Journal of Current Science and Technology, October-December 2025 Vol. 15 No. 4, Article 140

ISSN 2630-0656 (Online)

Cite this article: Nuryunarsih, D., Rauf, S., Okatiranti, Wahyuningsih, H. P., Zaidan, H., Arshad, A., & Rizvi, S. S. R. (2025). Predicting systolic and diastolic blood pressure response using machine learning: A 96-feature analysis in hypertensive patients with comorbidities. *Journal of Current Science and Technology*, *15*(4), Article 140. https://doi.org/10.59796/jcst.V15N4.2025.140

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Predicting Systolic and Diastolic Blood Pressure Response Using Machine Learning: A 96-Feature Analysis in Hypertensive Patients with Comorbidities

Desy Nuryunarsih[1,*], Sania Rauf[2,3], Okatiranti[4,5], Heni Puji Wahyuningsih[6], Mahija Zaidan[7], Lucky Herawati[8], Abida Arshad[1], and Syed Shakeel Raza Rizvi[1]

[1]Faculty of Medical Sciences, Population Health Sciences Institute, Newcastle University, United Kingdom
[2]Department of Zoology, Wildlife, and Fisheries, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan
[3]Department of Biosciences, University of Wah, Quaid Avenue, Wah Cantt, Pakistan
[4]School of Medicine, University of Nottingham, United Kingdom
[5]Adhirajasa Reswara Sanjaya University, Bandung, West Java, Indonesia
[6]Department of Midwifery, Health Polytechnic, Ministry of Health, Yogyakarta, Indonesia
[7]School of Law and Social Science, University of Nottingham, United Kingdom
[8]Nursing Department, Health Polytechnic, Ministry of Health, Yogyakarta, Indonesia

*Corresponding author; E-mail: desy.nuryunarsih@newcastle.ac.uk
Orcid ID 0000-0002-5306-0467

Received 7 May 2025; Revised 15 June 2025; Accepted 3 July 2025; Published online 20 September 2025

Abstract

Hypertension represents a complex condition that substantially increases the global cardiovascular disease burden and related deaths. This study compares three tree-based machine learning approaches-Decision Tree (DT), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost)-using 96 multi-domain features to predict reductions in both systolic and diastolic blood pressure following antihypertensive treatment in patients with varying comorbidity profiles. Our approach utilizes paired t-test analyses to examine blood pressure changes before and after medication across different patient categories, while employing comprehensive decision tree visualisation to create interpretable decision pathways that identifying predictive associations between medications and blood pressure outcomes. Analysis of 160 patients indicated significant blood pressure improvements in all studied patient groups, with systolic blood pressure reductions showing statistical significance ($p = 0.001$) and diastolic blood pressure changes demonstrating similar significance levels ($p = 0.02$). The Decision Tree method showed optimal performance for systolic blood pressure prediction, recording 93% F1-score and 83% AUC values, whilst Random Forest demonstrated excellence performance in diastolic blood pressure prediction with 98% F1-score and 92% AUC. XGBoost performed less effectively than the other two algorithms across metrics. Through decision tree analysis, we identified strong predictive associations between diuretics and ACE inhibitors with systolic blood pressure reduction, whilst nitrate compounds and combined medication regimens showed significant predictive relationships with diastolic blood pressure decrease. The machine learning models successfully integrated diverse patient characteristics across multiple domains, including demographics, clinical parameters, lifestyle factors, and socioeconomic determinants. Our findings from this 160-patient cohort demonstrate the clinical utility of interpretable machine learning models for medication response prediction, providing valuable insights that can guide personalized antihypertensive therapy selection and inform clinical decision-making through data-driven treatment approaches.

Keywords: *hypertension; decision tree; random forest; XG-Boost; hypertensive medication*

1. Introduction

Hypertension represents one of the most significant modifiable risk factors for cardiovascular disease, affecting approximately one-third of the global population and contributing substantially to cardiovascular morbidity and mortality worldwide (Mills et al., 2020). The clinical management of hypertension becomes increasingly complex when patients present with comorbidities such as cardiovascular disease and diabetes, necessitating personalized treatment approaches that consider multiple patient factors simultaneously (Elendu et al., 2024). However, predicting individual patient responses to specific medications remains challenging due to the multifactorial nature of blood pressure regulation and the complex interplay between patient characteristics, comorbidities, and treatment efficacy (Treebupachatsakul et al., 2022; Panyamit et al., 2022).

Machine learning approaches have shown promise in hypertension management, with previous studies employing various algorithms for treatment prediction. Existing investigations have primarily focused on general hypertension prediction or single-algorithm approaches. Nuryunarsih et al. (2023) found that logistic regression and neural networks performed better in Indonesian hypertensive patients (Nuryunarsih et al., 2023), while other studies have utilised RFECV and XGBoost for treatment outcomes (Chang et al., 2019). However, these approaches have shown limited attention to comparative analyses of tree-based methods for predicting both systolic and diastolic blood pressure responses in patients with multiple comorbidities.

