RESEARCH ARTICLE

Optimizing hypertension prediction using ensemble learning approaches

Isteaq Kabir Sifat, Md. Kaderi Kibria 65*

Department of Statistics, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh

* kibria.stt@tch.hstu.ac.bd



OPEN ACCESS

Citation: Sifat IK, Kibria M.K (2024) Optimizing hypertension prediction using ensemble learning approaches. PLoS ONE 19(12): e0315865. https://doi.org/10.1371/journal.pone.0315865

Editor: Tomo Popovic, University of Donja Gorica, MONTENEGRO

Received: July 11, 2024

Accepted: December 2, 2024

Published: December 23, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0315865

Copyright: © 2024 Sifat, Kibria. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: https://figshare.com/s/a709a390ecd276046607?file=41735691.

Funding: The author(s) received no specific funding for this work.

Abstract

Hypertension (HTN) prediction is critical for effective preventive healthcare strategies. This study investigates how well ensemble learning techniques work to increase the accuracy of HTN prediction models. Utilizing a dataset of 612 participants from Ethiopia, which includes 27 features potentially associated with HTN risk, we aimed to enhance predictive performance over traditional single-model methods. A multi-faceted feature selection approach was employed, incorporating Boruta, Lasso Regression, Forward and Backward Selection, and Random Forest feature importance, and found 13 common features that were considered for prediction. Five machine learning (ML) models such as logistic regression (LR), artificial neural network (ANN), random forest (RF), extreme gradient boosting (XGB), light gradient boosting machine (LGBM), and a stacking ensemble model were trained using selected features to predict HTN. The models' performance on the testing set was evaluated using accuracy, precision, recall, F1-score, and area under the curve (AUC). Additionally, SHapley Additive exPlanations (SHAP) was utilized to examine the impact of individual features on the models' predictions and identify the most important risk factors for HTN. The stacking ensemble model emerged as the most effective approach for predicting HTN risk, achieving an accuracy of 96.32%, precision of 95.48%, recall of 97.51%, F1-score of 96.48%, and an AUC of 0.971. SHAP analysis of the stacking model identified weight, drinking habits, history of hypertension, salt intake, age, diabetes, BMI, and fat intake as the most significant and interpretable risk factors for HTN. Our results demonstrate significant advancements in predictive accuracy and robustness, highlighting the potential of ensemble learning as a pivotal tool in healthcare analytics. This research contributes to ongoing efforts to optimize HTN prediction models, ultimately supporting early intervention and personalized healthcare management.

Introduction

Hypertension (HTN), characterized by elevated blood pressure beyond normal ranges, is a major public health issue affecting adult worldwide [1, 2]. It significantly increases the risk of

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: HTN, hypertension; HHTN, history of hypertension; ML, machine learning; LR, logistic regression; ANN, artificial neural network; RF, random forest; XGB, extreme gradient boosting; LGBM, light gradient boosting machine; NB, naïve base; AUC, area under the curve; BMI, body mass index; SHAP, SHalpey Aditive exPlanaions; WHO, World Health Organization; DT, decision Trees; NB, naïve bayes; BFS, Boruta-based feature selection; LASSO, Least Absolute Shrinkage and Selection Operator; ADASYN, Adaptive synthetic; CRT, classification and regression task; OR, odds ratio; GOSS, gradient-based one-side sampling; EFB, exclusive feature bundling.

cardiovascular disease, coronary heart disease, stroke, kidney damage, and other serious complications if left untreated [3, 4]. HTN is the leading causes of premature death globally and affects more than one in four men and one in five women [4]. Due to its high prevalence and association with chronic kidney disease, HTN poses a major global health challenge [5–7]. As a primary risk factor for cardiovascular diseases, it contributes to rising healthcare costs and loss of productivity [8]. According to the World Health Organization (WHO), HTN is the third greatest cause of death worldwide, responsible for one in every eight fatalities [9]. Currently, an estimated 1.3 billion people globally, including 116 million Americans, live with HTN [10]. Hypertensive individuals face a 2-4 times greater risk of developing heart disease, peripheral vascular disease, and stroke [11, 12], which exacerbate the economic burden of outof-pocket expenses and contribute to the leading causes of morbidity, mortality, and disability [13–15]. By 2025, approximately 1.56 billion adults aged 30 to 79 are projected to have HTN, with nearly two-thirds of them residing in low and middle income countries [16]. Given the high prevalence, controlling and predicting HTN at earlier stage is crucial. Early identification of interpretable risk factors in HTN patients is crucial for enabling timely prevention and intervention. Thus, detecting, diagnosis and understanding the risk factors associated with HTN are critical for effective management and treatment.

Modeling to predict the risk of acquiring HTN can aid in identifying significant risk factors contributing to HTN, offering reliable estimates of future HTN risk [17], and identifying individuals at high risk who may benefit from medical care and adopting healthy behaviors to prevent HTN [18-20]. Numerous prediction models have been created over time to forecast the risk of HTN in the general population. Models were created using either a contemporary machine learning technique or a conventional regression-based approach [21]. Primarily, prior research has examined conventional linear models, like logistic regression (LR) and the Cox proportional hazard model, in order to determine the risk factors significantly linked with HTN [22-24]. Also, several studies applied some machine learning techniques and measure their HTN prediction accuracy. A ML-base prediction study has explored a plethora of algorithms, with RF, K-Nearest Neighbors (KNN), DT, and Naive Bayes (NB) models showcasing promising results, with RF boasting an impressive accuracy of 80.12% [25]. In a subsequent study, RF, CatBoost, MLP Neural Network, and LR were used, with RF achieving an accuracy of 92% [26]. Medical data-based study, SVM, C4.5, RF and XGBoost methods were applied and 94.36% accuracy was achieved by using the XGBoost method [27]. However, a study utilizing RF, LR, ANN, and XGBoost models on Ethiopian data yielded a comparatively modest accuracy of 88.81% for XGBoost [28], which is lower than that reported in previous studies. Improving prediction performance for HTN data is therefore necessary. To enhance the performance of these classifiers, several strategies can be employed. Recent research has highlighted the efficacy of the Light Gradient Boosting Machine (LGBM) learning algorithm, which utilizes tree-based learning techniques and has outperformed existing machine learning algorithms [29]. Consequently, this study aims to apply this updated algorithm and their stacking model approach to enhance prediction accuracy using Ethiopia data, to advance the understanding and predictive capabilities in combating HTN. Although stacking models are wellestablished in machine learning, our study applies these methods in the specific context of HTN risk prediction, an area where their potential remains underexplored. The key contribution of our work lies in demonstrating how integrating multiple models through an ensemble approach significantly boosts predictive accuracy compared to single-model methods. Although the theoretical foundation of ensemble learning may not be novel, the improvement in predictive accuracy and its potential for real-world clinical application represents a meaningful contribution to the field.

