*Journal of the North for Basic and Applied Sciences,* Vol. (8), Issue (2), (November 2023/ Rabi' II 1445 H) 79-88

KINGDOM OF SAUDI ARABIA
**Northern Border University (NBU)**
**Journal of the North for Basic & Applied Sciences( JNBAS)**
**p - ISSN** : 1658 -7022 / **e-ISSN: 1658 - 7014**
www.nbu.edu.sa
s.journal@nbu.edu.sa

# Maximizing Hypertension Prevention through Machine Learning-Based Risk Prediction

**Saeed Awad M. Alqahtani ***

*(Received 5/2/2023 ; accepted 19/9/2023)*

**Abstract:** Background: Most current hypertension prediction models rely on conventional regression-based models, but this research aims to validate a straightforward and feasible model using various machine learning models such as Gradient Boosting Classifier, eXtreme Gradient Boosted (XGBoost), Adaptive Boosting (Adaboost), Multi-layer perceptron (MLP), Logistic Regression, Random Forest, eXtreme Gradient Boosted Random Forest (XGBRFBoost), K-Nearest Neighbors (KNN), and Decision Trees. Method: This study used data from the USA Centers for Disease Control and Prevention to predict hypertension using multiple machine learning algorithms. The preprocessed data set consists of 22 features and 70692 samples. Each individual and ensemble models were evaluated using ROC-AUC, accuracy, F1 score, precision, and recall. Results: The results show that the top 4 models that have high performance in terms of ROC-AUC, accuracy, F1 Score, precision, and recall are Gradient Boosting Classifier (Train ROC-AUC = 0.82, Test ROC-AUC = 0.81, Accuracy = 0.74, F1 Score = 0.78, Precision = 0.74, Recall = 0.83), XGBBoost (Train ROC-AUC = 0.82, Test ROC-AUC = 0.81, Accuracy = 0.74, F1 Score = 0.78, Precision = 0.74, Recall = 0.83), Adaboost (Train ROC-AUC = 0.81, Test ROC-AUC = 0.81, Accuracy = 0.74, F1 Score = 0.78, Precision = 0.75, Recall = 0.81) , MLP (Train ROC-AUC = 0.81, Test ROC-AUC = 0.81, Accuracy = 0.74, F1 Score = 0.77, Precision = 0.76, Recall = 0.79). Conclusion: This study shows Machine learning models outperform traditional statistical methods for complex hypertension risk prediction, offering improved screening for prevention.

**Keywords**: Machine learning Models, Hypertension, Evaluation metrics.



**DOI: 10.12816/0061646**

* **Corresponding Author:**
Associate Professor, Dept.: Physiology, Faculty: Medicine, University: Taibah, City: Medina, Kingdom of Saudi Arabia.

**e-mail:** dr_alqahtani@hotmail.com

المملكة العربية السعودية
جامعة الحدود الشمالية (NBU)
**مجلة الشمال للعلوم الأساسية والتطبيقية (JNBAS)**
طباعة ردمد: 1658-7022/ إلكتروني – ردمد: 1658-7014
www.nbu.edu.sa
http://jnbas.nbu.edu.sa

# الوقاية من ارتفاع ضغط الدم من خلال التنبؤ بالمخاطر القائم على التعلم الآلي

**سعيد عوض القحطاني**

(قدم للنشر في 1444/7/14هـ؛ وقبل للنشر في 1445/3/4هـ)

**مستخلص البحث**: هدفت هذه الدراسة إلى التحقق من صحة نموذج توقع ارتفاع ضغط الدم باستخدام نماذج متعددة للتعلم الآلي ، بما في ذلك Gradient Boosting Classifier و XGBoost و Adaboost و MLP و Random Forest و XGBRFBoost و KNN و Decision Trees. تم استخدام البيانات من المركز الأمريكي (CDC) المتخصص في السيطرة على الأمراض والوقاية منها ، مع مجموعة البيانات المعالجة مسبقًا التي تتكون من 22 متغير و 70692 عينة. تم إجراء التقييم باستخدام عدد من أدوات التقييم ومنها ROC-AUC ، وأظهرت النتائج أفضل النماذج التي تتمتع بأداء عالٍ. وخلصت الدراسة إلى أن نماذج التعلم الآلي تتفوق في الأداء على الأساليب الإحصائية التقليدية للتنبؤ بمخاطر ارتفاع ضغط الدم المعقد ، مما يوفر فحصًا محسنًا للوقاية.