This study addresses a critical research gap by being the first to employ three tree-based machine learning algorithms Decision Tree (DT), Random Forest (RF), and XGBoost to predict both SBP and DBP reductions following antihypertensive treatment in hypertensive patients with cardiovascular and diabetic complications. Our novel approach incorporates 96 features spanning clinical parameters, sociodemographic characteristics, lifestyle factors, and medication profiles to create comprehensive predictive models. Importantly, we provide interpretable decision tree visualizations that illustrate medication-specific pathways for blood pressure reduction, offering clinically actionable insights for personalized treatment selection. To our knowledge, this represents the first investigation to combine such an extensive multi-domain feature set with comparative tree-based algorithm analysis, whilst providing interpretable decision support tools for antihypertensive medication optimisation in complex patient populations.

2. Objective

This study aims to predict decreases in systolic blood pressure (SBP) and diastolic blood pressure (DBP) in hypertensive patients with cardiovascular and diabetic complications following antihypertensive medication treatment using three tree-based machine learning algorithms: Decision Tree (DT), Random Forest (RF), and XGBoost. Specifically, we will: (1) develop and compare the predictive performance of DT, RF, and XGBoost models using 96 clinical and demographic features, measuring accuracy, sensitivity, specificity, and AUC; (2) identify the most influential features contributing to blood pressure reduction and quantify their relative importance; (3) visualize decision pathways through decision tree models to illustrate how different antihypertensive medications and patient characteristics influence blood pressure reduction; and (4) provide interpretable decision rules to guide clinicians in selecting optimal treatment strategies for patients with these comorbidities. This research addresses a current gap in the literature, as no previous study has examined SBP and DBP reduction using these specific algorithms while providing clinical decision support through visualization of antihypertensive medication effects.

3. Methods and Analysis

This cross-sectional study was conducted on 160 hypertensive patients admitted to or visiting Fauji Foundation Hospital, Rawalpindi, with ethical approval from the Research Ethics Committee of Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi. Participants aged $\geq$ 19 years with diagnosed hypertension- with or without cardiovascular disease or diabetes-were included based on symptoms of high blood pressure, such as headache, dizziness, blurred vision, chest pain, shortness of breath, nausea, and sleep apnea. Hypertension was defined as three measured values of systolic blood pressure > 140 mmHg or diastolic blood pressure > 90 mmHg, or current use of antihypertensive medication. Written informed consent (in English/Urdu) was obtained from all participants after they were briefed about the study purpose and protocol.

Patient data were collected using a structured questionnaire specifically developed for this study, encompassing demographics (age, gender, home location, education, occupation), clinical symptoms, medication history, hereditary factors, dietary habits, lifestyle factors, socioeconomic status, smoking habits, and biometric information (height, weight, BMI). The primary outcomes were defined as binary

variables representing clinically meaningful reductions in systolic blood pressure (SBP) and diastolic blood pressure (DBP) following antihypertensive treatment. A reduction was defined as follows: SBP decrease ≥ 10 mmHg from baseline to post-treatment, and DBP decrease ≥ 5 mmHg from baseline to post-treatment.

These targeted binary thresholds were established through a review of current clinical evidence demonstrating clinically significant blood pressure improvements. Research indicates that systolic reductions of 10 mmHg or greater correlate with substantial decreases in cardiovascular events, including a 20% lower incidence of coronary heart disease (Hong, 2017). Correspondingly, diastolic reductions of 5 mmHg represent meaningful therapeutic benefits in cardiovascular risk mitigation (Canoy et al., 2022).

The dataset was systematically organized and divided using machine learning techniques into input features (X) and binary output labels representing reductions in SBP (y_1) and DBP (y_2), with comprehensive details provided in Tables 1 and 2. Three tree-based machine learning models were applied Decision Tree (DT), Random Forest (RF), and XGBoost to predict blood pressure reductions using various Python libraries including pandas, scikit-learn, NumPy, and matplotlib. Specifically, DecisionTreeClassifier and RandomForestClassifier algorithms from the scikit-learn library were utilised for DT and RF respectively, while XGBClassifier from the xgboost library was employed for XGBoost analysis. Decision tree visualisation was performed using Graphviz and Pydotplus libraries. Data analysis was conducted using exploratory techniques in Jupyter Notebook, and model performance evaluated using standard classification metrics.

3.1 Data Preparation and Sample Size Considerations

We processed survey responses from 160 participants, converting textual data into numerical, scaled datasets suitable for machine learning analysis. Missing data handling was performed systematically: one missing age entry was imputed using the median age value, while missing blood pressure values were addressed using established clinical ranges with mean or median imputation within these ranges to ensure data completeness and accuracy. We investigated the correlations between dependent and independent variables to understand their impact on both systolic and diastolic blood pressure reductions. Information about the input and output features is provided in Appendix Tables A1 and A2.