Materials and methods

Data collection and data processing

This study utilized secondary data on HTN collected from Paulose et al., (2022), where they investigated the prevalence and associated factors of HTN among the population of Hawassa City in Ethiopia [30]. The original study was a cross-sectional study carried out in the community by the Hawassa city administration. Participants had to be at least 30 years old and have resided in the city for a minimum of six months to be included. A total of 612 samples were collected using a multi-stage sampling technique. It was selected to represent diverse demographic and clinical characteristics associated with HTN risk in the Ethiopian population. Data collection involved the use of a structured questionnaire that captured demographic and socio-economic variables, as well as information on blood pressure history, co-morbidities, behavioral factors, and physical measurements. Blood pressure measurements were taken for all 612 participants to confirm the presence or absence of HTN. HTN diagnosis followed the WHO standard (systolic pressure at least 140/90 mmHg and/or diastolic pressure at least 90 mmHg) and was conducted by trained nurses. Individual risk factors for hypertension were identified based on different levels of explanatory variables, and the quantitative variables were classified according to prior sessions (see S1 Table) [28, 31–33]. The collected dataset didn't contain any missing values or outliers, as these issues were addressed during the original data collection process. Therefore, no additional steps for data cleaning, outlier detection, or imputation were required. The completeness and quality of the dataset ensured that the data was ready for analysis, enhancing the reliability of our results.

Feature selection

Feature selection techniques are essential in ML, as they allow for the extraction of the most relevant attributes for classification [34]. This not only improves the performance of the model during training but also facilitates easier interpretation of the model's outcomes [35]. To determine the most critical subset of features, we applied four feature selection algorithms: Borutabased feature selection (BFS) method, Least Absolute Shrinkage and Selection Operator (LASSO) regression, Forward and Backward Selection (FBS) and random forest (RF). Boruta is a wrapper-based feature selection method that utilizes the RF classifier algorithm, renowned for its unbiased and robust performance [36]. The LASSO algorithm introduces *L1* regularization to the regression model, penalizing the number of features to prevent overfitting [37]. Random Forest entails constructing multiple decision trees based on bootstrap samples [38]. Stepwise methods iteratively enhance the selected variables by including or excluding one variable at a time; examples include forward selection and backward selection algorithms [39]. After applying each feature selection approach, we intersected the results of all four methods to identify the most significant risk factors associated with hypertension (HTN).

Common features =
$$\bigcap_{i=1}^{r}$$
 Identified features form HTN dataset_i

Where, r is the number of utilizing feature selection methods (here, r = 4).

Machine learning algorithms

The study employed six machine learning algorithms to achieve its objectives, each chosen for their specific strengths in handling prediction tasks (see <u>S1 Appendix</u> for details). Logistic Regression (LR), a widely used supervised algorithm, was selected for its ability to predict the probability of binary outcomes, which aligns with the study's goal of distinguishing between

success and non-success outcomes [40, 41]. RF were chosen for their ability to improve decision trees by using bagging to build multiple trees, addressing overfitting through averaging predictions [40, 42-44]. ANN simulates the human brain's reasoning and pattern recognition, offering robust methods for identifying complex relationships [45, 46]. XGBoost, a gradientboosting algorithm, was included for its efficiency in predicting residuals from previous models, contributing to the overall predictive power [47, 48]. Similarly, LGBM, known for its speed and accuracy, was selected for its histogram-based approach and leaf-wise strategy, making it effective in large-scale data processing [49-51]. The stacking technique was employed as a meta-learning approach, where predictions from these diverse base models were used to train a new meta-learner, enhancing the overall performance [52]. The stacking classifier consisted of LR, ANN, RF, LGBM and XGBoost as base classifiers (level 0), with LR as the meta-learner (level 1). LR was chosen as the meta-learner due to its simplicity and interpretability, providing clear explanations of how base models contribute to final prediction [53]. This approach also mitigates overfitting in complex ensembles. While LR is effective in avoiding overfitting, its assumption of linearity may limit flexibility in datasets with non-linear relationships. In such cases, more complex meta-models like Gradient Boosting or RF could capture these intricate patterns more effectively, though they may introduce a higher risk of overfitting if the base models are already complex. The combination of these algorithms ensured that each model contributed meaningfully, while weaknesses in one algorithm were compensated for by the strengths of others, improving the overall accuracy and robustness of the predictions.

Data partition and balancing

The entire dataset was randomly split into two sets: 70% was used for training (HTN: 21.2%, non-HTN: 78.8%) and 30% was used for testing (HTN: 21.3%, non-HTN: 78.7%) and employing a technique for stratified sampling. Due to the class imbalance in the data, the majority class of the target variable may lead to biased results in classification tasks. To address this issue; several data balancing strategies can be employed. In our study, we used the Adaptive synthetic (ADASYN) balancing approach alongside under-sampling to balance the training set. Although our sample size may appear limited for machine learning applications, the ADA-SYN technique helped mitigate class imbalance and reduce potential overfitting. A larger dataset would improve the robustness and generalizability of the findings, and plans are in place to expand the dataset in future research or collaborate with larger cohorts to enhance the credibility and external applicability of the results.

Cross-validation and tuning hyperparameters

There are additional parameters, referred to as hyperparameters, in the ML algorithms discussed above. To increase the model's performance, the user can explicitly define hyperparameters before the learning process. The hyperparameter values in the training set were adjusted by the grid search technique using a repeating 10-fold (K10) cross-validation procedure. To carry out the K10 technique, a training subset and a verification set are separated from the training dataset in 7:3 ratios.

Kernel SHAP-based interpretability method

Kernel SHAP is a method that computes SHapley values using a specialized weighted linear regression function to estimate each feature's contribution [54]. SHapley values account for the various magnitudes and signs with which risk factors influence the model's result or prediction. As a result, SHapley values represent estimations of the contribution's feature important amount and direction (sign). In the model, risk factors with a positive SHAP value help

predict patients with HTN, whereas those with a negative SHAP value assist in forecasting patients under control [28]. Specifically, the significance of every risk factor, let's say the $k^{\rm th}$ risk factor, is determined by the SHapley value, which is determined by the following formula

$$\emptyset_k(\nu) = \frac{1}{M!} \sum_{S \subseteq M \setminus \{k\}} |S|! (M - |S| - 1)! [\nu(S \cup \{k\}) - \nu(S)]$$

where, S stands for the subset of risk factors that excludes the risk factor for which the value is being calculated; $\emptyset_k(v)$; $S \cup \{k\}$ is the subset of risk factors that includes the k^{th} risk factor in S; v(S) is the result of the ML-based model that explains using the risk factors of S; $S \subseteq M \setminus \{k\}$ and represents every set of S that is a subset of every risk factor set in all of M, which excludes the k^{th} risk factor.

Performance evaluation criteria

We assessed each method's prediction performance in terms of precision, sensitivity, and specificity. The following is a description of the equations:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

where true positives, true negatives, false positives, and false negatives are denoted, respectively, by TP, TN, FP, and FN. Additionally, metrics such as the area under the curve (AUC) and the receiver operating characteristic (ROC) curve are also evaluated. The calculation formula of AUC is as follows:

$$AUC = \int_{x=0}^{1} \text{TPR}(FPR^{-1}(x)) dx$$

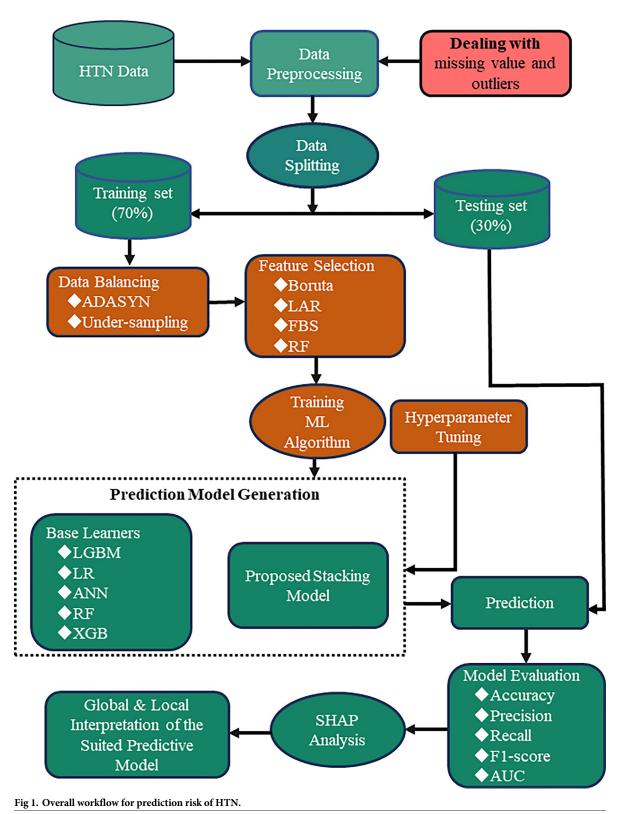
Ethical approval

To address the ethical considerations in this study, it is important to note that the data were sourced from a prior research project that received ethical approval from the Research and Ethics Committee of the University of South Africa (UNISA) (Reference number: REC-012714-039). The secondary analysis was conducted in compliance with ethical guidelines regarding the use of health-related data. As the original dataset was anonymized, no additional ethical approval was required for this analysis. We acknowledge the sensitivity of health-related information and have implemented measures to ensure participant confidentiality and data security throughout the study. Fig 1 illustrates the complete workflow of this study.

Results

Baseline characteristics

The study included 612 participants, of whom 130 (21.2%) had HTN and 482 (78.8%) did not. Among the participants, 53.4% were male, and over half resided in urban areas. The mean age of the cohort was 47.56 ± 13.40 years, with an average height and weight of 165.20 ± 8.87 cm and 66.589 ± 8.769 kg, respectively. HTN prevalence was notably higher among obese individuals compared to those with normal weight (50% vs. 13.4%). Additionally, individuals with



https://doi.org/10.1371/journal.pone.0315865.g001

| Table 1. Names, descriptions, and | d categorizations of the 13 commor | ı features selected usin | ng four feature selection methods. |
|-----------------------------------|------------------------------------|--------------------------|------------------------------------|
| | | | |

| S/N | Name | Description | Categorization |
|-----|-----------|-----------------------------|--|
| 1. | Age | Age of the respondents | Continuous variable (yearly) |
| 2. | Education | Educational level | Can't read and write, Read and write only, Primary, Secondary, Diploma and above |
| 3. | Walking | Walking at least 10 minutes | Yes, No |
| 4. | Weight | Weight of the respondents | Continuous variable (kg) |
| 5. | BMI | Body mass status | Underweight, Normal, Overweight, Obese |
| 6. | Smoking | Smoking status | Yes, No |
| 7. | Drinking | Drinking alcohol | Yes, No |
| 8. | Vegetable | Eat fruit at least per week | Yes, No |
| 9. | Fat | Eating animal fat | Yes, No |
| 10. | Salt | Eating habit salt | Yes, No |
| 11. | Transport | Mode of transport | On foot/pedal bicycle, Engine |
| 12. | HD | History of diabetes | Yes, No |
| 13. | HHTN | History of hypertension | Yes, No |

https://doi.org/10.1371/journal.pone.0315865.t001

diabetes (47.5% vs. 30.0%) and smokers (50.4% vs. 23.8%) exhibited higher prevalence rates of HTN. Having a family history of diabetes (41.8% vs. 11.2%) was linked to higher rates of HTN prevalence. Significant associations (P-value < 0.005) were observed between HTN and various factors including age, sex, residence, occupation, income, physical activity, diabetes, weight, height, BMI, smoking, alcohol consumption, dietary habits, transportation mode, and socioeconomic status (see \$1 Table).

Risk factors identification of HTN using different method

Feature selection involves identifying and removing inappropriate, irrelevant, or unnecessary features from a dataset to improve model accuracy. This study utilized four feature selection algorithms: BFS method, LASSO regression, FBS and RF, to identify important features that were common among these methods (see Table 1). The Boruta algorithm identified 18 features, while the forward-backward selection method identified 13, RF identified 19, and Lasso Regression identified 18. Each method revealed specific and important features associated with HTN. Thirteen features were found to be common across all four feature selection methods and were considered important risk factors for HTN (see Table 1 and S2 Table). These identified features are associated with hypertension, though their contribution to its development cannot be established based on the current data. The identified 13 risk factors were considered in the machine learning-based model for predicting HTN status.

Performance comparison of ML-based models

The performance of various machine learning models using imbalanced data, under-sampling and ADASYN for predicting hypertension is shown in Table 2 and S3 Table. The ADASYN technique significantly contributed to the improved performance metrics by addressing the class imbalance between hypertensive and non-hypertensive participants. By generating synthetic samples for the minority class (HTN), ADASYN enabled the model to better learn the patterns associated with hypertension, thereby reducing bias towards the majority class and enhancing predictive accuracy. The models' performance was evaluated using various metrics, including accuracy, precision, recall, F1-score, and AUC. The results indicated that the prediction performance of the ML models was comparatively low for imbalanced data (see S3 Table). However, for under-sampling and ADASYN data, prediction performance was comparatively better (see Table 2). The most accurate prediction was made by the LGBM model

| Balancing Method | Models | Accuracy | Precision | Recall | F1-Score | AUC |
|-------------------------|-------------------|----------|-----------|--------|----------|-------|
| Under Sampling | LGBM | 87.43 | 88.14 | 92.75 | 78.6 | 0.878 |
| | LR | 83.82 | 86.47 | 89.65 | 82.53 | 0.844 |
| | ANN | 86.34 | 81.54 | 93.06 | 84.08 | 0.879 |
| | RF | 82.51 | 86.84 | 94.48 | 53.01 | 0.777 |
| | XGB | 84.69 | 88.30 | 84.41 | 69.56 | 0.785 |
| ADASYN | LGBM | 92.61 | 91.24 | 94.12 | 92.46 | 0.943 |
| | LR | 86.53 | 83.01 | 90.07 | 86.39 | 0.912 |
| | ANN | 88.55 | 84.97 | 92.36 | 88.51 | 0.942 |
| | RF | 88.55 | 83.89 | 93.24 | 88.34 | 0.937 |
| | XGB | 90.61 | 92.05 | 90.25 | 91.14 | 0.933 |
| | Proposed Stacking | 96.32 | 95.48 | 97.51 | 96.48 | 0.971 |

Table 2. Performance of the five ML models and their stacking models with two-class balancing.

https://doi.org/10.1371/journal.pone.0315865.t002

using the ADASYN balancing technique with an accuracy of 92.61%, precision of 91.24%, recall of 94.12%, F1-score of 92.46% and AUC of 0.943 compared to other models. Additionally, to increase the prediction accuracy of HTN, we used stacking ensemble learning models. The results indicated that the best predictive discrimination ability was attained by our suggested stacking model with an accuracy of 96.32%, precision of 95.48%, recall of 97.51%, F1-score of 96.48%, and AUC of 0.971 compare to other five ML-models (see Table 2).

The ROC curves of the five predictive models and the proposed stacking model with ADA-SYN are shown in Fig 2. The ROC values also indicated that our proposed stacking model is significantly better than the LGBM, LR, ANN, RF and XGB models. Therefore, based on the prediction performance results, it is demonstrated that our proposed stacking model with ADASYN performed better.

