability and improve the reliability of our predictive models. Based on CDC [13] and AHA [14] guidelines, a comprehensive blood pressure classification system was created that considered both systolic and diastolic measurements. The criteria were:

1. If the systolic and diastolic readings were below 120 and 80 mm Hg, blood pressure was considered normal.
2. The "normal" classification was also given if the systolic reading was 120 to 129 mm Hg and the diastolic was below 80.
3. High blood pressure was defined as a systolic reading between 130 and 139 mm Hg or a diastolic reading between 80 and 89.
4. The "high" classification was also applied if either the systolic or diastolic reading was 140 or 90 mm Hg.
5. Finally, "high" status was given if the systolic or diastolic reading was 180 or 120 mm Hg.

A robust approach was applied to address nonsensical outliers in the dataset, thereby enhancing the reliability of the analysis. This approach involved the use of two complementary methods, aiming to identify and replace extreme values that lay outside the bounds of the upper and lower whiskers of the data distribution.

The first method was focused on handling outliers situated above the upper whisker. This boundary was determined by computing the third quartile (Q3) of the data distribution and adding a product of a predefined constant multiplier and the interquartile range (IQR). This is formally represented as $U = Q3 + k \times IQR$, where U is the upper boundary, $Q3$ is the third quartile, k is a predefined constant multiplier, and IQR is the interquartile range, calculated as $Q3 - Q1$. Upon identification of these upper-bound outliers, they were replaced with random numbers R_U that fell within the interquartile range, or alternatively, between the mean and Q3 (i.e., $\mu \leq R_U \leq Q3$), ensuring a more representative value in line with the general data distribution.

In a similar manner, the second method was aimed at outliers residing below the lower whisker. This boundary was established by calculating the first quartile (Q1) and subtracting a product of a predefined constant multiplier and the IQR. This is formally presented as $L = Q1 - k \times IQR$, where L is the lower boundary, $Q1$ is the first quartile, k and IQR are already defined. Once these lower-bound outliers were identified, they were replaced with random numbers either within the interquartile range or, alternatively, between Q1 and the mean of the data distribution (i.e., $Q1 \leq R_L \leq \mu$). This procedure ensured the replaced values were more harmonious with the overall data distribution, leading to a more accurate and reliable dataset for further analysis. For further reading on robust methods for handling outliers and their theoretical justification, see [15] and [16], which discuss the principles and application of these techniques in statistical analysis.

Model design and evaluation

A comprehensive approach was employed to predict blood pressure status, using multiple machine learning models. The methodology encompassed global, regional, and country-specific levels. This allowed for the tailoring of predictions to each level's unique characteristics.

The process began with data preprocessing. One-hot encoding was performed for categorical variables, a technique that converts categorical data into a binary matrix representation. This allows machine learning algorithms to work with categorical data in a numerical format. For example, a categorical variable "work status" with categories "employed," "unemployed," and "student" would be transformed into three binary columns.

Numerical variables were scaled using the StandardScaler method, which standardizes features by removing the mean and scaling to unit variance. This preprocessing step ensures that all numerical features contribute equally to the model and prevents features with larger magnitudes from dominating the learning process.

This preprocessing stage was crucial to ensure that the models could effectively learn from the data without being unduly influenced by the varying scales of different features. The dataset was split into training and testing subsets using the 'split' column, which was randomly assigned to each data point. This allowed for an unbiased evaluation of the models' performance on unseen data. Table 3 shows the data split for the countries.

At each level (global, regional, country), we applied a diverse set of machine learning algorithms to capture different aspects of the data and provide a comprehensive comparison. The chosen models represent a spectrum of approaches in machine learning:

Table 3 Country-wise test and train data distribution

Country	Test	Train	Country	Test	Train
Afghanistan	727	2876	Liberia	47	228
Algeria	1166	4939	Libya	616	2587
Armenia	348	1475	Madagascar	1021	4169
Azerbaijan	482	2013	Malawi	630	2590
Bahamas	242	1113	Maldives	279	1106
Bangladesh	1471	5964	Mali	204	834
Barbados	56	242	Micronesia	235	927
Belarus	979	3916	Moldova	696	2971
Benin	921	3528	Mongolia	992	4319
Bhutan	1088	4121	Mozambique	535	2106
Botswana	700	2755	Myanmar	1425	5542
Chad	357	1306	Namibia	635	2556
Comoros	856	3470	Nauru	196	808
Ecuador	893	3521	Nepal	1116	4319
Eritrea	1271	4879	Niger	445	1785
Eswatini	546	2200	Niue	153	622
Ethiopia	1864	7319	Palau	300	1086
Fiji	493	1982	Palestine	1286	5184
Gabon	453	1854	Qatar	431	1696
Gambia	667	2484	Rwanda	1385	5182
Georgia	804	2984	Samoa	312	1186
Ghana	507	1978	Tanzania	1011	4103
Grenada	181	673	Togo	730	2857
Guinea	449	1783	Tokelau	103	428
Guyana	513	1965	Tonga	681	2923
Kiribati	235	937	Tuvalu	216	798
Kyrgyzstan	488	2034	Uganda	727	2867
Lesotho	351	1385	Vanuatu	844	3525
Zambia	661	2654			

1. Logistic Regression: A linear model serving as a baseline and representing traditional statistical approaches [17].
2. K-Nearest Neighbours (KNN): A non-parametric method that can capture local patterns in the data [18].
3. Random Forest: An ensemble tree-based method known for handling non-linear relationships and interactions [19].
4. XGBoost: A gradient boosting algorithm that often achieves state-of-the-art performance in structured data problems [20].
5. Fully Connected Neural Network (FCNN): A deep learning approach capable of learning complex patterns and representations [21].

This selection allows us to compare linear (Logistic Regression) vs. non-linear (all others) models, tree-based ensemble methods (Random Forest, XGBoost) vs. other

approaches, and traditional machine learning (first four) vs. deep learning (FCNN) techniques. By including this diverse set, we aim to comprehensively evaluate different modeling paradigms and identify which approaches are most effective for blood pressure prediction across various geographical scales.

Model performance was evaluated using the testing subset, with metrics including accuracy, precision, recall, and F1 score. This multi-metric approach provides a holistic view of model performance, considering both the ability to correctly identify positive cases (precision and recall) and overall predictive accuracy.

For the global models, the entire dataset (comprising all countries) was used for training and evaluation. At the regional level, models were developed for each of the six WHO regions, allowing for a more nuanced understanding of factors influencing blood pressure status in different geographical areas. Country-specific models were trained using data from each individual country to capture unique aspects of blood pressure patterns, maximizing reliability and accuracy of results.

Class imbalance was addressed by computing class weights inversely proportional to class frequencies, which were then incorporated into model training. This ensured that the models were not biased towards the majority class and could effectively learn from minority classes.

By employing this comprehensive, multi-level approach to model design and evaluation, the study aimed to provide accurate and reliable predictions of blood pressure status at global, regional, and country-specific levels.

Model explainability

In this section, the interpretability of predictor variables in a random forest global model is investigated by assessing the feature importance of each respective variable. While machine learning literature often refers to this as "feature importance", it's crucial to distinguish this concept from the epidemiological notion of "risk factors associated with raised BP". Feature importance in a random forest model quantifies the statistical contribution of each variable to the model's predictive accuracy. A higher value indicates that the feature plays a more significant role in the model's predictions.

While feature importance can help identify variables that contribute substantially to the model's predictive power, it's important to note that high feature importance does not necessarily equate to clinical significance or causality in the context of blood pressure risk factors. Feature importance provides insights into the model's decision-making process but should be interpreted alongside clinical knowledge and epidemiological evidence.

In the present study, feature importance is computed through two complementary techniques: mean decrease in impurity (MDI) and feature permutation. MDI measures the total decrease in node impurity averaged over all trees of the forest, while feature permutation assesses the decrease in model performance when a feature's values are randomly shuffled.

Results

Study sample characteristics

The study dataset included 184,674 participants, with a mean age of 40.06 years and an average of 7.6 years of education. The average participant lived in a household with 3.01 members, and participants had an annual income of 1,727.08 USD.

On average participants started smoking at 18.65 years, consumed 7.62 tobacco products, and quit smoking at 29.87 years. They consumed 4.76 alcoholic drinks and had 10.91 servings of fruits and vegetables daily. Vigorous exercise occurred for 4.66 days and moderate exercise for 5.64 days per week. Participants spent 60.23 minutes walking or bicycling and 206.03 minutes sedentary daily.