Our study's sample size of 160 participants with 96 predictor features aligns with established statistical guidelines for developing robust predictive models with binary outcomes. When building predictive models, the relationship between sample size (n), number of outcome events (E), and predictor parameters (p) is crucial for model reliability. According to methodological standards, adequate sample size should satisfy three key criteria: (1) minimal optimism in effect estimates, indicated by a global shrinkage factor ≥ 0.9; (2) negligible overfitting, demonstrated by a difference ≤ 0.05 between apparent and adjusted model performance (Nagelkerke's R^2), and (3) precise population risk estimation (Riley et al., 2020). In our study, with approximately 80% of participants experiencing blood pressure reductions (128 events for SBP, 130 for DBP), we achieved an events-per-predictor (EPP) ratio of approximately 1.3-1.4. While traditional rules suggest 10-20 EPP, recent research indicates that lower EPP ratios can be acceptable when using machine learning algorithms with appropriate regularization and cross-validation techniques.

The 96 features included in this study were comprehensively selected through a systematic approach combining literature review and clinical expertise. A structured questionnaire was specifically developed for this research based on an extensive review of existing literature on hypertension prediction and treatment response. The questionnaire content and feature selection were validated through consultations with three clinical professors specializing in cardiovascular medicine and hypertension management.

All questionnaire items were retained as features to ensure comprehensive coverage of relevant clinical, demographic, and therapeutic variables without loss of potentially important predictive information. This approach was chosen over automated feature reduction techniques (such as PCA, LASSO, or mRMR) to maintain clinical interpretability and ensure that all expert-identified variables were available for analysis. The decision to include the full feature set was further supported by our use of tree-based algorithms, which inherently perform feature selection through their splitting mechanisms and can handle high-dimensional data effectively while identifying the most predictive variables through built-in importance rankings.

3.2 Model Development and Validation Strategy

Our use of tree-based algorithms (Decision Trees, Random Forest, XGBoost) with built-in feature selection and pruning mechanisms helps mitigate overfitting concerns. The 96 features were carefully selected to comprehensively capture relevant clinical, demographic, and therapeutic variables, allowing for nuanced prediction of treatment response while maintaining statistical validity through rigorous validation procedures. Multicollinearity among features was assessed using correlation analysis, with highly correlated features (correlation coefficient > 0.9) identified and addressed to prevent redundancy in the models. The binary classification targets (SBP/DBP reduction: yes/no) were assessed for class balance, and with 80% of participants showing reductions, class imbalance was addressed through appropriate weighting techniques during model training. We employed k-fold cross-validation (k = 5) to ensure robust model evaluation and prevent overfitting to specific data partitions. Hyperparameters were optimized using grid search with cross-validation to identify optimal model configurations. Performance metrics were calculated per class with confidence intervals reported across cross-validation folds to establish statistical reliability, whilst feature importance was calculated using built-in tree-based importance rankings to identify key predictors of blood pressure reductions.

3.3 Decision Tree

Decision Trees (DT) are widely used in data mining to discover patterns and extract insights from data. Once constructed, a DT can predict outcomes for new data based on patterns learned from the training dataset. Essentially, a DT serves as a structured framework for storing and applying learned experiences (Sheppard, 2019). In this study, we developed classification models to predict whether patients would experience decrease in systolic blood pressure (SBP) or diastolic blood pressure (DBP) based on available features. Our dataset comprised 160 participants with 96 features, encompassing demographic information, health status, medications, symptoms, hereditary factors, and dietary habits. Our objective was twofold: to train the algorithm to identify blood pressure reductions and to visualize the decision pathways that lead to SBP and DBP decreases following antihypertensive treatment.

We employed the Classification and Regression Trees (CART) algorithm, which is implemented in Scikit-Learn (Sheppard, 2019). Unlike other decision tree algorithms such as ID3 and C4.5 (developed by Quinlan), CART handles both categorical and continuous variables and incorporates pruning techniques to prevent overfitting.

CART is one of several decision tree algorithms, alongside others like ID3 and C4.5. While ID3 (developed in the 1980s) and its successor C4.5 use entropy as their splitting criterion, CART uses the Gini index (Che et al., 2011). The underlying principle is similar across these algorithms: they recursively partition data into subgroups based on available attributes using specific splitting criteria. Both entropy and the Gini index quantify the impurity of data partitions, where lower values indicating greater homogeneity within groups. An entropy value of 0 represents perfectly homogeneous data, while a value of 1 indicates maximum randomness (equal distribution between classes) (Fernández et al., 2016; Géron, 2017). We selected the Gini index for our study because it directly measures the probability of misclassification, making it particularly suitable for our binary classification problem (Rahmati et al., 2022). In our clinical application, these subgroups represent hypertensive patients with varying comorbidity profiles, including cardiovascular disease (CVD) and diabetes complications.

The primary advantage of DTs lies in their interpretability and transparent decision-making process. They excel at revealing patterns in complex datasets that lead to specific outcomes. However, this transparency comes with potential limitations, particularly when dealing with imbalanced data where the algorithm may favor majority classes while fragmenting minority classes. To mitigate this bias, we considered ensemble methods, such as bagging and boosting, which combine multiple decision trees to create more robust and balanced predictions (Fernández et al., 2016).

3.4 Random Forest

To overcome the limitations of decision trees, we used an ensemble method called Random Forest. The idea is to build multiple decision tree models and combine their results. We can use the “Sklearn.ensemble module” in “SCikit-Learn”, which provides the “ens.RandomForestClassifier” for creating Random Forest models. Random Forests are designed to create a more robust model by utilizing multiple decision trees, a technique known as bagging or bootstrap aggregation. This approach helps overcome overfitting and reduces bias by considering multiple prediction results (Sheppard, 2019).

