Indonesian Journal of Electrical Engineering and Computer Science
Vol. 31, No. 3, September 2023, pp. 1794~1802
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v31.i3.pp1794-1802

Evaluation of machine learning techniques for hypertension risk prediction based on medical data in Bangladesh

Md. Asadullah[1], Md. Murad Hossain[1,2], Sabrina Rahaman[1], Muhammad Saad Amin[3], Mst. Sharmin Akter Sumy[4,5], Md. Yasin Ali Parh[4,5], Mohammad Amzad Hossain[6]
[1]Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh
[2]Modeling and Data Science Program, University of Turin, Turin, Italy
[3]Department of Computer Science, University of Turin, Turin, Italy
[4]Department of Statistics, Islamic University, Kushtia, Bangladesh
[5]Department of Bioinformatics and Biostatistics, University of Louisville, Kentucky, USA
[6]Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali, Bangladesh

Article Info

Article history:

Received Sep 11, 2022
Revised Jun 4, 2023
Accepted Jun 17, 2023

Keywords:

Classification
Hypertension
Machine learning
Performance
Receiver operating characteristic

ABSTRACT

Hypertension in Bangladesh is a leading cause of cardiovascular diseases, stroke, and kidney failure, resulting in significant morbidity and mortality. Preventive measures and simple health practices can effectively reduce hypertension and its complications. This study utilizes machine learning algorithms (naive Bayes, support vector machine, logistic regression, random forest) to predict hypertension in high-risk individuals. The proposed hybrid model achieves a prediction accuracy of 78.17%, surpassing other machine learning methods. Random forest has the highest accuracy among the individual algorithms at 73.86%. Classification performance is evaluated using sensitivity, specificity, precision, and F-score, along with receiver operating characteristic analyses and confusion matrices through 10-fold cross-validation. These findings emphasize the importance of managing risk factors for better population health and highlight the efficacy of the hybrid model in hypertension prediction. The study underscores the significance of preventive measures in reducing the burden of hypertension-related diseases and improving overall well-being.

Corresponding Author:

Md. Murad Hossain
Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University
Gopalganj-8100, Bangladesh
Email: mdmurad.hossain@unito.it

1. INTRODUCTION

An educational clinical center used statistical and machine learning techniques to examine the digital health records of 14,360 adult hypertension patients. Finding predictors and the timing of lifestyle modifications was the goal [1]. Using data from Kuwait, the study developed classification models and risk assessment tools for diabetes, high blood pressure, and comorbidities [2]. Numerous techniques were used, including support vector machines (SVMs), multifactor dimensionality reduction, logistic regression, and K-nearest neighbors (KNNs). Fivefold cross-validation was used to get generalization errors and accuracies [3]. In this work, pulse waves from both the hypertensive and healthy groups were classified and predicted using a machine learning technique. By removing noise with K-means, the goal was to evaluate how pulse waves affected the accuracy and stability of the machine learning model [4]. In order to enhance therapy personalization and patient outcomes, the study used decision trees and neural networks to uncover parameters that contribute to the effectiveness of high blood pressure medicine treatment for a broad group of patients [5]. The study used supervised principal component analysis to identify systolic motion patterns that were highly predictive of survival. The researchers assessed the precision of survival prediction using the area under the curve with time-structured receiver operating characteristic analysis for 1-year survival [6]. The aim of the study was to assess the performance of the machine learning algorithms catboost, logistic regression, naive Bayes, random forest, and SVM in the screening of anxiety and depression among seafarers. The evaluation was performed using Python programming [7].