Interpretable hypertension risk factors

Using SHAP values, an in-depth analysis was conducted to identify interpretable predictive risk factors for HTN within the proposed stacking model. The SHAP summary plot provides a detailed view of how the input risk factors influence the predictions. The significance, impact, initial value, and connection of the risk factors to the high risk of HTN are illustrated by the swarm plot in Fig 3. The x-axis displays the direction and magnitude of the influence of each risk factor, with positive values suggesting an increased risk of HTN and negative values indicating a decreased risk. The color gradient represents the value of each risk factor, with red denoting high values and blue indicating low values. The analysis revealed that several factors are significantly associated with HTN risk such as weight, alcohol consumption, a history of HTN, salt intake, age, diabetes, BMI, and fat intake. High weight, excessive alcohol consumption, and a prior diagnosis of hypertension correlate strongly with increased HTN risk, while high salt intake and advancing age further elevate this risk. Additionally, a history of diabetes and elevated BMI contribute to susceptibility, consistent with established clinical knowledge. These findings underscore the importance of these risk factors in predicting HTN, suggesting targeted intervention strategies that could focus on weight management, dietary modifications, and lifestyle changes to mitigate the risk of developing HTN.

Discussion

Hypertension prediction is vital for effective preventive healthcare strategies [55]. Assessing current HTN risk is essential for identifying underlying risk factors that may not yet be

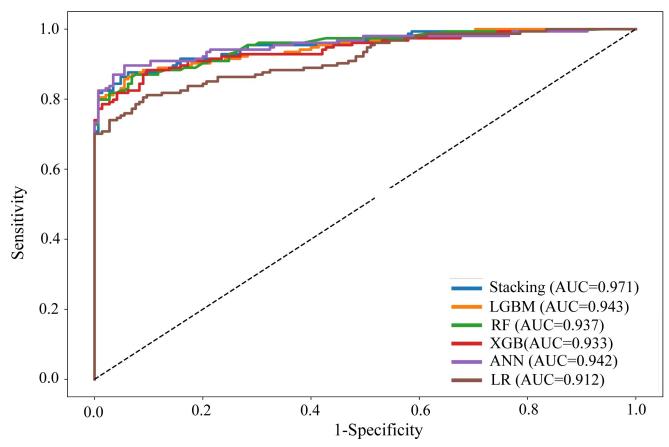


Fig 2. ROC curves of the six predictive models.

https://doi.org/10.1371/journal.pone.0315865.g002

reflected in elevated blood pressure. While predicting future risk aids in early detection and intervention, elevating current risk helps healthcare providers identify individuals at risk of developing HTN. This proactive approach enables timely lifestyle modifications and interventions to prevent its progression. Integrating both current and future risk assessments enhances HTN management strategies, addressing immediate concerns while mitigating long-term risk. This study aimed to improve the prediction of HTN using ensemble learning approaches. By integrating multiple ML algorithms, we sought to improve the accuracy of predictive models compared to traditional single-model methods. Our research utilized a comprehensive dataset from Ethiopia, consisting of 612 participants with 27 features potentially associated with HTN risk. The dataset underwent various balancing techniques, including under-sampling and ADASYN, to address class imbalance issues. We employed a comprehensive feature selection approach that integrated BFS, LASSO regression, FBS, and RF feature importance to explore the most relevant predictors.

Using 13 risk factors identified through this technique, we trained five ML algorithms (ANN, LR, RF, XGB, and LGBM) as well as a stacking model to predict HTN. The performance of these models was evaluated on the testing set using metrics such as AUC value, accuracy, precision, recall, and F1-score. The accuracy of the five ML-based models with ADASYN are as follows: 92.61% for LGBM, 86.53% for LR, 88.55% for both ANN and RF, and 90.61% for XGB. Our proposed stacking model achieved an accuracy of 96.32%. According to these performance metrics, we recommend the stacking model as the optimal classifier for predicting HTN. A previous study on the same dataset showed prediction accuracies of 86.43% for

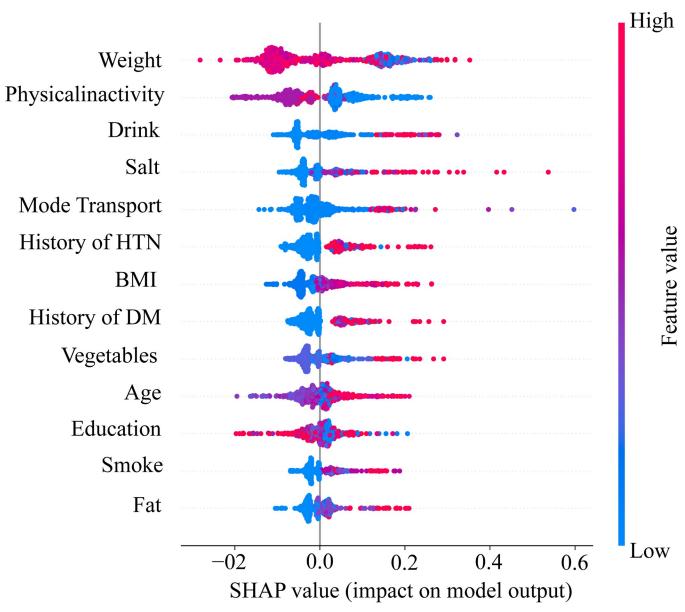


Fig 3. SHAP values: Impact of risk factors on prediction outcome.

https://doi.org/10.1371/journal.pone.0315865.g003

LR, 85.25% for ANN, 87.88% for RF, and 88.81% for XGB [28]. The XGB ML-model achieved highest accuracy (88.81%) in this study which is almost similar with our result. Another study showed that LGBM is more robust than other ML models [29], achieving the highest accuracy (92.61%) for our HTN data. To further increase prediction accuracy, we used the stacking model, which achieved an accuracy of 96.32%, outperforming the existing single-model prediction accuracies [28]. Additionally, a study on HTN risk detection selected features using FI-FA and MC-FA models, with MC-FA (10 Factor) and algorithms like RF, KNN, DT, and NB showing promising results, particularly RF with an AUC of 85.96% and accuracy of 80.12% [25]. In subsequent study, the LR approach was used to identify significant HTN risk factors, alongside RF, CatBoost, LR and MLP Neural Network models. Among these, RF achieved an accuracy of 82% and an AUC of 0.92. [26]. Another study found that LR

Table 3. Comparison of prediction performance of exiting models for hypertension risk prediction.

| Reference | Country | Study type | Data size | No. of riskfactors | Best model | AUC | SHAP |
|-----------------------------|-----------------------|------------|-----------|--------------------|------------|-------|------|
| Islam et al., 2023 [28] | Ethiopia | CS | 612 | 27 | XGB | 0.89 | yes |
| Zhao et al., 2021 [26] | China | CS | 29,700 | 11 | RF | 0.92 | No |
| Kurniawan et al., 2023 [56] | Indonesia | CS | 30,320 | 11 | LR | 0.829 | No |
| Wu Y et al., 2024 [57] | China | CS | 476 | 9 | RF | 0.95 | No |
| Chai et al., 2022 [58] | Malaysia | CS | 2461 | 11 | LGBM | 0.686 | No |
| Islam et al., 2021 [59] | Bangladesh | CS | 6965 | 13 | GB | 0.669 | No |
| AlKaabi et al., 2020[60] | Qatar | CS | 987 | 12 | RF | 0.869 | No |
| Islam et al., 2022 [61] | South Asian countries | CS | 8,18,603 | 7 | XGBoost | 0.900 | No |
| Proposed ensemble model | Ethiopia | CS | 612 | 27 | Stacking | 0.971 | Yes |