3.5 Boosting

Boosting is another ensemble method where predictions are made in stages and sequentially to achieve better results (Seni & Elder, 2010). This is different from Random Forests, which combine the results of many models built and executed in parallel (Kaur et al., 2022). In this study, we used XGBoost, which is one implementation of boosting. Scikit-learn also provides several implementations of XGBoost for machine learning algorithms. Our dataset containing 96 features across multiple domains (clinical, demographic, lifestyle, socioeconomic), likely involves complex non-linear relationships and feature interactions that tree-based methods naturally capture without requiring extensive feature engineering or transformation.

3.5.1 Measuring the Performance of the Model

There are many ways to measure the performance of the model that we made as shown in Figure 1, we will start with the confusion matrix, which includes true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

		Actual	
		Positive	Negative
Prediction	Positive	TP	FP
	Negative	FN	TN

Figure 1 Confusion matrix for model prediction performance

True positive (TP) refers to when the model successfully predicts a positive outcome because the actual outcome is indeed positive. It means that the model correctly identifies instances as positive when they are actually positive.

True negative (TN) refers to when the model correctly predicts a negative outcome because the actual outcome is indeed negative. It means that the model accurately identifies instances as negative when they are actually negative.

False positive (FP) refers to when the model incorrectly predicts a positive outcome when the actual outcome is negative. It means that the model mistakenly identifies instances as positive when they are actually negative.

False negative (FN) refers to when the model incorrectly predicts a negative outcome when the actual outcome is positive. It means that the model fails to identify instances as positive when they are actually positive.

3.5.2 Accuracy

Accuracy measures how often the model correctly predicts both positive and negative outcomes. It represents the overall correctness of the model's predictions, considering both false positives and false negatives. It is calculated by dividing the total number of correct predictions (true positives and true negatives) by the total number of instances in the dataset.

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

3.5.3 Precision

Precision, also known as positive predictive value, measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It focuses on the accuracy of positive predictions.

$$Precision = \frac{TP}{Total\ positive\ prediction} = \frac{TP}{TP+FP}$$

3.5.4 Sensitivity (Recall)

Sensitivity, also known as recall or true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances. It focuses on the ability of the model to correctly identify positive instances.

$$Sensitivity = \frac{TP}{Total\ positive\ actual} = \frac{TP}{TP+FN}$$

3.5.5 Area under the Curve

AUC, which stands for Area Under the Curve, refers to the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve plots the true positive rate (sensitivity) against the false positive rate. AUC measures how effectively the model distinguishes between positive and negative instances. A higher AUC value indicates better overall model performance in terms of classification accuracy.

3.5.6 F1-Score

The F1-score formula is calculated using the formula:

$F1 = (2 \times Precision \times Recall) / (Precision + Recall)$.

This metric combines both precision and recall (sensitivity) into a single value, providing a balanced evaluation of the model's performance

3.7 Analysis

In this research, we applied statistical methods and machine learning models. First, we compared systolic blood pressure before and after treatment using a paired t-test. Then, we implemented Decision Tree (DT), Random Forest (RF), and XGBoost models, developed using Python 3.7. The paired t-test was conducted using the scipy.stats library, while model training and evaluation were performed with scikit-learn (version 0.21.2). To evaluate different machine learning methods, the dataset was divided into a training set and a testing set, with 80% of the data randomly allocated for training and the remaining 20% used for testing.

3.8 Ethical Approval

Participants received a comprehensive explanation of the study's purpose and procedures before providing their written consent in either English or Urdu. The study population consisted of patients evaluated at Military Hospital (MH) Rawalpindi and Fauji Foundation Hospital (FFH) Rawalpindi, Pakistan.

This research adhered to the principles outlined in the Declaration of Helsinki (2013 revision). The Ethics Committee of Pir Mehr Ali Shah Arid Agriculture University Rawalpindi granted institutional approval for this study involving human subjects (Approval No. PMAS-AAUR/1406). All participants provided informed consent prior to enrollment.

4. Result

4.1 Descriptive statistics

Table 1 presents descriptive statistics for 160 participants with varying conditions: hypertension (HTN), hypertension with cardiovascular disease (HTN-CVD), hypertension with diabetes (HTN-DM), and a combination of all three (HTN-CVD-DM). Participant ages ranged from 13 to 72 years. The youngest diagnosed hypertensive patient was 13, while the oldest was 72. In the HTN-CVD group, ages ranged from 19 to 60, and among patients with diabetes, from 26 to 57. The average age at first diagnosis was 45 years for HTN, 46 for CVD, and 44 for diabetes.

Appendix Table A3 analyzed the connection between different factors and the decrease in DBP. The variables with the strongest correlation were DBP before and after therapy, followed by morning headache, foot numbness, shakiness, regular fish consumption, systolic blood pressure after therapy, and light-headedness.