The study's objective was to find out how machine learning methods may enhance pulmonary hypertension (PH) prediction. Five machine learning techniques, including random forests of classification trees, random forests of regression trees, lasso penalized logistic regression, boosted classification trees, and SVMs, were applied by the researchers. The study's goal was to assess how accurately these algorithms predicted PH [8]. This study provides a comprehensive review of machine learning and artificial intelligence (AI) approaches for non-invasive cuff-less blood pressure estimation using the photoplethysmography (PPG) method, including challenges and limitations. The study's goals were to assess the efficiency and practicality of blood pressure measurement methods based on PPGs and to show how machine learning and AI could improve accuracy and dependability [9]. The study's objective was to evaluate and contrast the performance of several machine learning approaches in locating those at risk of hypertension. Logit boost, Bayesian network classifier, locally weighted naive Bayes, artificial neural network, SVM, and random tree forest were some of the methods used. The study's objective was to assess the efficacy and precision of different methods for foretelling the likelihood of developing high blood pressure [10]. The potential for AI to offer clinical specialists in high blood pressure management useful insights is examined in this paper. In order to increase the accuracy and precision of high blood pressure management and patient outcomes, it focuses on the application of AI to predict clinical outcomes in vast and complicated datasets [11]. This study uses chest sound recordings and machine learning techniques to suggest Kullback-Leibler divergence as a possible method for anticipating hypertension. The objective of the study is to examine how well this method predicts hypertension and increases diagnostic precision [12].

This study's objective was to improve non-communicable disease intervention strategies through data analysis and the presentation of a practical, individualized, and predictive model. The goal of the model was to pinpoint individuals who would later be at risk for non-communicable diseases, enabling earlier intervention and better health outcomes [13]. Based on the prediction model, a mobile application has been created that combines clever tactics with an artificial neural network (ANN) that uses the multi-layer perceptron method. The purpose of the app is to help expectant mothers identify the type of high blood pressure they are dealing with. The accuracy and reliability of the diagnosis can be increased with the employment of intelligent methods and ANN [14]. This study analyzes four machine learning techniques to identify various forms of high blood pressure based on individual features and data (C4.5 direct torque control (DTC), random forest, latent dirichlet allocation (LDA), and least-squares support-vector machines (LSVM)). The objective is to increase the high blood pressure classification's accuracy and dependability [15]. In two racially and ethnically diverse urban groups, this study examines the precision of self-reported survey data in determining the prevalence of clinically diagnosed high blood pressure. The study also suggests a method for correcting self-reported data to represent the incidence of clinical hypertension more accurately [16].

The purpose of this study is to discover prehypertension and high blood pressure risk factors in middle-aged Korean people. The importance of prehypertension and high blood pressure was assessed using binary logistic regression analysis, and prediction models were created using logistic regression, naive Bayes, and decision trees. Improved knowledge of high blood pressure and prehypertension in this population is the aim [17]. To predict high blood pressure and diabetes risk, respectively, a combination of conditional decision-making and machine-learning methods is applied. Using supervised device learning, where a device is trained to anticipate the patient's diabetes and high blood pressure, it is possible to learn type algorithms [18]. In Bangladesh, where the condition is on the rise throughout Southeast Asia, this study examines the factors that contribute to high blood pressure in adults. The goal is to better understand the elements that lead to hypertension in this population [19]. This study proposes a hybrid prediction model (HPM) for early type 2 diabetes and high blood pressure prediction based on individual risk factors. The objective is to create a more accurate and effective prediction model under these conditions by utilizing data cleaning techniques and random forest classification [20]. Identifying the prevalence, underlying factors, and socioeconomic disparities of undiagnosed hypertension in Nepal is the aim of this study. Using logistic regression analysis, it is possible to identify the causes of hidden high blood pressure. The study provides information on how to diagnose and manage hypertension in Nepal, particularly in disadvantaged people [21] more effectively. Whether an increase in in-clinic mortality is associated with hypertensive patients' elevated systolic blood pressure (SBP) in the emergency department (ED) was the aim of this investigation [22]. The purpose of this study was to determine the reasons why non-Hispanic Whites and Hispanics and Latinos spent less money on hypertension management, detection, and treatment. The study's major objective was to examine the prevalence, awareness, management, and treatment of hypertension among Hispanic/Latino adults [23]. This paper's main objective is to explain hypertension and the factors influencing treatment adherence in male and female hypertensive patients living in rural Bangladesh [24]. The findings of this study point to a considerable increase in domestic violence, which is highly correlated with two factors: families' income levels during the COVID-19 epidemic and individual family individuals' levels of education [25].