The average height was 162.12 cm, with an average weight of 66.62 kg, waist circumference of 84.89 cm, and hip circumference of 95.89 cm. Mean fasting blood glucose was 39.67 mg/dL, total cholesterol 76.42 mg/dL, urinary sodium 121.13 mmol/L, and urinary creatinine 55.04 mg/dL. Triglycerides averaged 84.16 mg/dL, HDL cholesterol 17.67 mg/dL, systolic blood pressure 126.91 mmHg, diastolic blood pressure 80.27 mmHg, and resting heart rate 77.48 bpm.

The dataset was divided into training (n=147,739) and test (n=36,935) datasets for model development and validation, with similar characteristics in both sets, ensuring adequate representation (see Table 4).

In the study, 53.55% of participants were female, 36.40% had completed elementary school, 58.84% were married, 45.12% were employed. In addition, 80.00% did not smoke, 54.77% did not consume alcohol, salt consumption was reportedly normal among 29.11%, work intensity was vigorous-intensity 29.76%.

Blood pressure measurements were reported by 53.06% of the total population, while 80.60% did not respond about taking drugs for raised blood pressure. Blood sugar measurements were reported by 29.30% of the population, with 89.76% not responding about taking diabetes drugs. Cholesterol measurements were reported by 10.93% of the population, and only 0.07% reported taking oral cholesterol treatment. These characteristics are summarized in Table 5.

Table 4 Study sample characteristics between the train and test dataset for the numeric variables

Variables ($\mu \pm \sigma$)	Total Population (n=184674)	Train dataset (n=147739)	Test Dataset (n=36935)
Age	40.06 ± 13.27	40.08 ± 13.26	40.02 ± 13.31
Years at school	7.6 ± 5.33	7.6 ± 5.33	7.58 ± 5.32
People in household	3.01 ± 2.02	3.01 ± 2.02	3 ± 2.01
Earnings per year	1727.08 ± 1533.97	1734.25 ± 1538.65	1698.52 ± 1515
Age started smoking	18.65 ± 1.77	18.65 ± 1.77	18.63 ± 1.78
Length time smoking	7.38 ± 6.32	7.34 ± 6.33	7.52 ± 6.28
Number tobacco	7.62 ± 3.55	7.64 ± 3.47	7.57 ± 3.94
Age stopped smoking	29.87 ± 5.81	29.87 ± 5.82	29.88 ± 5.77
Number alcoholic drinks	4.76 ± 1.06	4.75 ± 1.06	4.76 ± 1.06
Number daily fruit vegetables	10.91 ± 6.73	10.91 ± 6.72	10.91 ± 6.75
Days vigorous exercise	4.66 ± 1.04	4.66 ± 1.04	4.67 ± 1.04
Days moderate exercise	5.64 ± 1.41	5.64 ± 1.41	5.64 ± 1.41
Time walking bicycling minutes	60.23 ± 34.33	60.33 ± 34.33	59.87 ± 34.32
Time sedentary	206.03 ± 172.04	205.89 ± 171.9	206.57 ± 172.59
Height	162.12 ± 10.29	162.12 ± 10.32	162.14 ± 10.17
Weight	66.62 ± 17.73	66.63 ± 17.73	66.59 ± 17.7
Waist circumference	84.89 ± 25.35	84.87 ± 25.13	84.98 ± 26.24
Hip circumference	95.89 ± 15.71	95.88 ± 15.7	95.9 ± 15.75
Fasting blood glucose	39.67 ± 37.09	39.6 ± 37.07	39.94 ± 37.17
Total cholesterol	76.42 ± 72.24	76.26 ± 72.23	77.06 ± 72.29
Urinary sodium	121.13 ± 32.8	121.09 ± 32.76	121.29 ± 32.95
Urinary creatinine	55.04 ± 38.3	55.06 ± 38.39	54.96 ± 37.93
Triglycerides	84.16 ± 23.99	84.13 ± 23.97	84.29 ± 24.08
Hdl cholesterol	17.67 ± 17.64	17.62 ± 17.64	17.87 ± 17.65
Systolic	126.91 ± 19.1	126.91 ± 19.09	126.89 ± 19.17
Diastolic	80.27 ± 11.7	80.28 ± 11.7	80.22 ± 11.71
Reading bpm	77.48 ± 12.32	77.48 ± 12.31	77.48 ± 12.33

Model performance

Performance globally

The global models were trained and evaluated on the entire dataset, encompassing all 57 countries from the six WHO regions. Table 6 presents the performance metrics for each model, including accuracy, F1 score, precision, and recall.

Among the five models, XGBoost achieved the highest accuracy of 68.52%, followed closely by the Fully Connected Neural Network (FCNN) with an accuracy of 68.25% and Logistic Regression with 68.20%. The Random Forest model performed comparably to Logistic Regression, with an accuracy of 68.14%. The K-Nearest neighbours (KNN) model had the lowest accuracy at 63.21%.

The F1 score, which is the harmonic mean of precision and recall, followed a similar trend to accuracy. XGBoost had the highest F1 score of 67.43%, while FCNN and Logistic Regression had scores of 67.66% and 67.19%, respectively. The Random Forest model had an F1 score of 66.67%, and KNN had the lowest score at 62.32%.

Precision, which measures the proportion of true positive predictions among all positive predictions, was highest for XGBoost at 67.85%, followed by FCNN at 67.64% and Logistic Regression at 67.51%. Random Forest had a precision of 67.56%, and KNN had the lowest precision at 62.39%.

Recall, which measures the proportion of true positive predictions among all actual positive instances, was highest for FCNN at 67.69%, followed by XGBoost at 67.27% and Logistic Regression at 67.06%. Random Forest had a recall of 66.51%, and KNN had the lowest recall at 62.28%.

The global model results demonstrate that XGBoost and FCNN consistently outperformed the other models across all performance metrics. Logistic Regression and Random Forest also showed competitive performance, while KNN had the lowest scores in all metrics.

Performance per region

The regional models were trained and evaluated on subsets of the dataset, each corresponding to one of the six

Table 5 Baseline characteristics between the train and test dataset for the categorical variables

Variable	Total population	Train dataset	Test dataset
Sex, *n*(%)			
female	98902.0 (53.55)	79050.0 (53.51)	19852.0 (53.75)
male	68096.0 (36.87)	54565.0 (36.93)	13531.0 (36.63)
no response	17676.0 (9.57)	14124.0 (9.56)	3552.0 (9.62)
level of education, *n*(%)			
elementary school	67226.0 (36.40)	53810.0 (36.42)	13416.0 (36.32)
high school	52337.0 (28.34)	41891.0 (28.35)	10446.0 (28.28)
no formal schooling	32767.0 (17.74)	26269.0 (17.78)	6498.0 (17.59)
tertiary	19993.0 (10.83)	15959.0 (10.80)	4034.0 (10.92)
no response	12351.0 (6.69)	9810.0 (6.64)	2541.0 (6.88)
marital status, *n*(%)			
married	108653.0 (58.84)	86865.0 (58.80)	21788.0 (58.99)
not married	29446.0 (15.94)	23569.0 (15.95)	5877.0 (15.91)
no response	21506.0 (11.65)	17250.0 (11.68)	4256.0 (11.52)
widowed	9677.0 (5.24)	7814.0 (5.29)	1863.0 (5.04)
cohabitating	6291.0 (3.41)	4984.0 (3.37)	1307.0 (3.54)
divorced	4977.0 (2.70)	3976.0 (2.69)	1001.0 (2.71)
separated	4124.0 (2.23)	3281.0 (2.22)	843.0 (2.28)
work status, *n*(%)			
employed	83332.0 (45.12)	66711.0 (45.15)	16621.0 (45.00)
homemaker	32866.0 (17.80)	26435.0 (17.89)	6431.0 (17.41)
unemployed	29313.0 (15.87)	23502.0 (15.91)	5811.0 (15.73)
no response	15891.0 (8.60)	12613.0 (8.54)	3278.0 (8.88)
student	13117.0 (7.10)	10396.0 (7.04)	2721.0 (7.37)
retired	10155.0 (5.50)	8082.0 (5.47)	2073.0 (5.61)
currently smoke tobacco, *n*(%)			
no	147737.0 (80.00)	118191.0 (80.00)	29546.0 (79.99)
yes	29116.0 (15.77)	23295.0 (15.77)	5821.0 (15.76)
no response	7821.0 (4.24)	6253.0 (4.23)	1568.0 (4.25)
type tobacco, *n*(%)			
no response	184414.0 (99.86)	147529.0 (99.86)	36885.0 (99.86)
cigarettes	238.0 (0.13)	192.0 (0.13)	46.0 (0.12)
shisha	19.0 (0.01)	16.0 (0.01)	3.0 (0.01)
cigars	3.0 (0.00)	2.0 (0.00)	1.0 (0.00)
smoke home workplace, *n*(%)			
no	111323.0 (60.28)	89033.0 (60.26)	22290.0 (60.35)
no response	37313.0 (20.20)	29897.0 (20.24)	7416.0 (20.08)
yes	36038.0 (19.51)	28809.0 (19.50)	7229.0 (19.57)
consumed alcohol, *n*(%)			
no	101138.0 (54.77)	80934.0 (54.78)	20204.0 (54.70)
yes	68693.0 (37.20)	54961.0 (37.20)	13732.0 (37.18)
no response	14843.0 (8.04)	11844.0 (8.02)	2999.0 (8.12)
quit drinking for health, *n*(%)			
no response	171569.0 (92.90)	137285.0 (92.92)	34284.0 (92.82)
no	7643.0 (4.14)	6068.0 (4.11)	1575.0 (4.26)
yes	5462.0 (2.96)	4386.0 (2.97)	1076.0 (2.91)
salt consumption, *n*(%)			
no response	99523.0 (53.89)	79551.0 (53.85)	19972.0 (54.07)
normal	53766.0 (29.11)	43082.0 (29.16)	10684.0 (28.93)