Appendix Table A4 examined the relationships between different features and decreases in SBP. The most strongly correlated variables were systolic blood pressure before and after therapy, followed by diastolic blood pressure, dry mouth, tingling sensation, morning headache, regular vegetable consumption, family history of hypertension, and headache.

Table 2 shows the results of the paired t-test analysis, which revealed significant differences between SBP and DBP values before and after medication: SBP ($p = 0.001$) and DBP ($p = 0.022$), with a significance level of 0.05.

Table 1 Descriptive statistics for participants HTN, HTN-CVD, HTN-DM, HTN-CVD-DM patients

	Age	AGE of HTN	Age of CVD	Age of Diabetes
Count	160	160	139	62
mean	52	45.2	46.8	44.3
std	8.6	9.8	8.8	7.6
min	19	13	19	26
25%	48	40	41	39.2
50%	53.5	48	49	45
75%	59	52	54	50
max	72	60	60	57

Table 2 Analysing the SBP and DBP before and after taking antihypertensive medications using a t-test

Variable		Before taking meds	After taking meds	Mean Diff	95% CI	Paired t_test p-value[a]
		Mean ± SD				
Hypertension (mmHg)	Systolic	146.7 ± 16.30	141.2± 15.8	5.5	(2.2-8.8)	0.001*
	Diastolic	91.4 ± 8.2	89.3 ± 9.8	2.1	(0.3-3.8)	0.022*

* Significant at 0.05 level, [a] paired *t-test*

Note: The confidence intervals shown are for the mean difference between before and after treatment values.

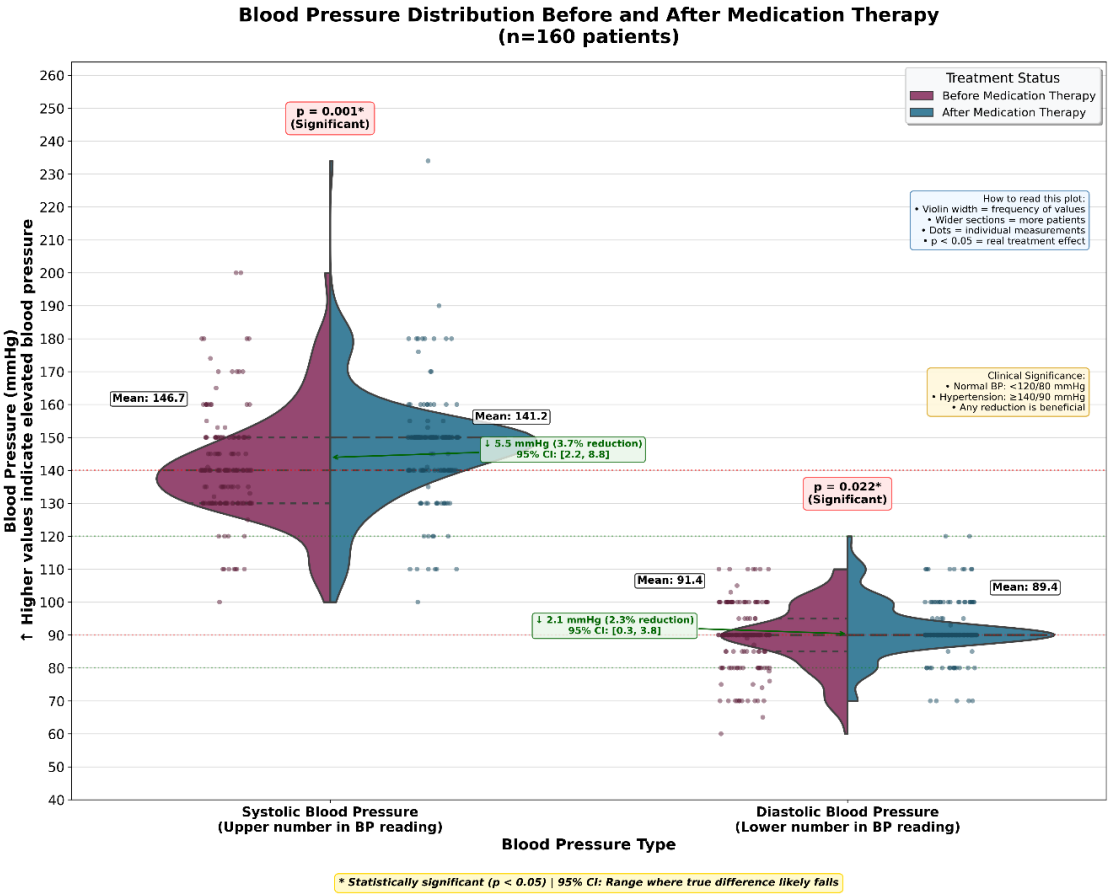
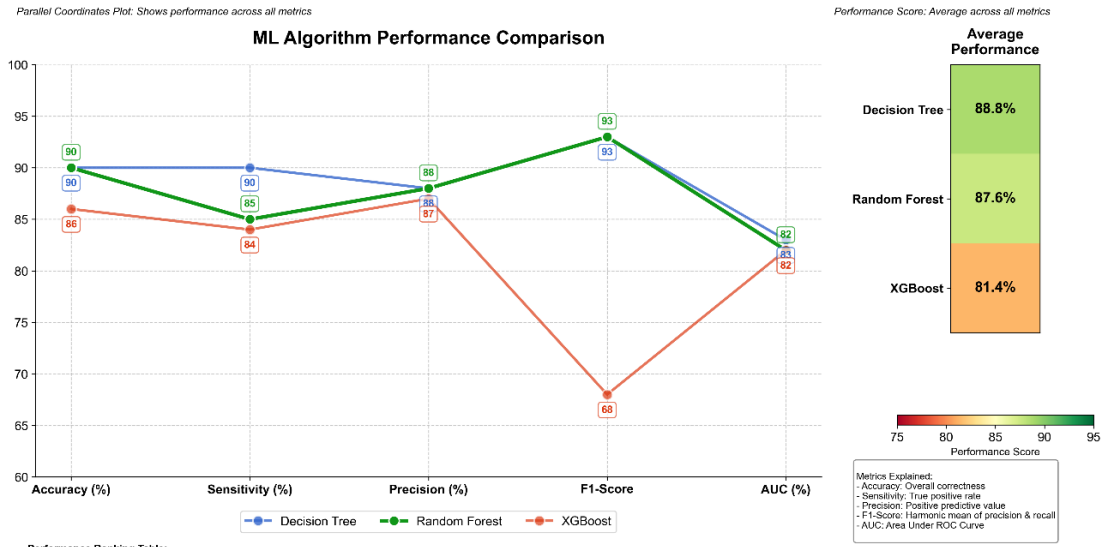