However, the above articles use different summarizing approaches, machine learning techniques, effectiveness along with SVM, KNN, ANN, recency frequency monetary (RFM), AI, and many others. Additionally, some papers used binary logistic regression and some predictive models for classification and forecasting hypertension. But there is no available paper that explained hypertension using a prediction model through machine learning techniques in the perspectives of Bangladeshi data. Hypertension is also caused by different kinds of serious diseases as like as heart disease and diabetes. Every 12 months, a large number of economic fees have to spend from our national price range for high blood pressure purposes. So, high blood pressure prevention is a first-rate problem these days in Bangladesh. During this paper, we try to predict hypertension through a proposed model. Even we strive to degree the overall performance of numerous machine learning tools within comparisons in the perspective of validation and accuracy all through the specified Bangladeshi records.

2. METHOD

2.1. Dataset collection and processing

In this paper, we used secondary data which is collected from a private hospital in the year 2014-2015. In our data set, there exists 9,620 patients and for each patient different information (which is known as a variable). In our dataset consists of some distinct medical variables, such as age, sex, heredity, body mass index (BMI), diabetes, and some demographical variable like occupation, income, and education. Describe all variables in Table 1.

Table 1. Dataset description

Variable names	Description
Age	Patients age in years
Sex	Sex of the patients (male or female)
Occupation	Occupation of the respondent (service, business, farmer, housewife, retired and others)
Heredity	Heredity means whether the previous generation had hypertension (yes or no)
Area	Respondence residence location (rural or urban)
BMI	Body mass index
Education level	Education qualification of the patients (literate or illiterate)
Diabetes	Patients have diabetes (yes or no)
Physical exercise	Whether the patient exercises regularly (yes or no)
Income	Income level of patients or family (low, medium, high)
Weight	Patients weight level (underweight, normal weight, overweight, obesity)
Hypertension	Decision by doctor having hypertension or not (yes or no)

This dataset has 3,853 patient data where all the patients are female, and 5,767 patients are male. Most of the patients in our dataset are in the urban area compared to the rural area. Using the wrapper method feature selection method, we select 11 variables in our analysis. We use R package version 3.6.3 for the data processing and analysis. Moreover, we used the wrapper method to find the important features for our predictive model.

2.2. Experimental setup

Figure 1 describes the different steps involved in our study. In the first step, we pre-processed the dataset for our predictive model—the next step the dataset into two parts, i.e., training, testing datasets. In our study, three prominent classification techniques and a hybrid model have been applied. Therefore, the best predictive model approves as a future predictive model.

2.3. Proposed hybrid model

The concept of hybrid or ensemble approaches involves utilizing multiple single models to generate effective discriminatory rules. To implement this, we can create ensembles of machine learning algorithms, which involves combining the predictions of four caret models through a technique called stacking. This suggests that each model is skilled in unique ways, and by combining their strengths, a new classifier can optimize accuracy.

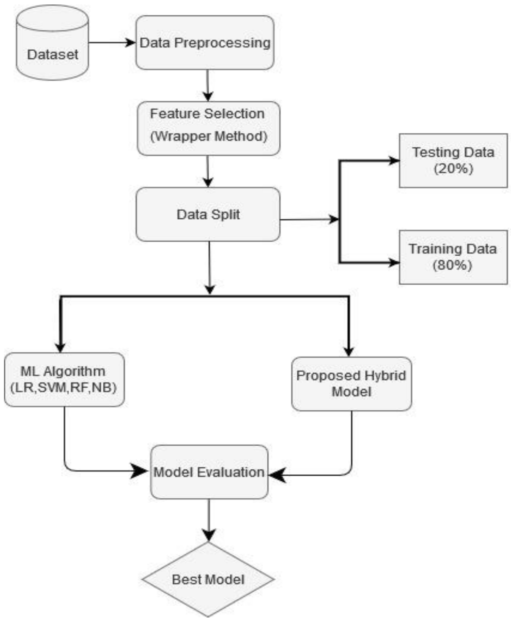

Figure 1. Flowchart for proposed hybrid model

2.4. Comparison of the previous research

The previous study's use different model and compared to the accuracy in the comparison of previous research (CPR) section. It's worth noting that we concentrated on the algorithms. The previous study focused on the machine learning algorithm but not proposed any hybrid model. We compared our work to the top six current studies on hypertension prediction. We displayed our comparison in Table 2.