Table 5 (continued)

Variable	Total population	Train dataset	Test dataset
low	16413.0 (8.89)	13091.0 (8.86)	3322.0 (8.99)
high	14972.0 (8.11)	12015.0 (8.13)	2957.0 (8.01)
work intensity, *n*(%)			
no response	65581.0 (35.51)	52339.0 (35.43)	13242.0 (35.85)
moderate-intensity	64134.0 (34.73)	51384.0 (34.78)	12750.0 (34.52)
vigorous-intensity	54959.0 (29.76)	44016.0 (29.79)	10943.0 (29.63)
had blood pressure measurement, *n*(%)			
yes	97983.0 (53.06)	78419.0 (53.08)	19564.0 (52.97)
no	67811.0 (36.72)	54229.0 (36.71)	13582.0 (36.77)
no response	18880.0 (10.22)	15091.0 (10.21)	3789.0 (10.26)
taken drugs for raised bp, *n*(%)			
no response	148841.0 (80.60)	119021.0 (80.56)	29820.0 (80.74)
no	22428.0 (12.14)	17921.0 (12.13)	4507.0 (12.20)
yes	13405.0 (7.26)	10797.0 (7.31)	2608.0 (7.06)
had blood sugar measurement, *n*(%)			
no	126432.0 (68.46)	101154.0 (68.47)	25278.0 (68.44)
yes	54113.0 (29.30)	43272.0 (29.29)	10841.0 (29.35)
no response	4129.0 (2.24)	3313.0 (2.24)	816.0 (2.21)
taken diabetes drugs, *n*(%)			
no response	165765.0 (89.76)	132600.0 (89.75)	33165.0 (89.79)
yes	17022.0 (9.22)	13607.0 (9.21)	3415.0 (9.25)
no	1887.0 (1.02)	1532.0 (1.04)	355.0 (0.96)
had cholesterol measurement, *n*(%)			
no	87445.0 (47.35)	69866.0 (47.29)	17579.0 (47.59)
no response	77040.0 (41.72)	61670.0 (41.74)	15370.0 (41.61)
yes	20189.0 (10.93)	16203.0 (10.97)	3986.0 (10.79)
taken cholesterol oral treatment, *n*(%)			
no response	179155.0 (97.01)	143288.0 (96.99)	35867.0 (97.11)
no	3835.0 (2.08)	3095.0 (2.09)	740.0 (2.00)
yes	1684.0 (0.91)	1356.0 (0.92)	328.0 (0.89)
had heart attack, *n*(%)			
no	98979.0 (53.60)	79176.0 (53.59)	19803.0 (53.62)
no response	76528.0 (41.44)	61254.0 (41.46)	15274.0 (41.35)
yes	9167.0 (4.96)	7309.0 (4.95)	1858.0 (5.03)
taking heart disease medication, *n*(%)			
no	96645.0 (52.33)	77291.0 (52.32)	19354.0 (52.40)
no response	83737.0 (45.34)	66963.0 (45.33)	16774.0 (45.41)
yes	4292.0 (2.32)	3485.0 (2.36)	807.0 (2.18)
treated for raised bp, *n*(%)			
no	160181.0 (86.74)	128063.0 (86.68)	32118.0 (86.96)
no response	12560.0 (6.80)	10068.0 (6.81)	2492.0 (6.75)
yes	11933.0 (6.46)	9608.0 (6.50)	2325.0 (6.29)
are you pregnant, *n*(%)			
no	105556.0 (57.16)	84437.0 (57.15)	21119.0 (57.18)
no response	74248.0 (40.20)	59427.0 (40.22)	14821.0 (40.13)
yes	4870.0 (2.64)	3875.0 (2.62)	995.0 (2.69)
blood pressure, *n*(%)			
high	105677.0 (57.22)	84523.0 (57.21)	21154.0 (57.27)
normal	78997.0 (42.78)	63216.0 (42.79)	15781.0 (42.73)

Table 6 Global model results

Model	Accuracy	F1	Precision	Recall
Logistic Regression	0.682000	0.671893	0.675080	0.670566
KNN	0.632135	0.623198	0.623900	0.622770
Random Forest	0.681374	0.666707	0.675614	0.665052
XGBoost	0.685157	0.674266	0.678514	0.672741
FCNN	0.682490	0.676615	0.676402	0.676859

WHO regions. Table 7 presents the performance metrics for each model in each region, including accuracy, precision, recall, and F1 score.

Sub-Saharan Africa In the Sub-Saharan Africa region, the Fully Connected Neural Network (FCNN) achieved the highest accuracy at 64.96%, followed by Logistic Regression at 64.89% and XGBoost at 64.44%. The Random Forest model had an accuracy of 64.39%, while the K-Nearest neighbours (KNN) model had the lowest accuracy at 59.99%.

East Asia and Pacific For the East Asia and Pacific region, FCNN outperformed the other models with an accuracy of 71.16%, followed by Random Forest at 71.02% and Logistic Regression at 70.09%. XGBoost had an accuracy of 70.04%, and KNN had the lowest accuracy at 64.66%.

South Asia In South Asia, FCNN achieved the highest accuracy at 68.70%, followed by Logistic Regression at 68.38% and XGBoost at 68.17%. Random Forest had an accuracy of 68.13%, and KNN had the lowest accuracy at 64.32%.

Table 7 Region model results

Region	Model	Accuracy	Precision	Recall	F1 Score
Sub-Saharan Africa	Logistic Regression	0.648861	0.644065	0.641447	0.642049
	KNN	0.599936	0.594048	0.592844	0.593083
	Random Forest	0.643888	0.639734	0.631392	0.631252
	XGBoost	0.644370	0.639408	0.635315	0.635860
	FCNN	0.649583	0.646307	0.647006	0.646550
East Asia and Pacific	Logistic Regression	0.700882	0.693892	0.688995	0.690592
	KNN	0.646561	0.637632	0.636127	0.636703
	Random Forest	0.710229	0.706572	0.692152	0.694847
	XGBoost	0.700353	0.693469	0.687624	0.689394
	FCNN	0.711640	0.707025	0.695255	0.697858
South Asia	Logistic Regression	0.683828	0.675252	0.667238	0.669180
	KNN	0.643239	0.633138	0.631687	0.632254
	Random Forest	0.681265	0.673443	0.661091	0.663098
	XGBoost	0.681692	0.672732	0.667324	0.668931
	FCNN	0.687033	0.678637	0.677180	0.677809
Middle East and North Africa	Logistic Regression	0.682332	0.681080	0.678263	0.678740
	KNN	0.627655	0.625866	0.625866	0.625866
	Random Forest	0.682332	0.681902	0.677070	0.677459
	XGBoost	0.659738	0.657889	0.656126	0.656465
	FCNN	0.685947	0.684707	0.685160	0.684859
Europe and Central Asia	Logistic Regression	0.777456	0.723471	0.680339	0.694339
	KNN	0.728733	0.650024	0.626934	0.634363
	Random Forest	0.775349	0.734304	0.646467	0.663054
	XGBoost	0.767711	0.708562	0.666829	0.679866
	FCNN	0.771925	0.718277	0.660310	0.675606
Latin America and the Caribbean	Logistic Regression	0.695491	0.695808	0.694014	0.694102
	KNN	0.640849	0.641394	0.638583	0.637973
	Random Forest	0.702918	0.702712	0.702018	0.702161
	XGBoost	0.668966	0.668650	0.667921	0.668012
	FCNN	0.693899	0.695581	0.691693	0.691448

Middle East and North Africa For the Middle East and North Africa region, FCNN had the highest accuracy at 68.59%, followed by Logistic Regression and Random Forest, both at 68.23%. XGBoost had an accuracy of 65.97%, and KNN had the lowest accuracy at 62.77%.