^{*} CS: cross-sectional study

https://doi.org/10.1371/journal.pone.0315865.t003

performed well in predicting HTN risk, achieving an AUC of 0.829 in an Indonesian study, indicating its potential for HTN risk assessment [56]. Furthermore, a study also employed a multi-pronged strategy for building a HTN prediction system in Malaysia, using feature selection and addressing class imbalance, achieving 74.39% accuracy with their LGBM-based model [58]. So, based on the previous discussion our proposed stacking model achieved the highest HTN prediction accuracy comparatively the existing studies (see Table 3). The existing study employed ML models to predict HTN risk without using clinical or genetic data in a cross-sectional study. In addition to comparing our machine learning model with other machine learning methods, it is important to consider its performance relative to classical risk prediction tools currently utilized in clinical practice, such as the Framingham Hypertension Risk Calculator [62–64]. The Framingham tools has been extensively validated in large cohorts and remains a reliable choice for long-term risk prediction, particularly for identifying individuals at risk up to four years in advance. However, it primarily relies on specific demographic and clinical variables and may lack generalizability to ethnically diverse populations, as it was developed in cohorts predominantly of European descent [62, 63]. Additionally, other HTN risk calculators, such as the Strong Heart Study, which is specifically tailored for Native-American populations, underscore the need for population-specific risk assessment tools [65]. This study focuses on risk factors prevalent in Native Americans, providing insights into hypertension prediction in a group often underrepresented in clinical studies [66]. In contrast, our MLbased model leverages advanced algorithms that integrate diverse data sources, allowing it to analyze a wider range of risk factors and capture complex interactions between them, potentially enhancing predictive accuracy. While our cross-sectional approach limits its ability to predict future hypertension onset, it identifies current risk profiles, enabling timely interventions for individuals who may otherwise be missed without a recent blood pressure check. Together, these tools highlight the importance of combining traditional clinical approaches, such as the Strong Heart Study, with advanced data-driven methods like ours, to improve hypertension prediction across diverse populations.

The SHAP analysis of the stacking model identified weight, alcohol consumption (drink), history of HTN, salt intake, age, diabetes, vegetables, BMI and fat as the key interpretable risk factors for HTN. Among these, weight and obesity (fat) were critical determinants, consistent with earlier studies showing that obese individuals are at a significantly higher risk of developing HTN compared to those with normal weight [67, 68]. Elevated BMI also emerged as a strong predictor, aligning with research linking BMI to HTN and cardiovascular diseases through mechanisms like renin-angiotensin system activation and endothelial dysfunction

[69, 70]. Dietary factors such as high salt intake and low vegetable consumption, along with alcohol consumption, were identified as significant lifestyle contributors, echoing previous finding [71, 72]. Individuals who consumed alcohol daily or after meals demonstrated a markedly higher risk of HTN compared to abstainers [73]. Additionally, age was a dominant risk factor, as older individuals (60+ years) were more likely to develop HTN compared to younger adults (18-40 years), supported by findings from Belay et al. (2022) and other systematic reviews [16, 74, 75]. This age-related risk is attributed to vascular changes, such as large artery stiffness, that occur with aging. History of hypertension (HHTN) further emerged as a significant variable, consistent with studies linking familial predisposition, shared lifestyle behaviors, and genetic factors to HTN risk [75]. Diabetes, with its bidirectional relationship to HTN, also plays a crucial role, as the two conditions share overlapping risk factors and can exacerbate each other [76]. The analysis highlights the compounded risk in obese hypertensive individuals, who have higher rates of coronary heart disease and mortality than those with either condition alone [68, 77]. Genetic predispositions, such as a higher Genetic Risk Score (GRS), were also significantly associated with increased odds of HTN, as shown in previous studies [68]. Furthermore, the relationship between salt intake and blood pressure, long established since Kempner's 1948 rice diet study, underscores the value of dietary modifications like salt restriction to mitigate HTN risk [78].

Limitations

The dataset used in this study presents certain limitations that could affect the generalizability and integrity of the findings. With only 612 instances, the relatively small sample size may limit the robustness of the machine learning model's predictions. Additionally, there is a notable class imbalance, with 21% of the cases being HTN and 79% non-hypertensive, which could introduce bias in the results despite employing data balancing techniques like ADASYN. This imbalance may still influence model accuracy, potentially leading to overfitting. While the current study offers valuable insights, further research with larger and more balanced datasets is necessary to enhance the external validity and real-world applicability of the model's predictions. Future work could involve collaboration with larger cohorts or utilizing additional secondary data sources to mitigate these limitations.

Conclusions

This study compared five ML algorithms and a stacking ensemble model for predicting HTN risk. The stacking model emerged as the most effective approach for identifying patients at risk of HTN. Furthermore, SHAP analysis of the stacking model revealed that physical inactivity, age, educational status, and a history of hypertension were the most significant contributing factors to HTN development. These findings suggest that the proposed integrated system can be a valuable tool in clinical settings for the early detection of patients at risk of HTN. By leveraging this information, healthcare professionals can make informed decisions that has the potential to decrease healthcare costs and improve patient outcomes. Additionally, early identification allows for the implementation of personalized interventions and targeted treatment strategies, ultimately contributing to a reduction in the overall burden of HTN in Ethiopia.

Supporting information

S1 Appendix. Machine learning algorithms. (DOCX)

S1 Table. Demographic profiles of the respondents.

(DOCX)

S2 Table. Risk factors for HTN identified using five feature selection techniques.

(DOCX)

S3 Table. Performance evaluation of prediction models on imbalanced data.

(DOCX)

S1 Text. Code for machine learning-based hypertension prediction.

(TXT)

Acknowledgments

Authors would like to acknowledge both reviewers for their valuable comments that were helpful to improve the quality of the manuscript.

Author Contributions

Conceptualization: Isteaq Kabir Sifat, Md. Kaderi Kibria.

Data curation: Isteaq Kabir Sifat, Md. Kaderi Kibria.

Formal analysis: Isteaq Kabir Sifat.

Funding acquisition: Md. Kaderi Kibria.

Investigation: Md. Kaderi Kibria. **Methodology:** Isteaq Kabir Sifat.

Project administration: Md. Kaderi Kibria.

Resources: Md. Kaderi Kibria.
Software: Isteaq Kabir Sifat.
Supervision: Md. Kaderi Kibria.

Validation: Md. Kaderi Kibria.
Visualization: Isteaq Kabir Sifat.

Writing - original draft: Isteaq Kabir Sifat, Md. Kaderi Kibria.

Writing - review & editing: Md. Kaderi Kibria.