Table 2. Comparison of previous research with our proposed hybrid model

Title of the previous study	Algorithm	Use proposed model
"Comparison of machine learning methods for the arterial hypertension diagnostics" [1]	DA KNN SVM DT NB	No
"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." [3]	LR KNN MDR SVM	No
"A study of machine-learning classifiers for hypertension based on radial pulse wave." [4]	RF SVM AdaBoost GBT KNN	No
"Machine learning of big data in gaining insight into successful treatment of hypertension." [5]	DT FCNN	No
"Uses and opportunities for machine learning in hypertension research." [11]	CART NN Boosting SVM CNN	No
"A predictive model for hypertension diagnosis using machine learning techniques." [14]	Fuzzy logic ANN Multilayer Perceptron DDT SVM	No
Our research.	LR SVM RF Hybrid	Yes

2.5. Support vector machines

Support vector networks and SVMs are supervised learning models that use learning techniques to classify and evaluate data for regression analysis. SVMs can do nonlinear classification in addition to linear classification by implicitly translating inputs into high-dimensional feature spaces using the kernel method. Introduced by Vladimir Vapnik and Alexey Chervonenkis, SVMs attempt to classify datasets into two classes by passing a linearly separable hyperplane. Ultimately, the model can accurately estimate the target groups or labels for new cases. Assuming perfect data separation, SVMs can effectively classify datasets. After that, we can improve the following. Minimize $\| \omega \|^2$, subject to:

$$(\mathrm{w}.\mathrm{x_i} + \mathrm{b}) \geq 1, \mathrm{if}\, \mathrm{y_i} = 1$$

$$(\mathrm{w}.\mathrm{x_i} + \mathrm{b}) \leq -1, \mathrm{if}\, \mathrm{y_i} = -1$$

the last two constraints can be compacted to:

$$y_i(w.x_i + b) \geq 1.$$

2.6. Naïve Bayes

The naive Bayesian model is a simple and effective technique for analyzing very large data sets. The Bayes theorem is used to categorize data under the presumption of predictor independence. The naive Bayes classifier thinks that whether a feature is present in a class has no bearing on whether another feature is present. The main challenge with this approach is calculating the class conditional density, which is typically based on the available data points. We can calculate the class conditional density of ambiguous data artifacts by estimating probability distributions for unknown classification issues. Bayes Theorem provides a way for P(c), P(x), and P(x) to measure posterior likelihood. Look at the equation underneath:

$$P(c|x) = \frac{P(X|C)P(c)}{P(x)}$$

there, P(c) is the posterior likelihood of class (target) given predictor (attribute), P(c) is the prior likelihood of class, P(x) is the likelihood of predictor given class, and P(x) is the predictor chance.

2.7. Random forest model

A flexible and user-friendly machine learning technique, the random forest approach can produce outstanding results even without hyper-parameter adjustment. Due to its versatility and simplicity, it is commonly used for problems involving classification and regression. This article gives a general summary of the algorithm's operation, distinguishing characteristics, and practical application. Being adaptable to both classification and regression issues, which is a major feature of many modern machine learning systems, is an additional advantage of random forest. The hyperparameters for decision trees and bagging classifiers are also applicable to a random forest model. Each tree in a random forest classifier is created using a random vector that is individually sampled from the input vector, and it mixes several tree classifiers. Then, each tree votes to categorize an input vector according to the class that appears the most frequently.