Europe and Central Asia In Europe and Central Asia, Logistic Regression achieved the highest accuracy at 77.75%, followed by Random Forest at 77.53% and FCNN at 77.19%. XGBoost had an accuracy of 76.77%, and KNN had the lowest accuracy at 72.87%.

Latin America and Caribbean For the Latin America and Caribbean region, Random Forest outperformed the other models with an accuracy of 70.29%, followed by Logistic Regression at 69.55% and FCNN at 69.39%. XGBoost had an accuracy of 66.90%, and KNN had the lowest accuracy at 64.08%.

The regional model results demonstrate that the performance of the models varies across regions, with FCNN and Logistic Regression generally performing well in most regions. Random Forest also showed strong performance in some regions, particularly in East Asia and Pacific and Latin America and the Caribbean. XGBoost and KNN consistently had lower accuracies compared to the other models across all regions.

Performance per country

The country-specific models were trained and evaluated on subsets of the dataset corresponding to each of the 57 countries included in the study. Table 8 presents the performance metrics for each model in each country, including accuracy, precision, recall, and F1 score.

The performance of the models varied considerably across countries, with some models consistently outperforming others in certain countries, while the reverse was true in other countries. For example, in Ethiopia, the Logistic Regression model achieved the highest accuracy at 62.55%, precision at 62.58%, recall at 62.31%, and F1 score at 62.22%. In contrast, the KNN model had the lowest scores across all metrics in Ethiopia, with an accuracy of 58.96%, precision of 58.88%, recall of 58.83%, and F1 score of 58.82%.

In some countries, such as Georgia and Belarus, the FCNN model performed well, with accuracies of 72.51% and 81.41%, respectively. However, in other countries like the Bahamas and Barbados, the Random Forest model achieved the highest accuracies at 71.49% and 73.21%, respectively.

XGBoost demonstrated strong performance in certain countries, such as Tokelau, where it achieved the highest accuracy (77.67%), precision (77.52%), recall (76.46%), and F1 score (76.80%) among all models. However, its performance was less impressive in other countries, like Chad and Grenada, where it had lower scores compared to other models.

The Logistic Regression model showed consistent performance across many countries, often ranking among the top models in terms of accuracy, precision, recall, and F1 score. For instance, in Mongolia, Logistic Regression achieved an accuracy of 72.98%, precision of 73.30%, recall of 72.92%, and F1 score of 72.86%.

The KNN model generally had lower scores compared to the other models across most countries. However, there were a few exceptions, such as in Grenada, where KNN achieved the highest accuracy at 66.85%, precision at 64.49%, recall at 61.81%, and F1 score at 61.94%.

Model explainability

Mean decrease in impurity (MDI)

Mean decrease in impurity (MDI) serves as a method for evaluating feature importance in decision tree-based models. This approach quantifies the average reduction in impurity-such as entropy or the Gini index-resulting from the utilization of a specific feature to partition the dataset. A greater decrease in impurity corresponds to a higher degree of importance for the feature. In essence, the feature that induces the largest reduction in impurity is deemed the most significant in the dataset. As illustrated in Fig. 2, the top five features include age, weight, hip circumference, waist circumference, and sex (male).

Feature permutation importance

Feature permutation importance evaluates the impact on model performance when a specific feature is randomly altered by introducing noise. The importance of that particular feature for the model's predictions can be estimated by contrasting the performance of the model utilizing the permuted or modified feature with the performance of the model employing the original feature. A greater change in performance implies increased importance for the feature. This method has been implemented for a random forest classifier in the global model.

For each feature within the dataset, the values of the feature were randomly permuted, and predictions were generated using the trained model. Among the features included in the model, age, heart reading in beats per minute, weight, waist circumference, and hip circumference emerged as the top five features contributing to the change in error of the model after training on the permuted features compared to the original model. The results for the feature permutation importance measure are depicted in Fig. 3.

Table 8 Country model results

Country	Model	Accuracy	Precision	Recall	F1 Score
Ethiopia	Logistic Regression	0.625536	0.625842	0.623066	0.622245
	KNN	0.589592	0.588780	0.588278	0.588184
	Random Forest	0.624464	0.623993	0.622877	0.622740
	XGBoost	0.598712	0.598074	0.597907	0.597934
	FCNN	0.622854	0.623633	0.619993	0.618681
Georgia	Logistic Regression	0.733831	0.705691	0.688705	0.694650
	KNN	0.662935	0.622973	0.616784	0.619009
	Random Forest	0.717662	0.690813	0.648612	0.655733
	XGBoost	0.702736	0.668253	0.653978	0.658629
	FCNN	0.725124	0.696775	0.668602	0.676023
Palestine	Logistic Regression	0.665630	0.657683	0.659452	0.658328
	KNN	0.644635	0.634411	0.634048	0.634220
	Random Forest	0.692068	0.682941	0.680831	0.681709
	XGBoost	0.671851	0.662777	0.662947	0.662861
	FCNN	0.676516	0.667312	0.666953	0.667125
Bahamas	Logistic Regression	0.681818	0.649082	0.617983	0.620913
	KNN	0.582645	0.530801	0.527957	0.527005
	Random Forest	0.714876	0.707520	0.638747	0.642311
	XGBoost	0.665289	0.630389	0.620208	0.623060
	FCNN	0.661157	0.625199	0.564034	0.545696
Barbados	Logistic Regression	0.696429	0.655050	0.659357	0.656937
	KNN	0.696429	0.649321	0.644737	0.646753
	Random Forest	0.732143	0.690244	0.671053	0.677791
	XGBoost	0.660714	0.615220	0.618421	0.616577
	FCNN	0.678571	0.619048	0.602339	0.606250
Eritrea	Logistic Regression	0.686074	0.659219	0.619708	0.621790
	KNN	0.638867	0.595351	0.579376	0.579375
	Random Forest	0.689221	0.662502	0.626358	0.629580
	XGBoost	0.651456	0.612989	0.597621	0.599444
	FCNN	0.684500	0.654385	0.637041	0.641096
Chad	Logistic Regression	0.672269	0.672207	0.658439	0.658373
	KNN	0.579832	0.571347	0.567640	0.566327
	Random Forest	0.613445	0.613805	0.590466	0.579814
	XGBoost	0.565826	0.557158	0.554949	0.554056
	FCNN	0.666667	0.662973	0.657471	0.658258
Palau	Logistic Regression	0.733333	0.733410	0.733493	0.733321
	KNN	0.633333	0.633273	0.633320	0.633268
	Random Forest	0.696667	0.699336	0.697679	0.696258
	XGBoost	0.693333	0.695670	0.694278	0.692992
	FCNN	0.740000	0.740991	0.740563	0.739954
Belarus	Logistic Regression	0.816139	0.738583	0.690975	0.708446
	KNN	0.781410	0.676625	0.642754	0.654732
	Random Forest	0.814096	0.740968	0.665448	0.687578
	XGBoost	0.803882	0.716959	0.668544	0.685272
	FCNN	0.814096	0.733316	0.700955	0.714092
Nauru	Logistic Regression	0.719388	0.716654	0.713805	0.714747
	KNN	0.566327	0.570330	0.570707	0.566225
	Random Forest	0.734694	0.744056	0.720328	0.721925
	XGBoost	0.663265	0.661066	0.651305	0.651509
	FCNN	0.704082	0.701258	0.702020	0.701565