References

- Katherine T., Mills A.S.& J.H. The Global Epidemiology of Hypertension _ Nature Reviews Nephrology. Nat. Rev. Nephrol. 2020, 16, 223–237, https://doi.org/10.1038/s41581-019-0244-2
- 2. Https://Www.Who.Int/News-Room/Fact-Sheets/Detail/Hypertension (WHO).
- Kalantari S.; Khalili D.; Asgari S.; Fahimfar N.; Hadaegh F.; Tohidi M.; et al. Predictors of Early Adult-hood Hypertension during Adolescence: A Population-Based Cohort Study. BMC Public Health 2017, 17, https://doi.org/10.1186/s12889-017-4922-3 PMID: 29183297
- KCDC Korea Centers for Disease Control and Prevention. Press Release. [Internet], http://Knhanes.cdc.Go.Kr. 2020.
- Lawes C.M.M.; Hoorn S. Vander; Rodgers A. Global Burden of Blood-Pressure-Related Disease, 2001; 2008; Vol. 371, https://doi.org/10.1016/S0140-6736(08)60655-8 PMID: 18456100
- Kearney P.M.; Whelton M.; Reynolds K.; Muntner P.; Whelton P.K.; He J. Global Burden of Hypertension: Analysis of Worldwide Data; 2005; Vol. 365; ISBN 1992932069, https://doi.org/10.1016/S0140-6736(05)17741-1 PMID: 15652604

- Forouzanfar M.H.; Afshin A.; Alexander L.T.; Biryukov S.; Brauer M.; Cercy K.; et al. Global, Regional, and National Comparative Risk Assessment of 79 Behavioural, Environmental and Occupational, and Metabolic Risks or Clusters of Risks, 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015. Lancet 2016, 388, 1659–1724, https://doi.org/10.1016/S0140-6736(16)31679-8 PMID: 27733284
- Lloyd-Jones D.; Adams R.J.; Brown T.M.; Carnethon M.; Dai S.; De Simone G.; et al. Executive Summary: Heart Disease and Stroke Statistics-2010 Update: A Report from the American Heart Association. Circulation 2010, 121, https://doi.org/10.1161/CIRCULATIONAHA.109.192666 PMID: 20177011
- Whelton P.K.; Carey R.M.; Aronow W.S.; Casey D.E.; Collins K.J.; Dennison Himmelfarb C.; et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Pr. Circulation 2018, 138, e484–e594. https://doi.org/10.1161/CIR.00000000000000596 PMID: 30354654
- CDC. Hypertension Prevalence in the U.S. | Million Hearts®. In: Centers for Disease Control and Prevention [Internet]. 22 Mar 2021. Available: Https://Millionhearts.Hhs.Gov/Data-Reports/Hypertensionprevalence.Html.
- Mroz T.; Griffin M.; Cartabuke R.; Laffin L.; Russo-Alvarez G.; Thomas G.; et al. Predicting Hypertension Control Using Machine Learning. PLoS One 2024, 19, https://doi.org/10.1371/journal.pone.0299932 PMID: 38507433
- Rocha E. Fifty Years of Framingham Study Contributions to Understanding Hypertension. Rev. Port. Cardiol. 2001, 20, 795–796, https://doi.org/10.1038/sj.jhh.1000949 PMID: 11582630
- Mehta R.; Mantri N.; Goel A.; Gupta M.; Joshi N.; Bhardwaj P. Out-of-Pocket Spending on Hypertension and Diabetes among Patients Reporting in a Health -Care Teaching Institute of the Western Rajasthan. J. Fam. Med. Prim. Care 2022, 11, 1083, https://doi.org/10.4103/jfmpc.jfmpc_998_21 PMID: 35495832
- Sorato M.M.; Davari M.; Kebriaeezadeh A.; Sarrafzadegan N.; Shibru T. Societal Economic Burden of Hypertension at Selected Hospitals in Southern Ethiopia: A Patient-Level Analysis. *BMJ Open* 2022, 12, https://doi.org/10.1136/bmjopen-2021-056627 PMID: 35387822
- Berek P.; Irawati D.; Hamid A. Hypertension_ A Global Health Crisis. Ann Clin Hypertens. 21AD, 5, 8– 11, https://doi.org/10.29328/journal.ach.1001027
- 16. Belay D.G.; Fekadu Wolde H.; Molla M.D.; Aragie H.; Adugna D.G.; Melese E.B.; et al. Prevalence and Associated Factors of Hypertension among Adult Patients Attending the Outpatient Department at the Primary Hospitals of Wolkait Tegedie Zone, Northwest Ethiopia. Front. Neurol. 2022, 13, https://doi.org/10.3389/fneur.2022.943595 PMID: 36034276
- Chowdhury M.Z.I.; Turin T.C. Precision Health through Prediction Modelling: Factors to Consider before Implementing a Prediction Model in Clinical Practice. *J. Prim. Health Care* 2020, 12, 3–9, https://doi.org/10.1071/HC19087 PMID: 32223844
- Usher-Smith J.A.; Silarova B.; Schuit E.; Moons K.G.M.; Griffin S.J. Impact of Provision of Cardiovascular Disease Risk Estimates to Healthcare Professionals and Patients: A Systematic Review. *BMJ Open* 2015, 5, https://doi.org/10.1136/bmjopen-2015-008717 PMID: 26503388
- Lopez-Gonzalez A.A.; Aguilo A.; Frontera M.; Bennasar-Veny M.; Campos I.; Vicente-Herrero T.; et al. Effectiveness of the Heart Age Tool for Improving Modifiable Cardiovascular Risk Factors in a Southern European Population: A Randomized Trial. Eur. J. Prev. Cardiol. 2015, 22, 389–396, https://doi.org/10. 1177/2047487313518479 PMID: 24491403
- Chowdhury M.Z.I.; Naeem I.; Quan H.; Leung A.A.; Sikdar K.C.; O'Beirne M.; et al. Summarising and Synthesising Regression Coefficients through Systematic Review and Meta-Analysis for Improving Hypertension Prediction Using Metamodelling: Protocol. BMJ Open 2020, 10, https://doi.org/10.1136/ bmjopen-2019-036388 PMID: 32276958
- Chowdhury M.Z.I.; Yeasmin F.; Rabi D.M.; Ronksley P.E.; Turin T.C. Prognostic Tools for Cardiovascular Disease in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis of C-Statistics. *J. Diabetes Complications* 2019, 33, 98–111, https://doi.org/10.1016/j.jdiacomp.2018.10.010 PMID: 30446478
- Chowdhury M.Z.I.; Naeem I.; Quan H.; Leung A.A.; Sikdar K.C.; OBeirne M.; et al. Prediction of Hypertension Using Traditional Regression and Machine Learning Models: A Systematic Review and Meta-Analysis. PLoS One 2022, 17, https://doi.org/10.1371/journal.pone.0266334 PMID: 35390039
- Chowdhury M.Z.I.; Leung A.A.; Sikdar K.C.; O'Beirne M.; Quan H.; Turin T.C. Development and Validation of a Hypertension Risk Prediction Model and Construction of a Risk Score in a Canadian Population. Sci. Rep. 2022, 12, https://doi.org/10.1038/s41598-022-16904-x PMID: 35896590