**كلمات مفتاحية**: نماذج التعلم الآلي ، ارتفاع ضغط الدم ، مقاييس التقييم.



**للمراسلة:**
أستاذ مشارك. قسم: علم وظائف الأعضاء ، جامعة طيبة، المدينة المنورة.

**e-mail:** dr_alqahtani@hotmail.com

**DOI: 10.12816/0061646**

## 1. INTRODUCTION

Hypertension is a critical public health challenge that affects a diverse range of demographic groups globally and is the leading risk factor for preventable cardiovascular morbidity and mortality (Bromfield & Muntner, 2013). Despite efforts to improve hypertension detection, treatment, and control, there has been little focus on primary prevention (Meinert & Thomopoulos, 2023).

Identifying individuals at an elevated risk of developing hypertension and target them for early prevention and treatment, health and clinical research initiatives aim to screen and predict hypertension risk. A prediction model can screen for high-risk individuals by estimating their probability of developing hypertension within a certain time frame (Chowdhury & Turin, 2020). While machine learning algorithms have proven successful in various fields, most hypertension prediction models still rely on conventional regression-based models (Chen, Wang, Liu, Yuan, Zhang, Li, et al., 2016; Framingham & Study, 2017; Kadomatsu, Tsukamoto, Sasakabe, Kawai, Naito, Kubo, et al., 2019; Kanegae, Oikawa, Suzuki, Okawara, & Kario, 2018; Lim, Son, Lee, Park, & Cho, 2013; Otsuka, Kachi, Takada, Kato, Kodani, Ibuki, et al., 2015; Paynter, Cook, Everett, Sesso, Buring, & Ridker, 2009; Pearson, LaCroix, Mead, & Liang, 1990; Wang, Liu, Sun, Yin, Li, Ren, et al., 2021; Zhang, 2015).

Hence, the purpose of this research is to construct a straightforward and feasible hypertension risk prediction model and validate it internally trying different machine learning models either as standalone models or in an ensemble fashion.

## 2. METHODS

### *Data source*

This study conducted a cross-sectional analysis utilizing secondary data from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) of the USA Centers for Disease Control and Prevention (CDC). The BRFSS is an open-access online database freely available to the public under the CC0 1.0 Universal (CC0 1.0) Public Domain Dedication license. The original data set consisted of 330 features and a total of 441,456 records.

### *Data analysis*

The preprocessing of data for this study involved several steps including data cleaning, normalization, feature selection, and engineering. Data cleaning involved removing outliers and missing values, while data normalization scaled the data to a specific range. Feature selection involved selecting the most relevant features, and feature engineering combined existing features or created new ones. The final result was 70692 samples selected with 22 features, table 1. Out of the total records, 39,832 subjects were diagnosed with hypertension.

Multiple supervised machine learning algorithms have been evaluated for their ability to predict hypertension by using models either alone or in a combined manner. The models used for this purpose include Gradient Boosting Classifier, XGBoost, Adaboost, MLP, Logistic Regression, Random Forest (RF), XGBRFBoost, KNN, and Decision Trees(DT). These models were constructed using Google Colab and the Python programming language, with the aid of libraries such as numpy, pandas, matplotlib, seaborn, and sklearn. The code imports a dataset, performs data analysis and preprocessing, and separates the data into X and Y variables for machine learning model training and testing. The code evaluates various machine learning models metrics like ROC-AUC, accuracy, F1 score, precision, and recall. Additionally, the code uses an ensemble approach to combine the predictions of several models for a final prediction and calculates the ROC-AUC score. The code also determines the Mutual Information scores for the dataset's features.