Table 8 (continued)

Country	Model	Accuracy	Precision	Recall	F1 Score
Botswana	Logistic Regression	0.647143	0.637373	0.636700	0.637002
	KNN	0.584286	0.571909	0.571253	0.571482
	Random Forest	0.657143	0.646250	0.637345	0.638738
	XGBoost	0.662857	0.653093	0.651185	0.651935
	FCNN	0.641429	0.629754	0.625674	0.626709
Malawi	Logistic Regression	0.688889	0.678105	0.653439	0.656297
	KNN	0.623810	0.598810	0.582011	0.579260
	Random Forest	0.680952	0.674762	0.636243	0.636182
	XGBoost	0.666667	0.650107	0.641534	0.643684
	FCNN	0.701587	0.708848	0.653439	0.653830
Fiji	Logistic Regression	0.673428	0.647494	0.645718	0.646529
	KNN	0.643002	0.608170	0.599642	0.601558
	Random Forest	0.705882	0.684696	0.652801	0.658282
	XGBoost	0.653144	0.618364	0.605335	0.607574
	FCNN	0.713996	0.696499	0.659212	0.665328
Niger	Logistic Regression	0.665169	0.609734	0.588531	0.590520
	KNN	0.638202	0.580668	0.571469	0.572904
	Random Forest	0.669663	0.611601	0.577175	0.575020
	XGBoost	0.653933	0.598522	0.584972	0.587129
	FCNN	0.676404	0.623507	0.592090	0.593463
Vanuatu	Logistic Regression	0.645735	0.642009	0.626388	0.624881
	KNN	0.614929	0.606436	0.600939	0.600760
	Random Forest	0.635071	0.636640	0.608446	0.600585
	XGBoost	0.637441	0.630802	0.622186	0.622109
	FCNN	0.654028	0.649372	0.637857	0.637924
Qatar	Logistic Regression	0.663573	0.662685	0.662433	0.662527
	KNN	0.587007	0.585272	0.583531	0.582855
	Random Forest	0.654292	0.653333	0.652903	0.653030
	XGBoost	0.619490	0.618318	0.617904	0.617990
	FCNN	0.647332	0.647693	0.643320	0.642714
Azerbaijan	Logistic Regression	0.713693	0.699672	0.667344	0.673018
	KNN	0.668050	0.642422	0.631758	0.634516
	Random Forest	0.740664	0.737720	0.692253	0.700241
	XGBoost	0.738589	0.722867	0.708956	0.713744
	FCNN	0.742739	0.737575	0.697161	0.705199
Bhutan	Logistic Regression	0.715074	0.695368	0.669524	0.675488
	KNN	0.658088	0.629602	0.624850	0.626589
	Random Forest	0.697610	0.675012	0.645892	0.650623
	XGBoost	0.690257	0.664531	0.652934	0.656516
	FCNN	0.705882	0.682779	0.666355	0.671059
Togo	Logistic Regression	0.639726	0.635969	0.633385	0.633782
	KNN	0.595890	0.590321	0.586691	0.585970
	Random Forest	0.602740	0.599778	0.599965	0.599853
	XGBoost	0.656164	0.653118	0.652706	0.652881
	FCNN	0.643836	0.642635	0.631910	0.630636
Tokelau	Logistic Regression	0.718447	0.713716	0.716680	0.714572
	KNN	0.669903	0.661943	0.654083	0.655586
	Random Forest	0.766990	0.763889	0.756163	0.758782
	XGBoost	0.776699	0.775240	0.764638	0.767950
	FCNN	0.766990	0.763889	0.756163	0.758782

Table 8 (continued)

Country	Model	Accuracy	Precision	Recall	F1 Score
Gambia	Logistic Regression	0.667166	0.644245	0.620177	0.621805
	KNN	0.650675	0.624230	0.611214	0.613045
	Random Forest	0.674663	0.660153	0.615211	0.613135
	XGBoost	0.638681	0.610284	0.599272	0.600582
	FCNN	0.668666	0.648221	0.614764	0.614510
Guyana	Logistic Regression	0.703704	0.706216	0.700129	0.700000
	KNN	0.615984	0.615574	0.612321	0.611353
	Random Forest	0.697856	0.698430	0.695233	0.695407
	XGBoost	0.692008	0.691562	0.690337	0.690567
	FCNN	0.684211	0.685751	0.680772	0.680538
Micronesia	Logistic Regression	0.736170	0.734393	0.736142	0.734782
	KNN	0.685106	0.683812	0.685425	0.683818
	Random Forest	0.719149	0.717334	0.718956	0.717671
	XGBoost	0.710638	0.709366	0.711204	0.709455
	FCNN	0.748936	0.746934	0.744405	0.745321
Bangladesh	Logistic Regression	0.692046	0.692115	0.692152	0.692041
	KNN	0.643100	0.643257	0.643257	0.643100
	Random Forest	0.682529	0.683077	0.682866	0.682491
	XGBoost	0.664854	0.665434	0.665208	0.664804
	FCNN	0.693406	0.693341	0.693363	0.693349
Mali	Logistic Regression	0.666667	0.656082	0.660343	0.657143
	KNN	0.671569	0.660462	0.664408	0.661608
	Random Forest	0.691176	0.676740	0.674345	0.675381
	XGBoost	0.656863	0.642637	0.643782	0.643143
	FCNN	0.710784	0.697601	0.688497	0.691503
Tuvalu	Logistic Regression	0.782407	0.758696	0.702644	0.718368
	KNN	0.703704	0.639287	0.596314	0.599954
	Random Forest	0.763889	0.808107	0.631724	0.639139
	XGBoost	0.782407	0.761892	0.698538	0.715080
	FCNN	0.773148	0.737654	0.708254	0.718743
Mongolia	Logistic Regression	0.729839	0.732957	0.729222	0.728563
	KNN	0.642137	0.644449	0.641420	0.639968
	Random Forest	0.730847	0.731299	0.730586	0.730549
	XGBoost	0.706653	0.706958	0.706410	0.706366
	FCNN	0.721774	0.722880	0.721380	0.721175
Eswatini	Logistic Regression	0.628205	0.622158	0.621067	0.621438
	KNN	0.622711	0.615863	0.613536	0.614005
	Random Forest	0.624542	0.617705	0.607775	0.606472
	XGBoost	0.635531	0.629408	0.620223	0.619818
	FCNN	0.635531	0.629808	0.628930	0.629262
Myanmar	Logistic Regression	0.710175	0.697167	0.697047	0.697106
	KNN	0.649825	0.632651	0.630654	0.631472
	Random Forest	0.704561	0.692208	0.674789	0.678739
	XGBoost	0.692632	0.677698	0.666727	0.669762
	FCNN	0.703158	0.690313	0.691536	0.690880
Maldives	Logistic Regression	0.713262	0.716073	0.710268	0.710130
	KNN	0.670251	0.675646	0.665775	0.663679
	Random Forest	0.731183	0.733325	0.728641	0.728828
	XGBoost	0.684588	0.685456	0.682115	0.682035
	FCNN	0.731183	0.730839	0.730340	0.730504

Table 8 (continued)