Figure 2 Violin plots showing systolic blood pressure (SBP) and diastolic blood pressure (DBP) values before and after therapy (mmHg)

Figure 2 shows the violin plot comparing SBP and DBP values before and after therapy. It illustrates the impact of the given medication on reducing SBP and DBP levels. The y-axis represents blood pressure in mmHg. The graph highlights slight differences in both SBP and DBP before and after medication.

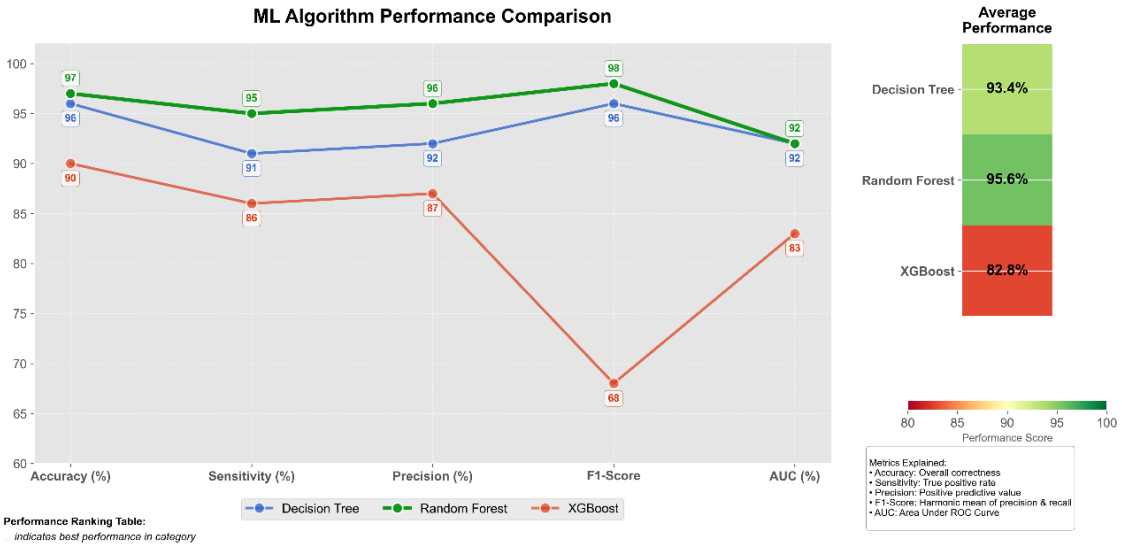
4.2 Prediction Model for a Decrease in Diastolic and Systolic Blood Pressure

The parallel coordinates plot in Figure 3 compares three machine learning algorithms for classification. The Decision Tree model achieved the highest overall average performance (88.8%) with the best metrics in Accuracy (90%), Sensitivity (90%), and AUC (83%). Random Forest followed with a strong average performance (87.6%) and the highest F1-score (93%). In contrast, XGBoost performed comparatively lower, with an overall average of 81.4% and a notably poor F1-score (68%), despite showing similar Precision scores across all three algorithms.