**Table 1: Variables descriptions.**

| Variable | Definition |
|---|---|
| Diabetes_binary | 0 (No Diabetes), 1 (Diabetes) |
| HighBP | 0 (No High Blood Pressure), 1 (High Blood Pressure) |
| HighChol | 0 (No High Cholesterol), 1 (High Cholesterol) |
| CholCheck | 0 (No Check), 1 (Cholesterol Check Done) |
| BMI | 1: Underweight (BMI < 18.5 $Kg/m^2$), 2: Normal weight (BMI 18.5-24.9 $Kg/m^2$), 3: Overweight (BMI 25 - 29.9 $Kg/m^2$), 4: Obese (BMI ≥ 30 $Kg/m^2$) |
| Smoker | 0 (No), 1 (Yes) |
| Stroke | 0 (No), 1 (Yes) |
| HeartDiseaseorAttack | 0 (No), 1 (Yes) |
| PhysActivity | 0 (No), 1 (Yes) |
| Fruits | 0 (No), 1 (Yes) |
| Veggies | 0 (No), 1 (Yes) |
| HvyAlcoholConsump | 0 (No), 1 (Yes) |
| AnyHealthcare | 0 (No), 1 (Yes) |
| NoDocbcCost | 0 (No), 1 (Yes) |
| GenHlth | Response to: you say that in general your health is: 1 (Excellent), 2 (Very Good), 3 (Good), 4 (Fair), 5 (Poor) |
| MentHlth | 0 to 30 (number of days in the past 30 days that an individual reported poor mental health) |
| PhysHlth | 0 to 30 (number of days in the past 30 days that an individual reported physical illness or injury) |
| DiffWalk | 0 (No Difficulty), 1 (Serious Difficulty) |
| Sex | 0 (Female), 1 (Male) |
| Age | 13-level category (1: 18-24 y, 2: 25-29 y, 3: 30-34 y, 4: 35-39 y, 5: 40-44 y, 6: 45-49 y, 7: 50-54 y, 8: 55-59 y, 9: 60-64 y, 10: 65-69 y, 11: 70-74 y, 12: 75-79 y, 13: 80 y and above) |
| Education | 6-level category (1-Never attended school or only attended kindergarten, 2, Grades 1 through 8 (Elementary), 3- Grades 9 through 11 (Some high school), 4- Grade 12 or GED (High school graduate), 5- College 1 year to 3 years (Some college or technical school), 6- College 4 years or more (College graduate) |
| Income | 1: <$10 K, 2: $10–$15 K, 3: $15–$20 K, 4: $20–$25 K, 5: $25–$35 K, 6: $35–$50 K, 7: $50–$75 K, 8: >$75 K |

## 3. RESULTS AND DISCUSSION

The data analyzed in this study involves the examination of health-related variables of individuals, with all variables being categorical, table 1. There are 22 features, including Diabetes_binary, which indicates whether an individual has diabetes, with 0 being no diabetes, and 1 diabetes. The most common value is 0, appearing in 50% of the data. HighBP represents whether an individual has high blood pressure, with the most common value being 1 (indicating high blood pressure) appearing in 56% of the data. HighChol indicates high cholesterol levels, with 53% of individuals having high cholesterol. CholCheck represents whether an individual has had a cholesterol check in the past 5 years, with 98% of individuals having had a check. Smoker indicates whether an individual has smoked at least 100 cigarettes in their lifetime, with 52% of individuals reporting no. Stroke and Heart Disease or Attack represent if an individual has had a stroke or coronary heart disease, respectively, with 94% and 85% reporting no. PhysActivity and Fruits represent physical activity and daily fruit consumption, respectively, with 70% and 61% of individuals reporting positive. Veggies indicates daily vegetable consumption, with 79% of individuals reporting positive. HvyAlcoholConsump represents heavy alcohol consumption, with 96% of individuals reporting no. Other features include AnyHealthcare (95% have healthcare coverage), NoDocbcCost (91% have not been unable to see a doctor due to cost), GenHlth (33% rate their health as good), MentHlth (68% report no days of poor mental health), PhysHlth (56% report no days of physical illness), DiffWalk (75% have no difficulty walking) and Sex (54% female). The age distribution shows that the top group comprises 60-64 years old, accounting for 15% of the study population. The average BMI of the sample group is 29.86, signaling that the majority are classified as overweight. The distribution of the BMIs is well spread with a standard deviation of 7.11.