2.8. Logistic regression

A popular machine learning algorithm is a logistic regression, second only to linear regression in popularity. While there are similarities between the two, the primary distinction is in their applications. Linear regression is used for predicting or forecasting numerical values, whereas logistic regression is commonly used for classification tasks. Let $Y_i, i = 1,2, \dots\dots. N$ be a binary outcome $(0|1)$ from Bernoulli$(1, \pi_i)$ with $\pi_i = Pr[Y_i = 1]$. The logistic regression model can be defined as:

$$logit = [Pr(Y_i = 1|x_i)] = log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i = \beta^T x_i$$

2.9. Performance evaluation

We need to assess performance assessment after fitting various machine learning methods. Based on the following standards, we assess machine learning algorithms for categorization using different techniques of performance measures mentioned in Table 3. For the binary classification problem, we have samples of hypertension risk belonging to two classes: YES or NO. Here are four important terms which are used to calculate different evaluation criteria.

Table 3. Model evaluation technique

Technique names	Formula
Accuracy =	$\frac{TP+TN}{TP+FP+TN+FN}$
Precision =	$\frac{TP}{TP+FP}$
Sensitivity/Recall =	$\frac{TP}{TP+FN}$
Specificity =	$\frac{TN}{TN+FP}$
F1-Score =	$2*\frac{Precision*Recall}{Precision+Recall}$

3. RESULTS AND DISCUSSION

This study employed a novel approach to assess the effectiveness of four machine learning classification algorithms and one hybrid model in predicting hypertension. The performance of each model was evaluated using five criteria, including specificity, recall, precision, F1 score, and area under curve (AUC) value. In Table 4 represent the different machine learning algorithm performance evaluation criteria.

Table 4. Performance measurements for classification technique

Algorithm	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
RF	0.7386	0.8432	0.5793	0.8432	0.8432	0.8099
SVM	0.7344	0.8449	0.5662	0.8449	0.8449	0.7976
NB	0.6913	0.7227	0.6435	0.7227	0.7227	0.7546
LR	0.6887	0.7812	0.5478	0.7812	0.7812	0.7406
Hybrid	0.7817	0.8751	0.6396	0.8751	0.8751	0.8634

Table 4 represents the performance of four supervised machine learning techniques and one hybrid model for hypertension prediction. We used a 10-fold cross-validation approach to evaluate the performance of the prediction model. As a result, random forest exhibits the highest performance among four machine learning algorithms with an accuracy of 73.86% and the lowest performance in logistic regression with an accuracy of 68.87%. We compare all performance among four machine learning algorithms in the existing criteria, and in all cases, the random forest model is the much better performer algorithm. But our proposed hybrid model provides better accuracy (78.18%) than the random forest model. We also show the machine learning algorithms' performance through the ROC curve. From Figure 2, five different colors represent five different predictive models in our medical dataset.

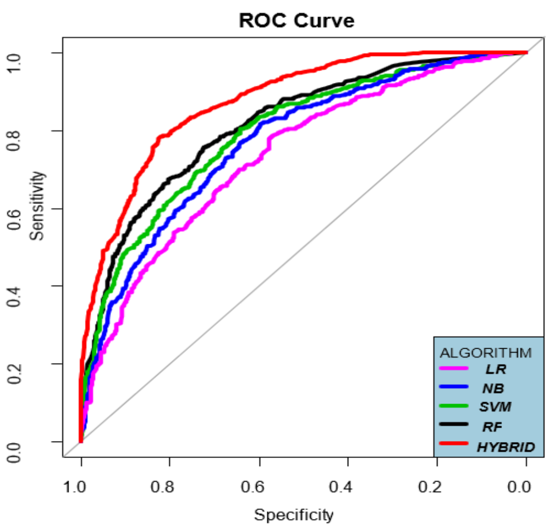

Figure 2. ROC Curve for different machine learning algorithm compared to hybrid model

The hybrid model we've proposed is represented by the red hue and performs better than any other predictive models. In ROC curve analysis, it provides the AUC value, which is included in Table 4. It is clear that AUC=0.8634, which indicates that our proposed model predictions are 86.34% correctly identify the patients who have hypertension. That means our proposed hybrid model accurately predicts who has the possibility of hypertension. But logistic regressions is less accurate in predicting or identifying the patients who have hypertension. The AUC value for logistic regression is AUC=0.7406 which is the lowest among all predictive models.