Country	Model	Accuracy	Precision	Recall	F1 Score
Uganda	Logistic Regression	0.647868	0.642360	0.611689	0.606781
	KNN	0.570839	0.555112	0.553520	0.553566
	Random Forest	0.627235	0.628617	0.577554	0.555532
	XGBoost	0.602476	0.583961	0.573499	0.570627
	FCNN	0.647868	0.660842	0.599872	0.582858
Mozambique	Logistic Regression	0.663551	0.638310	0.614308	0.615575
	KNN	0.605607	0.575077	0.571241	0.571959
	Random Forest	0.635514	0.601050	0.577543	0.573503
	XGBoost	0.659813	0.633941	0.620690	0.623088
	FCNN	0.656075	0.629357	0.599802	0.598230
Benin	Logistic Regression	0.649294	0.623345	0.610178	0.611866
	KNN	0.630836	0.600436	0.587284	0.587344
	Random Forest	0.665581	0.650284	0.603476	0.598268
	XGBoost	0.628664	0.597301	0.583417	0.582925
	FCNN	0.644951	0.620728	0.614520	0.616219
Moldova	Logistic Regression	0.839080	0.781997	0.676812	0.706276
	KNN	0.791667	0.672123	0.626687	0.640873
	Random Forest	0.826149	0.777489	0.625866	0.651087
	XGBoost	0.806034	0.701556	0.645841	0.663475
	FCNN	0.821839	0.734792	0.668418	0.689889
Nepal	Logistic Regression	0.664875	0.629546	0.601241	0.601854
	KNN	0.637993	0.596695	0.583224	0.584017
	Random Forest	0.668459	0.640692	0.584701	0.575610
	XGBoost	0.633513	0.586801	0.568709	0.566421
	FCNN	0.672939	0.645476	0.594506	0.589560
Comoros	Logistic Regression	0.689252	0.689676	0.689551	0.689237
	KNN	0.600467	0.600530	0.600552	0.600459
	Random Forest	0.689252	0.692012	0.690125	0.688702
	XGBoost	0.657710	0.659065	0.658326	0.657463
	FCNN	0.679907	0.679987	0.679515	0.679513
Tonga	Logistic Regression	0.723935	0.672291	0.608892	0.614364
	KNN	0.666667	0.587003	0.571578	0.573714
	Random Forest	0.707783	0.642865	0.579910	0.577661
	XGBoost	0.712188	0.650688	0.615081	0.621744
	FCNN	0.709251	0.644995	0.598299	0.602426
Grenada	Logistic Regression	0.624309	0.589517	0.576799	0.576171
	KNN	0.668508	0.644887	0.618077	0.619428
	Random Forest	0.646409	0.616013	0.591874	0.590266
	XGBoost	0.629834	0.595878	0.581263	0.580627
	FCNN	0.624309	0.587409	0.571234	0.568685
Madagascar	Logistic Regression	0.631734	0.605470	0.574197	0.564566
	KNN	0.619980	0.592991	0.583253	0.583279
	Random Forest	0.671890	0.671190	0.611675	0.604290
	XGBoost	0.670911	0.653440	0.631797	0.633806
	FCNN	0.647405	0.625052	0.602382	0.601359
Ecuador	Logistic Regression	0.715566	0.707130	0.689602	0.693620
	KNN	0.645017	0.627281	0.608577	0.608124
	Random Forest	0.702128	0.693695	0.672091	0.675682
	XGBoost	0.678611	0.665848	0.648793	0.651281
	FCNN	0.707727	0.700692	0.677236	0.681109

Table 8 (continued)

Country	Model	Accuracy	Precision	Recall	F1 Score
Gabon	Logistic Regression	0.686534	0.686788	0.677684	0.678177
	KNN	0.609272	0.607233	0.595057	0.590087
	Random Forest	0.679912	0.678897	0.671969	0.672533
	XGBoost	0.637969	0.634601	0.631053	0.631283
	FCNN	0.668874	0.668249	0.669465	0.668018
Libya	Logistic Regression	0.738636	0.717506	0.692677	0.700026
	KNN	0.691558	0.660090	0.646970	0.650998
	Random Forest	0.743506	0.734682	0.681313	0.690963
	XGBoost	0.712662	0.685109	0.670455	0.675283
	FCNN	0.745130	0.726864	0.696717	0.705081
Zambia	Logistic Regression	0.650530	0.652300	0.644968	0.643626
	KNN	0.571861	0.569341	0.566721	0.564833
	Random Forest	0.617247	0.615781	0.614002	0.613849
	XGBoost	0.594554	0.592675	0.591325	0.591071
	FCNN	0.659607	0.661574	0.654231	0.653177
Lesotho	Logistic Regression	0.700855	0.647063	0.612088	0.617310
	KNN	0.643875	0.576465	0.567747	0.569367
	Random Forest	0.675214	0.597547	0.553674	0.542593
	XGBoost	0.663818	0.600459	0.587101	0.589921
	FCNN	0.669516	0.592877	0.561092	0.556818
Namibia	Logistic Regression	0.681890	0.682331	0.588432	0.572695
	KNN	0.656693	0.618461	0.588699	0.586904
	Random Forest	0.650394	0.608842	0.568190	0.557455
	XGBoost	0.636220	0.589815	0.568912	0.565621
	FCNN	0.677165	0.665168	0.587453	0.574426
Tanzania	Logistic Regression	0.639960	0.629931	0.602395	0.597317
	KNN	0.553907	0.527434	0.524584	0.520730
	Random Forest	0.613254	0.596192	0.569989	0.557808
	XGBoost	0.601385	0.582168	0.574107	0.572687
	FCNN	0.629080	0.615445	0.611444	0.612355
Liberia	Logistic Regression	0.702128	0.663095	0.638105	0.643939
	KNN	0.595745	0.543750	0.542339	0.542755
	Random Forest	0.617021	0.495935	0.497984	0.465909
	XGBoost	0.787234	0.763105	0.763105	0.763105
	FCNN	0.723404	0.693182	0.654234	0.662244
Samoa	Logistic Regression	0.756410	0.756872	0.754120	0.754707
	KNN	0.669872	0.672818	0.664552	0.663480
	Random Forest	0.750000	0.750836	0.747363	0.747970
	XGBoost	0.733974	0.734327	0.731460	0.731967
	FCNN	0.778846	0.778256	0.778428	0.778334
Afghanistan	Logistic Regression	0.661623	0.648159	0.647146	0.647605
	KNN	0.617607	0.608101	0.610741	0.608480
	Random Forest	0.672627	0.658818	0.655292	0.656613
	XGBoost	0.658872	0.648610	0.651388	0.649481
	FCNN	0.662999	0.650104	0.649939	0.650020
Ghana	Logistic Regression	0.708087	0.665085	0.651338	0.656177
	KNN	0.641026	0.589921	0.587628	0.588591
	Random Forest	0.706114	0.660886	0.630063	0.636402
	XGBoost	0.682446	0.632649	0.620033	0.623924
	FCNN	0.694280	0.644149	0.615146	0.620129

Table 8 (continued)

Country	Model	Accuracy	Precision	Recall	F1 Score
Rwanda	Logistic Regression	0.615162	0.614989	0.614894	0.614902
	KNN	0.549458	0.549088	0.548373	0.547238
	Random Forest	0.610830	0.611647	0.609647	0.608537
	XGBoost	0.597834	0.597648	0.597586	0.597591
	FCNN	0.622383	0.626220	0.623707	0.620921
Niue	Logistic Regression	0.738562	0.734921	0.731170	0.732517
	KNN	0.679739	0.675731	0.677195	0.676197
	Random Forest	0.718954	0.719697	0.703835	0.706098
	XGBoost	0.705882	0.701882	0.695505	0.697136
	FCNN	0.738562	0.738978	0.742711	0.737654
Armenia	Logistic Regression	0.755747	0.687307	0.661098	0.670510
	KNN	0.744253	0.669384	0.640046	0.649412
	Random Forest	0.781609	0.737388	0.659164	0.677070
	XGBoost	0.747126	0.676745	0.661743	0.667867
	FCNN	0.772989	0.737395	0.626940	0.641017
Kiribati	Logistic Regression	0.736170	0.691840	0.661407	0.670362
	KNN	0.693617	0.634327	0.615746	0.620764
	Random Forest	0.765957	0.739412	0.683146	0.697172
	XGBoost	0.727660	0.681974	0.673451	0.677087
	FCNN	0.731915	0.686502	0.650999	0.660132
Algeria	Logistic Regression	0.663808	0.663813	0.663801	0.663799
	KNN	0.613208	0.613269	0.613184	0.613125
	Random Forest	0.661235	0.662057	0.661294	0.660856
	XGBoost	0.643225	0.643734	0.643274	0.642956
	FCNN	0.669811	0.669832	0.669800	0.669792
Kyrgyzstan	Logistic Regression	0.735656	0.632479	0.586549	0.592868
	KNN	0.721311	0.606199	0.571750	0.575511
	Random Forest	0.754098	0.680889	0.565839	0.560255
	XGBoost	0.743852	0.649783	0.602297	0.611264
	FCNN	0.750000	0.661911	0.614108	0.624663
Guinea	Logistic Regression	0.645880	0.640829	0.643661	0.641065
	KNN	0.636971	0.631348	0.633801	0.631643
	Random Forest	0.634744	0.624379	0.623209	0.623676
	XGBoost	0.634744	0.624379	0.623209	0.623676
	FCNN	0.654788	0.645787	0.645574	0.645677

Discussion

This study applied several machine learning models to predict blood pressure status using the WHO STEPS dataset with a nationally-representative sample of 184,674 participants from 57 low- and middle-income countries. The XGBoost and FCNN models performed slightly better than logistic regression and random forest models across various metrics, while KNN consistently underperformed. Notably, model performance varied significantly across regions and countries, highlighting the need for context-specific approaches.