To evaluate the performance of the models used to predict hypertension, a variety of metrics were employed, including ROC-AUC, accuracy, F1 Score, precision, and recall. Table 2 shows the evaluation of all models used in this study. The given statistics in the table show the performance of the individual and ensemble models. For the individual models, the mean values ± standard deviation of Train ROC-AUC, Test ROC-AUC, Accuracy, F1 Score, Precision, and Recall are 0.85 ± 0.08, 0.77 ± 0.05, 0.72 ± 0.03, 0.76 ± 0.03, 0.73 ± 0.02, and 0.79 ± 0.05, respectively. These results indicate that the models have a good average performance, with Train ROC-AUC having the highest mean and accuracy having the lowest mean, but with relatively small standard deviations.

The Gradient Boosting Classifier, XGBBoost, Adaboost and MLP have the highest test ROC-AUC with a value of 0.81, followed by logistic with a value of 0.80. The accuracy of all four models is also the same with a value of 0.74 followed by logistic with a value of 0.73. The Gradient Boosting Classifier, XGBBoost, Adaboost have the highest F1 Score followed by MLP and logistic. Logistic have the highest precision followed by adaboost then both Gradient Boosting Classifier, XGBBoost. Gradient Boosting Classifier, XGBBoost have the highest recall followed by adaboost and logistic then MLP. XGBRFBoost model has slightly lower test ROC-AUC, accuracy, F1 Score and precision values compared to the top five models, but, similar recall as MLP.

The high difference between the Train ROC-AUC and Test ROC-AUC values for the RF and DT models suggests that these models are overfitting to the training data and not performing well on the test data. This means that these models are not generalizing well to new unseen data and therefore, they are not the best models for this task. This is a common issue with decision tree-based models, as they tend to memorize the training data and perform poorly on new unseen data. The Ensemble model which has high Train ROC-AUC value but it also has a large difference between the Train ROC-AUC and Test ROC-AUC values and low-test accuracy, F1 Score, Precision, and Recall values, indicating that it also suffers from the same overfitting issue. The DT model has the lowest test ROC-AUC and accuracy, with a value of 0.64, and it also has the lowest F1 Score, precision, and recall values.

The KNN model has a slightly higher test ROC-AUC and accuracy compared to the DT model, but still lower than the top models.
Overall, the top four models, Gradient Boosting Classifier, XGBBoost, Adaboost and, MLP are the best performance models in this case, with almost similar performance in terms of test ROC-AUC, accuracy, F1 Score, precision, and recall.

**Table 2: Models metrics.**

| Model | Train ROC-AUC | Test ROC-AUC | Accuracy | F1 Score | Precision | Recall |
|---|---|---|---|---|---|---|
| **Gradient Boosting Classifier** | 0.82 | 0.81 | 0.74 | 0.78 | 0.74 | 0.83 |
| **XGBBoost** | 0.82 | 0.81 | 0.74 | 0.78 | 0.74 | 0.83 |
| **Adaboost** | 0.81 | 0.81 | 0.74 | 0.78 | 0.75 | 0.81 |
| **MLP** | 0.81 | 0.81 | 0.74 | 0.77 | 0.76 | 0.79 |
| **Logistic Regression** | 0.80 | 0.80 | 0.73 | 0.77 | 0.74 | 0.81 |
| **Random Forest** | 1.00 | 0.78 | 0.73 | 0.77 | 0.74 | 0.81 |
| **XGBRFBoost** | 0.79 | 0.78 | 0.72 | 0.76 | 0.74 | 0.79 |
| **KNN** | 0.87 | 0.74 | 0.70 | 0.74 | 0.72 | 0.76 |
| **Decision Tree** | 1.00 | 0.64 | 0.64 | 0.67 | 0.69 | 0.66 |
| **mean** | 0.85 | 0.77 | 0.72 | 0.76 | 0.73 | 0.79 |
| **SD** | 0.08 | 0.05 | 0.03 | 0.03 | 0.02 | 0.05 |
| **Ensemble** | 0.96 | 0.80 | 0.74 | 0.77 | 0.76 | 0.79 |