4. CONCLUSION

By analyzing the experimental result, it is concluded that our proposed hybrid model obtains AUC's highest accuracy, 78.18%, compared to all machine algorithms and also provides the highest AUC value is 86.34%. The AUC value in the proposed model is close to 1, which indicates the more accurate prediction. Therefore, our proposed hybrid model more accurately identifies the patients who actually have hypertension. The proposed model shows the best results under the ROC curve regarding classification accuracy. The outcomes show that the suggested model performs better than other models with greater accuracy and less classification error. However, the literature assessment suggests that there has only been minimal progress in developing a predictive model for identifying hypertension patients. Finally, we can say that for all machine learning algorithms, our proposed hybrid model reduces classification error, and the best predictive model compares to all machine learning algorithms.

ACKNOWLEDGEMENTS

We thank Dr. MD Matiur Rahaman for assistance with guidance and suggestion, and for comments that greatly improved the manuscript.

REFERENCES

[1] V. S. Kublanov, A. Y. Dolganov, D. Belo, and H. Gamboa, "Comparison of machine learning methods for the arterial hypertension diagnostics," *Applied Bionics and Biomechanics*, vol. 2017, pp. 1–13, 2017, doi: 10.1155/2017/5985479.

[2] P. Santhanam and R. S. Ahima, "Machine learning and blood pressure," *Journal of Clinical Hypertension*, vol. 21, no. 11, pp. 1735–1737, Nov. 2019, doi: 10.1111/jch.13700.

[3] B. Farran, A. M. Channanath, K. Behbehani, and T. A. Thanaraj, "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*, vol. 3, no. 5, p. e002457, May 2013, doi: 10.1136/bmjopen-2012-002457.

[4] Z. Y. Luo *et al.*, "A study of machine-learning classifiers for hypertension based on radial pulse wave," *BioMed Research International*, vol. 2018, pp. 1–12, Nov. 2018, doi: 10.1155/2018/2964816.

[5] G. Koren, G. Nordon, K. Radinsky, and V. Shalev, "Machine learning of big data in gaining insight into successful treatment of hypertension," *Pharmacology Research and Perspectives*, vol. 6, no. 3, p. e00396, Jun. 2018, doi: 10.1002/prp2.396.

[6] T. J. W. Dawes *et al.*, "Machine learning of threedimensional right ventricular motion enables outcome prediction in pulmonary hypertension: A cardiac MR imaging study," *Radiology*, vol. 283, no. 2, pp. 381–390, May 2017, doi: 10.1148/radiol.2016161315.

[7] A. Sau and I. Bhakta, "Screening of anxiety and depression among the seafarers using machine learning technology," *Informatics in Medicine Unlocked*, vol. 16, p. 100149, 2019, doi: 10.1016/j.imu.2018.12.004.

[8] T. Seidler *et al.*, "A machine learning approach for the prediction of pulmonary hypertension," *Journal of the American College of Cardiology*, vol. 73, no. 9, p. 1589, Mar. 2019, doi: 10.1016/s0735-1097(19)32195-3.

[9] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," *Biomedical Signal Processing and Control*, vol. 58, p. 101870, Apr. 2020, doi: 10.1016/j.bspc.2020.101870.

[10] S. Sakr *et al.*, "Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford exercise testing (FIT) Project," *PLoS ONE*, vol. 13, no. 4, p. e0195344, Apr. 2018, doi: 10.1371/journal.pone.0195344.