- 24. Ghosh S.; Kumar M. Prevalence and Associated Risk Factors of Hypertension among Persons Aged 15–49 in India: A Cross-Sectional Study. *BMJ Open* 2019, 9, https://doi.org/10.1136/bmjopen-2019-029714 PMID: 31848161
- Khongorzul D.; Kim M. Comparison of Feature Selection Methods Applied on Risk Prediction for Hypertension. KIPS Transactions on Software and Data Engineering, 11 (3), 107–114, https://doi.org/10.3745/KTSDE.2022.11.3.107
- Zhao H.; Zhang X.; Xu Y.; Gao L.; Ma Z.; Sun Y.; et al. Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method. Front. Public Heal. 2021, 9, https://doi.org/10.3389/fpubh.2021.619429 PMID: 34631636
- Chang W.; Liu Y.; Xiao Y.; Yuan X.; Xu X.; Zhang S.; et al. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. *Diagnostics* 2019, 9, https://doi.org/10. 3390/diagnostics9040178 PMID: 31703364
- Islam M.M.; Alam M.J.; Maniruzzaman M.; Ahmed N.A.M.F.; Ali M.S.; Rahman M.J.; et al. Predicting the Risk of Hypertension Using Machine Learning Algorithms: A Cross Sectional Study in Ethiopia. PLoS One 2023, 18, https://doi.org/10.1371/journal.pone.0289613 PMID: 37616271
- Rufo D.D.; Debelee T.G.; Ibenthal A.; Negera W.G. Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (Lightgbm). *Diagnostics* 2021, 11, https://doi.org/10.3390/diagnostics11091714 PMID: 34574055
- Paulose T.; Nkosi Z.Z.; Endriyas M. Prevalence of Hypertension and Its Associated Factors in Hawassa City Administration, Southern Ethiopia: Community Based Crosssectional Study. PLoS One 2022, 17, https://doi.org/10.1371/journal.pone.0264679 PMID: 35231073
- Mika M.; Kenneth S.; Orvalho A.; Yoshito K.; Kristjana Á.; Falume C.; et al. The Prevalence of Hypertension and Its Distribution by Sociodemographic Factors in Central Mozambique: A Cross Sectional Study. BMC Public Health 2020, 20, https://doi.org/10.1186/s12889-020-09947-0 PMID: 33261617
- Sharma JR; Mabhida SE; Myers B; Apalata T; Nicol E; Benjeddou M; et al Prevalence of Hypertension and Its Associated Risk Factors in aRural Black Population of Mthatha Town, South Africa. *Int. J. Envi*ron. Res. Public Health 2021, 10, https://doi.org/10.3390/ijerph18031215
- 33. Manios Y.; Androutsos O.; Lambrinou C.P.; Cardon G.; Lindstrom J.; Annemans L.; et al. A School- and Community-Based Intervention to Promote Healthy Lifestyle and Prevent Type 2 Diabetes in Vulnerable Families across Europe: Design and Implementation of the Feel4Diabetes-Study. Public Health Nutr. 2018, 21, 3281–3290, https://doi.org/10.1017/S1368980018002136 PMID: 30207513
- Ghosh P.; Azam S.; Jonkman M.; Karim A.; Shamrat F.M.J.M.; Ignatious E.; et al. Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and Lasso Feature Selection Techniques. IEEE Access 2021, 9, 19304–19326, https://doi.org/10.1109/ACCESS.2021.3053759
- Liu X.; Zhang Y.; Fu C.; Zhang R.; Zhou F. EnRank: An Ensemble Method to Detect Pulmonary Hypertension Biomarkers Based on Feature Selection and Machine Learning Models. Front. Genet. 2021, 12, https://doi.org/10.3389/fgene.2021.636429 PMID: 33986767
- Pudjihartono N.; Fadason T.; Kempa-Liehr A.W.; O'Sullivan J.M. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. Front. Bioinforma. 2022, 2, https://doi.org/10.3389/fbinf.2022.927312 PMID: 36304293
- Deshpande S.; Shuttleworth J.; Yang J.; Taramonli S.; England M. PLIT: An Alignment-Free Computational Tool for Identification of Long Non-Coding RNAs in Plant Transcriptomic Datasets. *Comput. Biol. Med.* 2019, 105, 169–181, https://doi.org/10.1016/j.compbiomed.2018.12.014 PMID: 30665012
- Gharsalli S.; Emile B.; Laurent H.; Desquesnes X. Feature Selection for Emotion Recognition Based on Random Forest.; Scitepress, April 2016; pp. 610–617, https://doi.org/10.5220/0005725206100617
- 39. Borboudakis G. Forward-Backward Selection with Early Dropping; 2019; Vol. 20;.
- Kantharaju V; R. Pavithra; Nisarga H; Karishma S Prediction of Chronic Kidney Disease-A Machine Learning Perspective. Int. J. Sci. Res. Sci. Eng. Technol. 2022, 37–43, https://doi.org/10.32628/ iisrset22924
- Priya Ranganathan C. S. Pramesh; Rakesh Aggarwal Common Pitfalls in Statistical Analysis: Logistic Regression. Perspect. Clin. Res. 2017, 148–151, https://doi.org/10.4103/picr.PICR_87_17 PMID: 28828311
- Bauer E.; Kohavi R. Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. Mach. Learn. 1999, 36, 105–139, https://doi.org/10.1023/a:1007515423169
- **43.** Random Decision Forests. *Encycl. Mach. Learn. Data Min.* 2017, 1054–1054, https://doi.org/10.1007/978-1-4899-7687-1_100391
- Liaw A.; Prasad A.M.; Iverson L.R. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 2006, 9, 181–199, https://doi.org/10.1007/s10021-005-0054-1.