To improve the accuracy of predictions in a model, it is important to identify the most important features. This can be done by analyzing the mutual information scores between various features and the dependent variable, hypertension, figure 1. A high mutual information score indicates a strong relationship between the feature and hypertension, while a low score indicates a weak relationship.
Diabetes_binary and Age have a high mutual information score of 0.09 and 0.07, respectively, signifying a strong relationship with hypertension. GenHlth and HighChol have a moderate relationship with hypertension. BMI has a lower relationship with hypertension compared to HighChol. DiffWalk, HeartDiseaseorAttack, Income, PhysHlth, Education, CholCheck, Stroke, PhysActivity, Veggies, AnyHealthcare, Smoker, Fruits, MentHlth, Sex have low scores, indicating a very weak relationship with hypertension. NoDocbcCost and HvyAlcoholConsump have a score of 0, indicating no relationship.

When deciding which features to include in a model, the mutual information scores should be taken into consideration. Typically, features with low scores are dropped as they do not provide significant information to the model. Thereore, droping BMI HighChol. DiffWalk, HeartDiseaseorAttack, Income, PhysHlth, Education, CholCheck, Stroke, PhysActivity, Veggies, AnyHealthcare, Smoker, Fruits, MentHlth, Sex will improve the perfomacne of the prediction models.