Rank	Algorithm	Accuracy	Sensitivity	Precision	F1-Score	AUC	Average
1	Decision Tree	90	90	88	93	83	88.8
2	Random Forest	90	85	88	93	82	87.6
3	XGBoost	86	84	87	68	82	81.4

Figure 3 Parallel coordinates plot comparing performance metrics of three machine learning algorithms for classifying decreases in systolic blood pressure (SBP)



Rank	Algorithm	Accuracy	Sensitivity	Precision	F1-Score	AUC	Average
1	Random Forest	97	95	96	98	92	95.6
2	Decision Tree	96	91	92	96	92	93.4
3	XGBoost	90	86	87	68	83	82.8

Figure 4 Parallel coordinates plot comparing performance metrics of three machine learning algorithms for classification (decrease DBP)

This parallel coordinates plot in Figure 4 compares three algorithms (Decision Tree, Random Forest, and XGBoost) across five performance metrics in term of decrease of diastolic blood pressure, with Random Forest demonstrating superior overall performance (95.6% average) compared to Decision Tree (93.4%) and XGBoost (82.8%). Random Forest excels particularly in Accuracy (97%), Precision (96%), and F1-Score (98%), while maintaining strong performance in Sensitivity (95%) and AUC (92%, tied with Decision Tree). XGBoost significantly underperforms in F1-Score (68%) compared to the other algorithms (96-98%), while all three models show their strongest metrics in different areas - Random Forest in F1-Score (98%), Decision Tree in F1-Score (96%), and XGBoost in Accuracy (90%).

Figure 5 illustrates the AUC values that reflect how the various classifiers are performing. A higher AUC signifies better predictive ability. DT leads with the highest AUC of 0.83, followed by RF with AUCs of 0.82. XGBoost trails with the lowest AUC of 0.82.

Figures 5 and 6 show the area under the curve (AUC) c-statistic for predicting SBP and DBP if the value is near 1, it means the model is almost perfect, but if the AUC c-statistic is below 0.5 means not a good model.

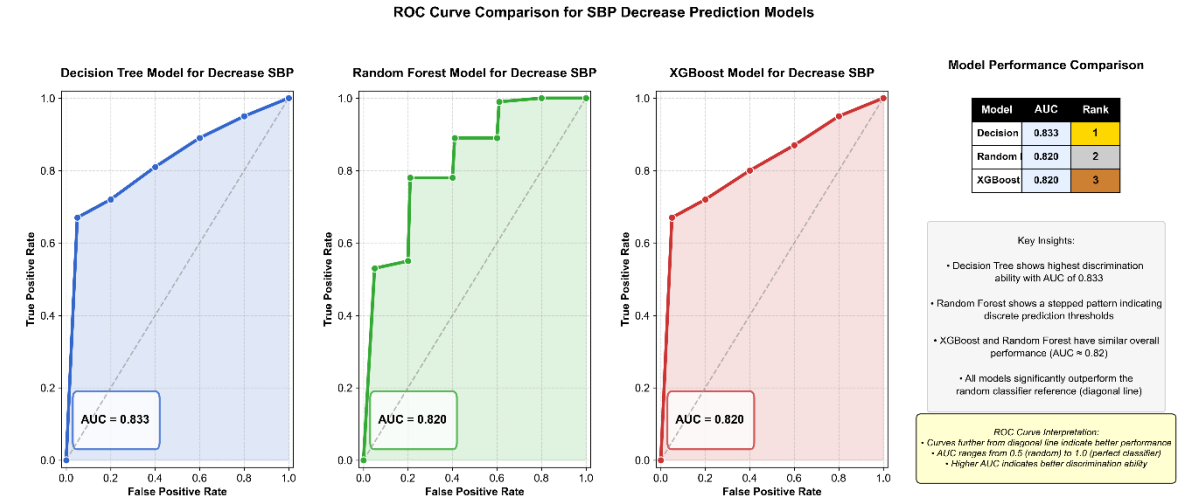


Figure 5 The AUC values that reflect how the various classifiers are performing

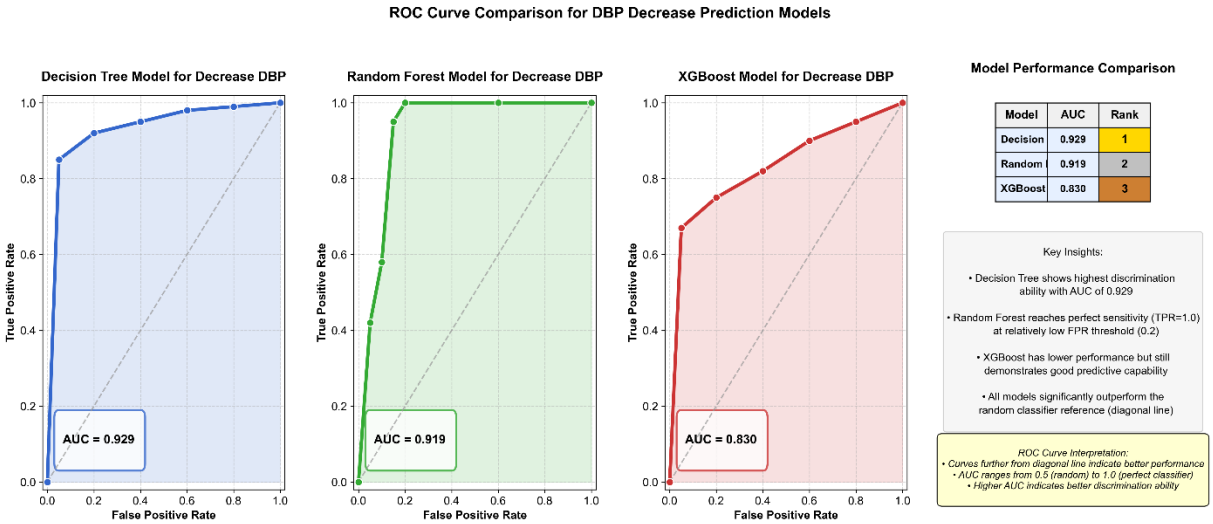


Figure 6 AUC values for different classification algorithms showing predictive performance