- Kulkarni P.S.; Londhe S.N.; M.C.Deo Artificial Neural Networks Management: A Review for Construction. J. Soft Comput. Civ. Eng. 2017, 1, 70–88, https://doi.org/10.22115/scce.2017.49580
- Andrew A.M. (1999), "The Handbook of Brain Theory and Neural Networks", Kybernetes, Vol. 28 No. 9, pp. 1084 1094. https://doi.org/10.1108/k.1999.28.9.1084.1
- Chen T.; Guestrin C. XGBoost: A Scalable Tree Boosting System. Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2016, 13-17-Augu, 785–794, https://doi.org/10.1145/2939672.2939785
- **48.** Nielsen D. Tree Boosting With XGBoost Why Does XGBoost Win "Every" Machine Learning Competition? *Glob. Policy* 2012, 3, 24–34.
- **49.** Ke G.; Meng Q.; Finley T.; Wang T.; Chen W.; Ma W.; et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* 2017, 2017-Decem, 3147–3155.
- Liang W.; Luo S.; Zhao G.; Wu H. Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics* 2020, 8, https://doi.org/10.3390/MATH8050765
- Basha S.M.; Rajput D.S.; Vandhan V. Impact of Gradient Ascent and Boosting Algorithm in Classification. Int. J. Intell. Eng. Syst. 2018, 11, 41–49, https://doi.org/10.22266/ijies2018.0228.05
- Dey R.; Mathur R. Ensemble Learning Method Using Stacking with Base Learner, A Comparison. Lect. Notes Networks Syst. 2023, 727 LNNS, 159–169, https://doi.org/10.1007/978-981-99-3878-0_14
- Barton M.; Lennox B. Model Stacking to Improve Prediction and Variable Importance Robustness for Soft Sensor Development. *Digit. Chem. Eng.* 2022, 3, https://doi.org/10.1016/j.dche.2022.100034
- Scott M. Lundberg; Su-In Lee A Unified Approach to Interpreting Model Predictions. Adv. Neural Inf. Process. Syst. 2017, 30, https://doi.org/10.48550/arXiv.1705.07874.
- 55. Ojurongbe T.A.; Afolabi H.A.; Oyekale A.; Bashiru K.A.; Ayelagbe O.; Ojurongbe O.; et al. Predictive Model for Early Detection of Type 2 Diabetes Using Patients' Clinical Symptoms, Demographic Features, and Knowledge of Diabetes. *Heal. Sci. Reports* 2024, 7, https://doi.org/10.1002/hsr2.1834 PMID: 38274131
- Kurniawan R.; Utomo B.; Siregar K.N.; Ramli K.; Besral Suhatril R.J; Pratiwi O.A. Hypertension Prediction Using Machine Learning Algorithm among Indonesian Adults. *IAES Int. J. Artif. Intell.* 2023, 12, 776–784, https://doi.org/10.11591/ijai.v12.i2.pp776-784
- 57. Wu Y.; Xin B.; Wan Q.; Ren Y.; Jiang W. Risk Factors and Prediction Models for Cardiovascular Complications of Hypertension in Older Adults with Machine Learning: A Cross-Sectional Study. *Heliyon* 2024, 10, https://doi.org/10.1016/j.heliyon.2024.e27941 PMID: 38509942
- Chai S.S.; Goh K.L.; Cheah W.L.; Chang Y.H.R.; Ng G.W. Hypertension Prediction in Adolescents
 Using Anthropometric Measurements: Do Machine Learning Models Perform Equally Well? *Appl. Sci.* 2022, 12, https://doi.org/10.3390/app12031600
- Islam M.M.; Rahman M.J.; Chandra Roy D.; Tawabunnahar M.; Jahan R.; Ahmed N.A.M.F.; et al. Machine Learning Algorithm for Characterizing Risks of Hypertension, at an Early Stage in Bangladesh. Diabetes Metab. Syndr. Clin. Res. Rev. 2021, 15, 877–884, https://doi.org/10.1016/j.dsx.2021.03.035
 PMID: 33892404
- AlKaabi L.A.; Ahmed L.S.; Al Attiyah M.F.; Abdel-Rahman M.E. Predicting Hypertension Using Machine Learning: Findings from Qatar Biobank Study. *PLoS One* 2020, 15, https://doi.org/10.1371/journal. pone.0240370 PMID: 33064740
- 61. Islam S.M.S.; Talukder A.; Awal M.A.; Siddiqui M.M.U.; Ahamad M.M.; 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, https://doi.org/10.3389/fcvm.2022.839379 PMID: 35433854
- 62. Kivimäki M.; Batty G.D.; Singh-Manoux A.; Ferrie J.E.; Tabak A.G.; Jokela M.; et al. Validating the Framingham Hypertension Risk Score: Results from the Whitehall II Study. *Hypertension* 2009, 54, 496–501, https://doi.org/10.1161/HYPERTENSIONAHA.109.132373 PMID: 19597041
- Parikh N.I.; Pencina M.J.; Wang T.J.; Benjamin E.J.; Lanier K.J.; Levy D.; et al. A Risk Score for Predicting Near-Term Incidence of Hypertension: The Framingham Heart Study. *Ann. Intern. Med.* 2008, 148, 102–110, https://doi.org/10.7326/0003-4819-148-2-200801150-00005 PMID: 18195335
- Bloch M.J.; Basile J. Analysis of Recent Papers in Hypertension Jan Basile, MD, Senior Editor. J. Clin. Hypertens. 2008, 10, 160–163, https://doi.org/10.1111/j.1751-7176.2008.00028.x PMID: 19090884
- 65. Howard B. V.; Lee E.T.; Yeh J.L.; Go O.; Fabsitz R.R.; Devereux R.B.; et al. Hypertension in Adult American Indians: The Strong Heart Study. *Hypertension* 1996, 28, 256–264, https://doi.org/10.1161/01.HYP.28.2.256 PMID: 8707391
- 66. Hicks Patrice M, Collazo Melendez Samuel A, Albert Vitale, William Self, Mary Elizabeth Hartnett, Paul Bernstein, et al. Genetic Epidemiologic Analysis of Hypertensive Retinopathy in an Underrepresented and Rare Federally Recognized Native American Population of the Intermountain West. J. Community Med. Public Heal. 2019, 3, https://doi.org/10.29011/2577-2228.100052

- 67. Leggio M.; Lombardi M.; Caldarone E.; Severi P.; D'emidio S.; Armeni M.; et al. The Relationship between Obesity and Hypertension: An Updated Comprehensive Overview on Vicious Twins. Hypertens. Res. 2017, 40, 947–963, https://doi.org/10.1038/hr.2017.75 PMID: 28978986
- 68. Solomon M.; Shiferaw B.Z.; Tarekegn T.T.; Gebre Eyesus F.A.; Mengist S.T.; Mammo M.; Mewahegn A.A.; Mengiste B.T.; Terefe T.F. Prevalence and Associated Factors of Hypertension Among Adults in Gurage Zone, Southwest Ethiopia, 2022. SAGE Open Nurs. 2023, 9, https://doi.org/10.1177/23779608231153473 PMID: 36761364
- 69. Hall J.E.; do Carmo J.M.; da Silva A.A.; Wang Z.; Hall M.E. Obesity, Kidney Dysfunction and Hypertension: Mechanistic Links. *Nat. Rev. Nephrol.* 2019, 15, 367–385, https://doi.org/10.1038/s41581-019-0145-4 PMID: 31015582
- Imai Y. A Personal History of Research on Hypertension From an Encounter with Hypertension to the Development of Hypertension Practice Based on Out-of-Clinic Blood Pressure Measurements. *Hypertens. Res.* 2022, 45, 1726–1742, https://doi.org/10.1038/s41440-022-01011-1 PMID: 36075990
- 71. Nguyen T.T.; Nguyen M.H.; Nguyen Y.H.; Nguyen T.T.P.; Giap M.H.; Tran T.D.X.; et al. Body Mass Index, Body Fat Percentage, and Visceral Fat as Mediators in the Association between Health Literacy and Hypertension among Residents Living in Rural and Suburban Areas. Front. Med. 2022, 9, https://doi.org/10.3389/fmed.2022.877013 PMID: 36148456
- Choi J.W.; Han E.; Kim T.H. Risk of Hypertension and Type 2 Diabetes in Relation to Changes in Alcohol Consumption: A Nationwide Cohort Study. *Int. J. Environ. Res. Public Health* 2022, 19, https://doi.org/10.3390/ijerph19094941 PMID: 35564335
- 73. Klatsky A.L. Alcohol-Associated Hypertension When One Drinks Makes a Difference. *Hypertension* 2004, 44, 805–806, https://doi.org/10.1161/01.HYP.0000146538.26193.60 PMID: 15492132
- Legese N.; Tadiwos Y. Epidemiology of Hypertension in Ethiopia: A Systematic Review. Integr. Blood Press. Control 2020, 13, 135–143, https://doi.org/10.2147/IBPC.S276089 PMID: 33116810
- 75. Koya S.F.; Pilakkadavath Z.; Chandran P.; Wilson T.; Kuriakose S.; Akbar S.K.; et al. Hypertension Control Rate in India: Systematic Review and Meta-Analysis of Population-Level Non-Interventional Studies, 2001–2022. Lancet Reg. Heal.—Southeast Asia 2023, 9, https://doi.org/10.1016/j.lansea. 2022.100113 PMID: 37383035
- 76. Mayl J.J.; German C.A.; Bertoni A.G.; Upadhya B.; Bhave P.D.; Yeboah J.; et al. Association of Alcohol Intake with Hypertension in Type 2 Diabetes Mellitus: The Accord Trial. J. Am. Heart Assoc. 2020, 9, https://doi.org/10.1161/JAHA.120.017334 PMID: 32900264
- Overweight and Hypertension. Nutr. Rev. 1969, 27, 168–171, https://doi.org/10.1111/j.1753-4887. 1969.tb05009.x
- Kempner W. Treatment of Hypertensive Vascular Disease with Rice Diet. Am. J. Med. 1948, 4, 545–577, https://doi.org/10.1016/0002-9343(48)90441-0 PMID: 18909456