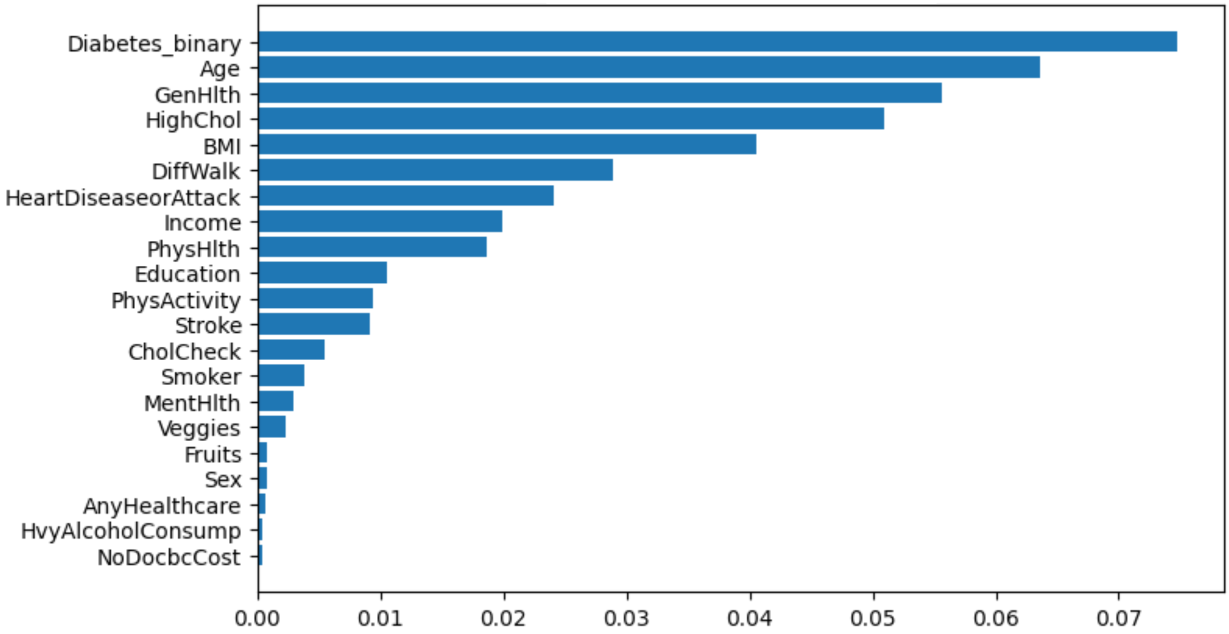


**Figure 1: Mutual information scores of the selected variables.**

In this study, predictive models for hypertension were built using various machine learning models instead of conventional methods. This study reports that Gradient Boosting Classifier, XGBoost, Adaboost, and MLP performed best with a ROC-AUC of 0.81 and accuracy of 0.74 on the test dataset followed by logistic regression. Among these, Gradient Boosting Classifier and XGBoost had slightly higher recall, while MLP had the highest precision followed by Adaboost. Both precision and recall are important when predicting hypertension, as they provide different but complementary measures of accuracy. Precision measures how many of the predicted labels are correct, while recall measures how many of the actual labels are correctly identified. In other words, precision measures how many true positives were identified, while recall measures how many false negatives were identified. Both measures are important when predicting hypertension, as accuracy is key when diagnosing and treating the condition. Therefore, building an ensemble model using these four models may deliver better accuracy.

Several studies have used mathematical techniques and machine learning models to predict risk in healthcare, including decision trees, statistical algorithms, and neural networks (Islam, Ahmed, Uddin, Siddiqui, Malekahmadi, Al Mamun, & Nahavandi, 2021). One study found that neural networks were the best predictor of hypertension, but its results were limited by missing data on obesity (Ture, Kurt, Kurum, & Ozdamar, 2005). Another study used decision trees, logistic regression, and Naive Bayes classifiers to predict hypertension using variables such as obesity, biomarkers, and spirometry indices, but was limited by a lack of data on other factors such as wealth index, education levels, smoking, alcohol use, and physical activity (Heo & Ryu, 2018).

The selected features in this study, including age, diabetes, high cholesterol, and general health, were found to have the strongest relationship with hypertension. This aligns with previous research, where age (Ren, Rao, Xie, Li, Wang, Cui, et al., 2020; Sakr, Elshawi, Ahmed, Qureshi, Brawner, Keteyian, et al., 2018; Kanegae et al., 2018), diabetes (Kshirsagar, Chiu, Bomback, August, Viera, Colindres, et al., 2010; Farran, Channanath, Behbehani, & Thanaraj, 2013; Sakr et al., 2018), cholesterol level (Tayefi, Esmaeili, Karimian, Zadeh, Ebrahimi, Safarian, et al., 2017; (Wu, Pang, & Kwong, 2014,& Wu, Kwong, & Pang, 2015) and BMI (Kshirsagar et al., 2010; Ren et al., 2020; Akdag, Fenkci, Degirmencioglu, Rota, Sermez, & Camdeviren, 2006). were identified as predictors of hypertension in various hypertension risk assessment models.

According to the findings of this study, general health was found to be a predictor of hypertension for the first time. This study showed that subjects who reported poor general health had a higher likelihood of developing hypertension. This study aimed to use non-invasive data to develop machine learning (ML) models to predict hypertension, utilizing the effectiveness and cost-effectiveness of mobile phones and digital technologies, which have been demonstrated in previous studies (Islam & Maddison, 2021; Islam, Peiffer, Chow, Maddison, Lechner, Holle, et al., 2020; Islam, Farmer, Bobrow, Maddison, Whittaker, Dale, et al., 2019; Islam & Tabassum, 2015; Krittanawong, Zhang, Wang, Aydar, & Kitai, 2017).

The results of this study should be interpreted with caution, taking into consideration several limitations. Firstly, only a limited number of variables were included in the models, and data on other risk factors such as family history, race, alcohol consumption, waist-hip ratio, physical activity levels, dietary intake, and biochemical parameters were unavailable, which might have affected the measurement precision. Secondly, the risk factors may have changed since some of the study data was from the 2016 survey. Thirdly, ML models have a weakness in claiming causation. Finally, the models were not externally validated using other data sources, so their results should be interpreted with caution.

Despite these limitations, the primary strength of this study is the use of large-scale nationally representative survey data using ML approaches to predict hypertension.