[11] D. Amaratunga, J. Cabrera, D. Sargsyan, J. B. Kostis, S. Zinonos, and W. J. Kostis, "Uses and opportunities for machine learning in hypertension research," *International Journal of Cardiology: Hypertension*, vol. 5, p. 100027, Jun. 2020, doi: 10.1016/j.ijchy.2020.100027.

[12] A. Clim, R. D. Zota, and G. Tinica, "The kullback-leibler divergence used in machine learning algorithms for health care applications and hypertension prediction: a literature review," *Procedia Computer Science*, vol. 141, pp. 448–453, 2018, doi: 10.1016/j.procs.2018.10.144.

[13] M. Hu, Y. Nohara, Y. Wakata, A. Ahmed, N. Nakashima, and M. Nakamura, "Machine learning based prediction of non-communicable diseases to improving intervention program in Bangladesh," *European Journal for Biomedical Informatics*, vol. 14, no. 4, 2018, doi: 10.24105/ejbi.2018.14.4.5.

[14] M. A. J. Tengnah, R. Sooklall, and S. D. Nagowah, "A predictive model for hypertension diagnosis using machine learning techniques," in *Telemedicine Technologies*, 2019, pp. 139–152, doi: 10.1016/B978-0-12-816948-3.00009-X.

[15] M. Nour and K. Polat, "Automatic classification of hypertension types based on personal features by machine learning algorithms," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–13, Jan. 2020, doi: 10.1155/2020/2742781.

[16] G. Mentz, A. J. Schulz, B. Mukherjee, T. E. Ragunathan, D. W. Perkins, and B. A. Israel, "Hypertension: Development of a prediction model to adjust self-reported hypertension prevalence at the community level," *BMC Health Services Research*, vol. 12, no. 1, p. 312, Dec. 2012, doi: 10.1186/1472-6963-12-312.

[17] B. M. Heo and K. H. Ryu, “Prediction of prehypertenison and hypertension based on anthropometry, blood parameters, and spirometry,” *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, p. 2571, Nov. 2018, doi: 10.3390/ijerph15112571.
[18] S. P. Chatrati *et al.*, “Smart home health monitoring system for predicting type 2 diabetes and hypertension,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 3, pp. 862–870, Mar. 2022, doi: 10.1016/j.jksuci.2020.01.010.
[19] M. A. B. Chowdhury, M. J. Uddin, M. R. Haque, and B. Ibrahimou, “Hypertension among adults in Bangladesh: evidence from a national cross-sectional survey,” *BMC Cardiovascular Disorders*, vol. 16, no. 1, Dec. 2016, doi: 10.1186/s12872-016-0197-3.
[20] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, “Hybrid prediction model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, synthetic minority over sampling technique (SMOTE), and random forest,” *Applied Sciences (Switzerland)*, vol. 8, no. 8, p. 1325, Aug. 2018, doi: 10.3390/app8081325.
[21] M. R. Haider and R. Das Gupta, “Inequalities in undiagnosed hypertension among adult Nepalese population: evidence from a nationally representative survey,” *International Journal of Cardiology: Hypertension*, vol. 5, p. 100026, Jun. 2020, doi: 10.1016/j.ijchy.2020.100026.
[22] Klang, Eyal, *et al.* "Association of normal systolic blood pressure in the emergency department with higher in-hospital mortality among hypertensive patients." *The Journal of Clinical Hypertension*, vol. 21, no. 12, p. 1841-1848, 2019, doi: 10.1111/jch.13727.
[23] C. M. Lora *et al.*, “Prevalence, awareness, and treatment of hypertension in hispanics/Latinos with CKD in the hispanic community health study/study of Latinos,” *Kidney Medicine*, vol. 2, no. 3, pp. 332–340, May 2020, doi: 10.1016/j.xkme.2020.02.005.
[24] M. A. Khanam, W. Lindeboom, T. L. P. Koehlmoos, D. S. Alam, L. Niessen, and A. H. Milton, “Hypertension: Adherence to treatment in rural Bangladesh - findings from a population-based study,” *Global Health Action*, vol. 7, no. 1, p. 25028, Dec. 2014, doi: 10.3402/gha.v7.25028.
[25] M. M. Hossain *et al.*, “Prediction on domestic violence in bangladesh during the covid-19 outbreak using machine learning methods,” *Applied System Innovation*, vol. 4, no. 4, p. 77, Oct. 2021, doi: 10.3390/asi4040077.