Our feature importance analysis identified age, weight, heart rate, waist circumference, and height as the most important blood pressure predictors, aligning with previous research findings [22–24]. The model's explainability is crucial for facilitating its adoption and trust among healthcare professionals.

Our findings have several implications for policymakers, clinical care, and researchers focused on NCD prevention and control in resource-limited settings. Predictive machine learning models calibrated using health information systems can identify high-risk populations and prioritise sub-national regions for in-person hypertension screening and interventions. Similarly, clinicians can use these models to screen their patient database and

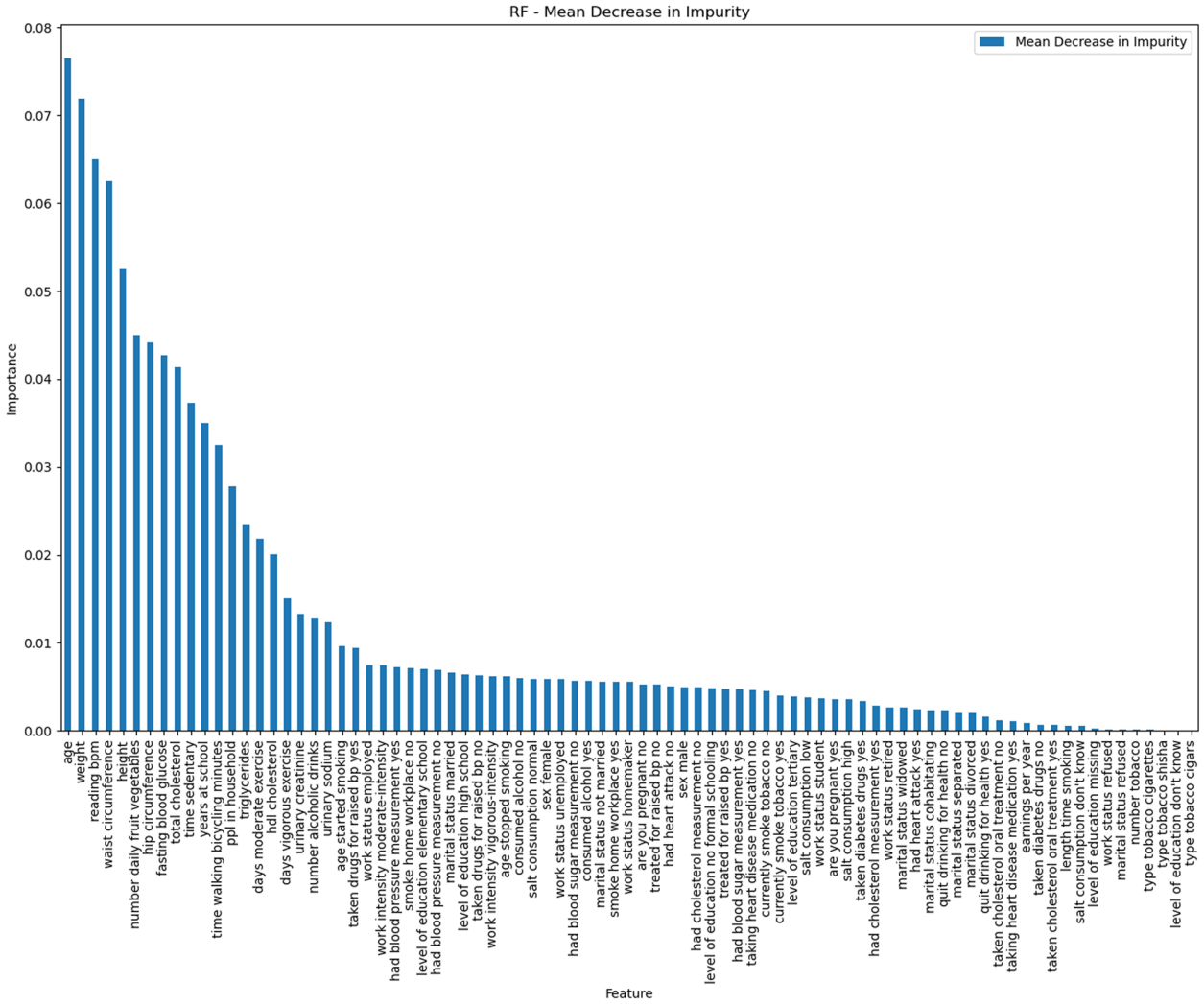

Fig. 2 Using mean decrease in impurity on the global model

identify those with a higher risk of raised blood pressure for intensive monitoring or earlier intervention.

The variability in model performance across countries reinforces the importance of developing and validating country-specific models for increased accuracy and more tailored interventions. While this study does not introduce new ML algorithms, it highlights the potential of applying existing techniques to large-scale datasets in LMICs to advance public health objectives.

The strengths of this study include the use of a large, nationally representative dataset [25, 26], evaluation of multiple machine learning models, and validation using a separate testing dataset. While this study provides valuable insights, it also has limitations. The models were applied to retrospective data and not tested prospectively in clinical settings or population settings. Future studies should validate these findings in prospective, real-world scenarios. Additionally, the study focused solely on predicting blood pressure status and did not extend to other NCD risk factors or blood pressure control over time.

Despite these limitations, this study contributes to the literature on using machine learning for chronic disease management. Future research should focus on validation of these models in clinical settings, developing country-specific models to improve prediction accuracy, expanding the target to other NCD risk factors and long-term blood pressure control, and more broadly, the integration of ML-based prediction tools into health systems in LMICs.