The findings of this study indicate that machine learning models can effectively predict hypertension using simple information such as age and diabetes, which were found to be among the most significant risk factors in our study population (Ye, Fu, Hao, Zhang, Wang, Jin, et al., 2018; Weng, Reps, Kai, Garibaldi, & Qureshi, 2017). However, future research is necessary to incorporate additional risk factors and biomarkers related to hypertension. These models could be made accessible online or through mobile phone applications, allowing individuals to check their hypertension risk at home by answering basic questions such as age, BMI, and sex. A two-step approach can also be implemented in clinical practice, where the ML model first identifies individuals at risk of hypertension and then a physician confirms the diagnosis and provides appropriate treatment (Ye et al., 2018).

## Conclusion

This study highlights the superiority of ML models compared to traditional statistical techniques when it comes to dealing with complex relationships between variables that cannot be fully comprehended using standard statistics. This has significant implications for hypertension prevention, as these ML models can be applied to population-level data for hypertension screening.

## References:


Akdag, B., Fenkci, S., Degirmencioglu, S., Rota, S., Sermez, Y., & Camdeviren, H. (2006). Determination of risk factors for hypertension through the classification tree method. Advances in therapy, 23, 885-892.

Bromfield S, Muntner P. High blood pressure: The leading global burden of disease risk factor and the need for worldwide prevention programs. Curr. Hypertens. Rep. 2013;15(3):134–136.

Chen, Y., Wang, C., Liu, Y., Yuan, Z., Zhang, W., Li, X., ... & Zhang, C. (2016). Incident hypertension and its prediction model in a prospective northern urban Han Chinese cohort study. Journal of human hypertension, 30(12), 794-800.

Chowdhury, M. Z., & Turin, T. C. (2020). Precision health through prediction modelling: factors to consider before implementing a prediction model in clinical practice. Journal of primary health care, 12(1), 3-9.

Farran, B., Channanath, A. M., Behbehani, K., & Thanaraj, T. A. (2013). Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study. BMJ open, 3(5), e002457.

Framingham, T., & Study, H. (2017). Article annals of internal medicine a risk score for predicting near-term incidence of hypertension. Annals of Internal Medicine, 148(2), 102–110. https://doi.org/10.7326/M16-1751.

Heo, B. M., & Ryu, K. H. (2018). Prediction of prehypertenison and hypertension based on anthropometry, blood parameters, and spirometry. International journal of environmental research and public health, 15(11), 2571.

Islam, S. M. S., & Maddison, R. (2021). Digital health approaches for cardiovascular diseases prevention and management: lessons from preliminary studies. Mhealth, 7.

Islam, S. M. S., & Tabassum, R. (2015). Implementation of information and communication technologies for health in Bangladesh. Bulletin of the World Health Organization, 93, 806-809.

Islam, S. M. S., Ahmed, S., Uddin, R., Siddiqui, M. U., Malekahmadi, M., Al Mamun, A., ... & Nahavandi, S. (2021). Cardiovascular diseases risk prediction in patients with diabetes: Posthoc analysis from a matched case-control study in Bangladesh. Journal of Diabetes & Metabolic Disorders, 20, 417-425.

Islam, S. M. S., Farmer, A. J., Bobrow, K., Maddison, R., Whittaker, R., Dale, L. A. P., ... & Chow, C. K. (2019). Mobile phone text-messaging interventions aimed to prevent cardiovascular diseases (Text2PreventCVD): systematic review and individual patient data meta-analysis. Open Heart, 6(2), e001017.

Islam, S. M. S., Peiffer, R., Chow, C. K., Maddison, R., Lechner, A., Holle, R., ... & Laxy, M. (2020). Cost-effectiveness of a mobile-phone text messaging intervention on type 2 diabetes—A randomized-controlled trial. Health Policy and Technology, 9(1), 79-85.

Kadomatsu, Y., Tsukamoto, M., Sasakabe, T., Kawai, S., Naito, M., Kubo, Y., ... & Wakai, K. (2019). A risk score predicting new incidence of hypertension in Japan. Journal of Human Hypertension, 33(10), 748-755.