BIOGRAPHIES OF AUTHORS

Md. Asadullah born in Rajshahi, Bangladesh in 1993. He received B.Sc., M.Sc. degrees in Statistics from Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh in 2016 and 2018 respectively. He also worked as an data intelligence officer in a private research firm in Bangladesh. His research interests include machine learning, public health and natural language processing. He can be contacted at email: asadullahstat@gmail.com.

Md. Murad Hossain received the B.Sc. and M.Sc. degrees in statistics from the Jahangirnagar University, Savar, Dhaka, Bangladesh, in 2010 and 2011, respectively. He is currently working toward the Ph.D. degree in modeling and data science with the Modeling and Data Science Program, University of Turin, Turin, Italy. He is currently an Assistant Professor (on leave) with the Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh. His research interests include NLP, statistics, data science, machine learning, and public health. He can be contacted at email: mdmurad.hossain@unito.it.

Sabrina Rahaman was born in Faridpur, Bangladesh in 1990. She received B.Sc., M.Sc. in statistics from Jahangirnagar University in 2012 and 2013 respectively. She also worked as an Assistant Professor in Department of Statistics at Bangabandhu Sheikh Mujibur Rahman Science and Technology University. Her research interest includes time series analysis, probability distribution, and biostatistics. She can be contacted at email: sabrinarahaman.ju.7@gmail.com.

Muhammad Saad Amin was born in Pakistan in 1996. He received his bachelor's and master's degree in computer engineering from University of Engineering and Technology Taxila and University of Lahore, Pakistan in 2017 and 2020 respectively. Currently he is doing his Ph.D. in Computer Science from University of Turin, Italy. His research interests include machine learning, neural networks, and natural language processing and generation. He can be contacted at email: muhammadsaad.amin@unito.it.

Mst. Sharmin Akter Sumy received the B.S., and M.S., degrees in Statistics from Jahangirnagar University, Bangladesh and Ball State University (BSU), USA, in 2010, and 2020 respectively. She worked on subgroup identification of individuals that have a differential effect of cardiorespiratory fitness (CRF) on mortality, meta-analysis, and several projects on Biostatistics. She has more than 10 publications, and her interests include causal Inference, subgroup identification for personalized medicine, survival analysis, and longitudinal data analysis. She can be contacted at email: m0sumy01@louisville.edu.

Md. Yasin Ali Parh received the B.S., and M.S. degrees in Applied Statistics from University of Dhaka, Bangladesh in 2011 and 2012 respectively. He completed his second MS degree from Ball State University, US in 2020. Currently, he is working in Louisville Metro Public Health and Wellness vaccine inequality project. He has authored a number of peer-reviewed articles, and his research focuses are probability theory, Bayesian inference, Markov chain Monte Carlo, subgroup identification for personalized medicine, survival analysis, and data mining. He can be contacted at email: mdyasinali.parh@louisville.edu.

Dr. Mohammad Amzad Hossain received B.Sc. and M.Sc. in Information and Communication Technology (ICT) from the Islamic University (IU), Kushtia, Bangladesh, in 2010 and 2011, respectively. He also received his Structured Ph.D. in the School of Computer Science, National University of Ireland Galway (NUIG), Galway, Ireland. He is currently an Associate Professor in the Department of Information and Communication Engineering (ICE), Noakhali Science and Technology University, Noakhali, Bangladesh. In 2018, Amzad has awarded the prestigious College of Science and Engineering postgraduate research Scholarship. His research interests include spectrum sensing, machine learning and data analysis, image processing, MIMO based cognitive radio networks, cognitive radio based internet of things (CR-IoT) networks. He can be contacted at email: amzad@nstu.edu.bd.