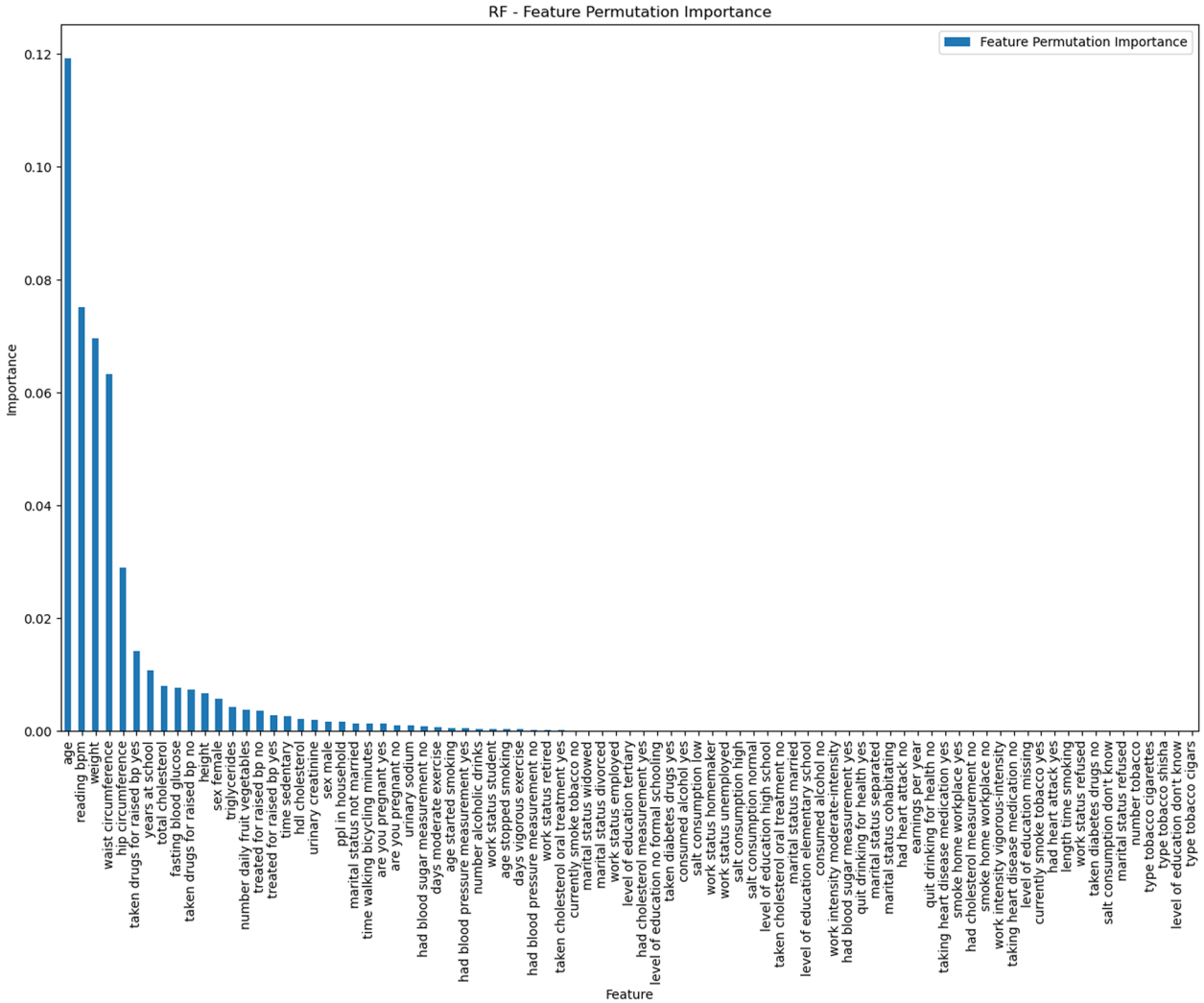

Fig. 3 Using feature permutation importance on the global model

Conclusion

This study demonstrates the potential of applying machine learning techniques to large-scale health datasets for predicting blood pressure status in LMICs. The variability in model performance across countries underscores the need for context-specific approaches in addressing hypertension. Policymakers and healthcare providers in LMICs could potentially use these models as tools for population-level risk stratification and resource allocation, complementing rather than replacing direct blood pressure measurements.

By addressing the identified limitations and expanding the geographical coverage to include more diverse populations, researchers can develop more comprehensive and reliable models for predicting blood pressure control. The integration of such models into clinical practice, coupled with further validation and refinement, has the potential to revolutionize the management of hypertension and other non-communicable diseases in resource-limited settings.

In conclusion, this study lays the foundation for future research on the use of machine learning in the context of global health and non-communicable disease management. The explainable machine learning model developed herein serves as a valuable tool for supporting clinical decision-making and improving blood pressure control in low- and middle-income countries. With continued efforts to address the limitations and expand upon this work, the application of machine learning in healthcare can contribute significantly to the achievement of better health outcomes for populations worldwide.

Acknowledgements
Not applicable.

Authors' contributions
E.B. and A.C. conceptualized the study. E.B. and G.O. acquired and processed the data. E.B. and A.C. designed and performed the data analyses with support from N.J., P.P., and E.B. E.B. and A.C., together with N.J., P.P., and E.B., drafted the manuscript. All authors contributed to the interpretation of the results and critically revised the manuscript for important intellectual content. A.C. supervised the study. All authors approved the final version of the manuscript for submission.

Authors' information
Not applicable.

Funding
Not applicable.

Availability of data and materials
Data and code that support the findings of this study are available on Github: https://github.com/SiliconBlast/bpc-prediction-lmics.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 1 May 2024 Accepted: 12 August 2024
Published online: 23 August 2024

References

1. World Health Organization, et al. Noncommunicable diseases progress monitor. 2022. https://iris.who.int/bitstream/handle/10665/353048/9789240047761-eng.pdf.
2. Centers for Disease Control and Prevention. About Global NCDs. 2021. https://www.cdc.gov/globalhealth/healthprotection/ncd/global-ncd-overview.html#one. Accessed 6 July 2024.
3. UNICEF. Non-communicable diseases. 2023. https://www.unicef.org/health/non-communicable-diseases. Accessed 6 July 2024.
4. Aikaeli F, Njim T, Gissing S, Moyo F, Alam U, Mfinanga SG, et al. Prevalence of microvascular and macrovascular complications of diabetes in newly diagnosed type 2 diabetes in low-and-middle-income countries: A systematic review and meta-analysis. PLOS Global Public Health. 2022;2(6):e0000599.
5. Gheorghe A, Griffiths U, Murphy A, Legido-Quigley H, Lamptey P, Perel P. The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: a systematic review. BMC Public Health. 2018;18(1):1–11.
6. World Health Organization. Noncommunicable Disease Surveillance, Monitoring and Reporting; 2023. https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/steps. Accessed 6 July 2024.
7. Wamai RG, Kengne AP, Levitt N. Non-communicable diseases surveillance: overview of magnitude and determinants in Kenya from STEPwise approach survey of 2015. BMC Public Health. 2018;18(3):1–8.
8. World Health Organization, et al. The WHO STEPwise approach to chronic disease risk factor surveillance. Geneva: World Health Organization; 2005.
9. Riley L, Guthold R, Cowan M, Savin S, Bhatti L, Armstrong T, et al. The World Health Organization STEPwise approach to noncommunicable disease risk-factor surveillance: methods, challenges, and opportunities. Am J Public Health. 2016;106(1):74–8.
10. Bonita R, Winkelmann R, Douglas KA, de Courten M. The WHO stepwise approach to surveillance (Steps) of non-communicable disease risk factors. In: McQueen DV, Puska P, editors. Global behavioral risk factor surveillance. Boston: Springer; 2003. pp. 9–22. https://doi.org/10.1007/978-1-4615-0071-1_3.
11. Kalaivani B, Ranichitra A. Unveiling the impact of outliers: an improved feature engineering technique for heart disease prediction. In: International Conference on IoT based control networks and intelligent systems. Singapore: Springer Nature Singapore; 2023. pp. 469–78.
12. Aggarwal V, Gupta V, Singh P, Sharma K, Sharma N. Detection of spatial outlier by using improved Z-score test. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). Tirunelveli: IEEE; 2019. pp. 788–90. https://doi.org/10.1109/ICOEI.2019.8862582.
13. Centers for Disease Control and Prevention. High blood pressure symptoms and causes. 2021. https://www.cdc.gov/bloodpressure/about.htm. Accessed 6 July 2024.
14. American Heart Association. Understanding blood pressure readings. 2022. https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings. Accessed 6 July 2024.
15. Rousseeuw PJ, Hampel FR, Ronchetti EM, Stahel WA. Robust statistics: the approach based on influence functions. New York: Wiley; 1986.
16. Huber PJ. Robust estimation of a location parameter. In: Kotz S, Johnson NL, editors. Breakthroughs in statistics. New York: Springer Series in Statistics; 1992. pp. 492–518. https://doi.org/10.1007/978-1-4612-4380-9_35.
17. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Wiley; 2013.
18. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am Stat. 1992;46(3):175–85.
19. Breiman L. Random Forests. Mach Learn. 2001;45(1):5–32.
20. 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 (KDD '16). New York: Association for Computing Machinery; 2016. pp. 785–94. https://doi.org/10.1145/2939672.2939785.
21. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016. http://www.deeplearningbook.org.
22. Islam SMS, Talukder A, Awal MA, Siddiqui MMU, Ahamad MM, Ahammed B, et al. Machine learning approaches for predicting hypertension and its associated factors using population-level data from three south asian countries. Front Cardiovasc Med. 2022;9:839379. https://doi.org/10.3389/fcvm.2022.839379.
23. Martinez-Ríos E, Montesinos L, Alfaro-Ponce M, Pecchia L. A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. Biomed Signal Process Control. 2021;68:102813.
24. Wu X, Yuan X, Wang W, Liu K, Qin Y, Sun X, et al. Value of a machine learning approach for predicting clinical outcomes in young patients with hypertension. Hypertension. 2020;75(5):1271–8.
25. Nasir N, Oswald P, Barneih F, Alshaltone O, AlShabi M, Bonny T, et al. Hypertension classification using machine learning part II. In: 2021 14th International Conference on Developments in eSystems Engineering (DeSE). Sharjah: IEEE; 2021. pp. 459–63. https://doi.org/10.1109/DeSE54285.2021.9719408.
26. Bani-Salameh H, Alkhatib SM, Abdalla M, Al-Hami M, Banat R, Zyod H, et al. Prediction of diabetes and hypertension using multi-layer perceptron neural networks. Int J Model Simul Sci Comput. 2021;12(02):2150012.

Publisher's Note