Kanegae, H., Oikawa, T., Suzuki, K., Okawara, Y., & Kario, K. (2018). Developing and validating a new precise risk-prediction model for new-onset hypertension: The Jichi Genki hypertension prediction model (JG model). Journal of Clinical Hypertension, 20(5), 880–890.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. Journal of the American College of Cardiology, 69(21), 2657-2664.

Kshirsagar, A. V., Chiu, Y. L., Bomback, A. S., August, P. A., Viera, A. J., Colindres, R. E., & Bang, H. (2010). A hypertension risk score for middle-aged and older adults. The Journal of Clinical Hypertension, 12(10), 800-808.

Lim, N. K., Son, K. H., Lee, K. S., Park, H. Y., & Cho, M. C. (2013). Predicting the risk of incident hypertension in a Korean middle-aged population: Korean genome and epidemiology study. Journal of Clinical Hypertension, 15(5), 344–349.

Meinert, F., Thomopoulos, C., & Kreutz, R. (2023). Sex and gender in hypertension guidelines. Journal of Human Hypertension, 1-8.

Otsuka, T., Kachi, Y., Takada, H., Kato, K., Kodani, E., Ibuki, C., ... & Kawada, T. (2015). Development of a risk prediction model for incident hypertension in a working-age Japanese male population. Hypertension Research, 38(6), 419-425.

Paynter, N. P., Cook, N. R., Everett, B. M., Sesso, H. D., Buring, J. E., & Ridker, P. M. (2009). Prediction of incident hypertension risk in women with currently normal blood pressure. The American journal of medicine, 122(5), 464-471.

Pearson, T. A., LaCroix, A. Z., Mead, L. A., & Liang, K. Y. (1990). The prediction of midlife coronary heart disease and hypertension in young adults: The Johns Hopkins multiple risk equations. American Journal of Preventive Medicine, 6(2 SUPPL.), 23–28.

Ren, Z., Rao, B., Xie, S., Li, A., Wang, L., Cui, G., ... & Ding, S. (2020). A novel predicted model for hypertension based on a large cross-sectional study. Scientific Reports, 10(1), 1-9.

Sakr, S., Elshawi, R., Ahmed, A., Qureshi, W. T., Brawner, C., Keteyian, S., ... & Al-Mallah, M. H. (2018). Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. PLoS One, 13(4), e0195344.

Tayefi, M., Esmaeili, H., Karimian, M. S., Zadeh, A. A., Ebrahimi, M., Safarian, M., ... & Ghayour-Mobarhan, M. (2017). The application of a decision tree to establish the parameters associated with hypertension. Computer methods and programs in biomedicine, 139, 83-91.

Ture, M., Kurt, I., Kurum, A. T., & Ozdamar, K. (2005). Comparing classification techniques for predicting essential hypertension. Expert Systems with

Wang, B., Liu, Y., Sun, X., Yin, Z., Li, H., Ren, Y., ... & Hu, D. (2021). Prediction model and assessment of probability of incident hypertension: the Rural Chinese Cohort Study. Journal of Human Hypertension, 35(1), 74-84.

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. PloS one, 12(4), e0174944.

Wu, T. H., Kwong, E. W. Y., & Pang, G. K. H. (2015, March). Bio-medical application on predicting systolic blood pressure using neural networks. In 2015 IEEE first international conference on big data computing service and applications (pp. 456-461). IEEE.

Wu, T. H., Pang, G. K. H., & Kwong, E. W. Y. (2014, December). Predicting systolic blood pressure using machine learning. In 7th international conference on information and automation for sustainability (pp. 1-6). IEEE.

Ye, C., Fu, T., Hao, S., Zhang, Y., Wang, O., Jin, B., ... & Ling, X. (2018). Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. Journal of medical Internet research, 20(1), e22.

Zhang, W. (2015). Identification of hypertension predictors and application to hypertension prediction in an urban Han Chinese population: A longitudinal study, 2005–2010. Preventing chronic disease, 12.