Decision tree in Figure 7 shows factors associated with reductions in systolic blood pressure (SBP) based on medication usage patterns. These patterns represent statistical associations rather than causal relationships. In this study, we can observe that the machine learning process considers the most statistically important feature to be whether the participant does not consume diuretics. Out of 124 patients, 4 individuals did not experience a decrease in systolic blood pressure, while 120 individuals did experience a decrease. If we further analyse the data, in the next branch, if the patient does not consume diuretics, the decision tree tests whether the patient does not use ACE inhibitors as their treatment. In this case, there are 57 patients who do not use ACE inhibitors, and among them, 14 individuals do not use Ascard Plus. Then, in the next branch, the decision tree tests whether the patient uses β-blockers. Among those who do not use diuretics, Ascard Plus, and β-blockers, there are 2 individuals whose blood pressure decreases.

In the analysis of patients who use diuretics on the right branch, we can observe more subdivisions. We can see that out of the patients who use diuretics, 3 individuals did not experience a decrease in blood pressure, while 62 did. From this, we can conclude that there are a total of 12 leaf nodes consisting of individuals whose blood pressure decreased, and 3 leaf nodes involving those whose blood pressure did not. If sorted from top to bottom, we can see that the statistically important factors identified by the machine learning decision tree in this study are the consumption of diuretics, ACE inhibitors, Nitromint, Ascard Plus, angised, β-blockers, combination medications, and Cardnit, according to their respective branches. It is important to note that these represent correlational patterns within our dataset, not established causal efficacy rankings. This interpretation also applies to the graph showing the decrease in diastolic blood pressure (Figure 8) below. This decision tree analysis helps us understand the statistical relationships between different variables and their association with decrease in systolic and diastolic blood pressure.

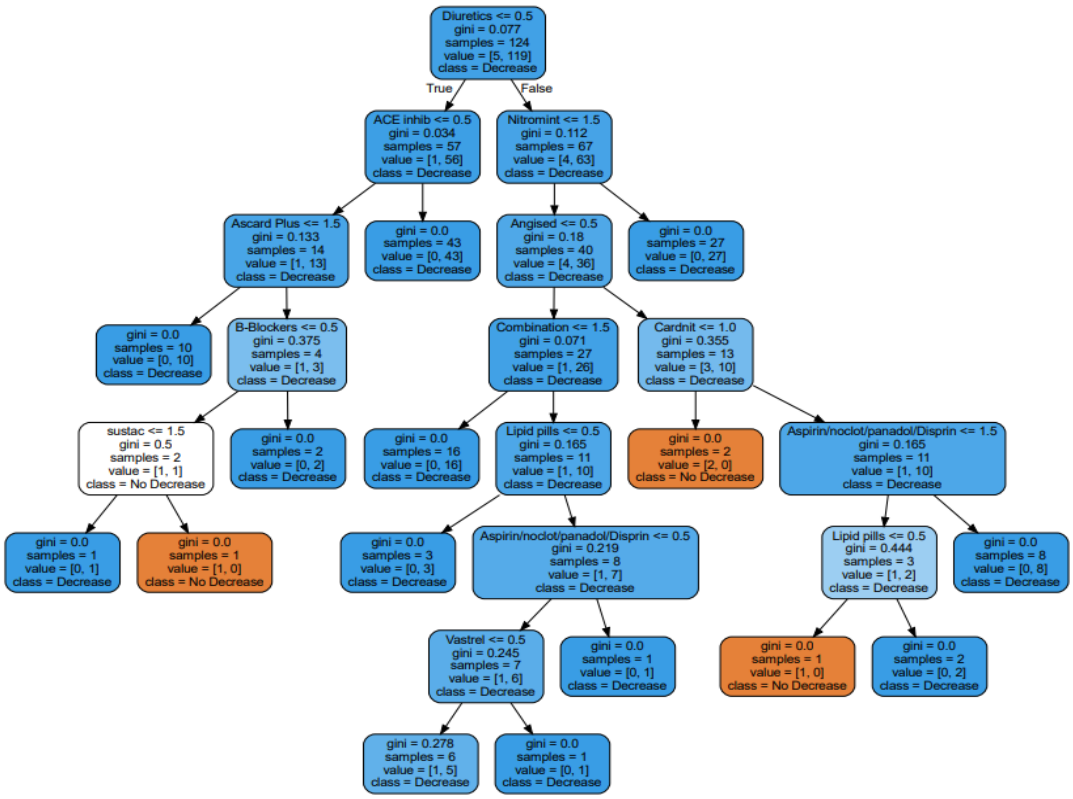


Figure 7 Decision tree showing factors associated with a decrease in systolic blood pressure based on medication usage patterns

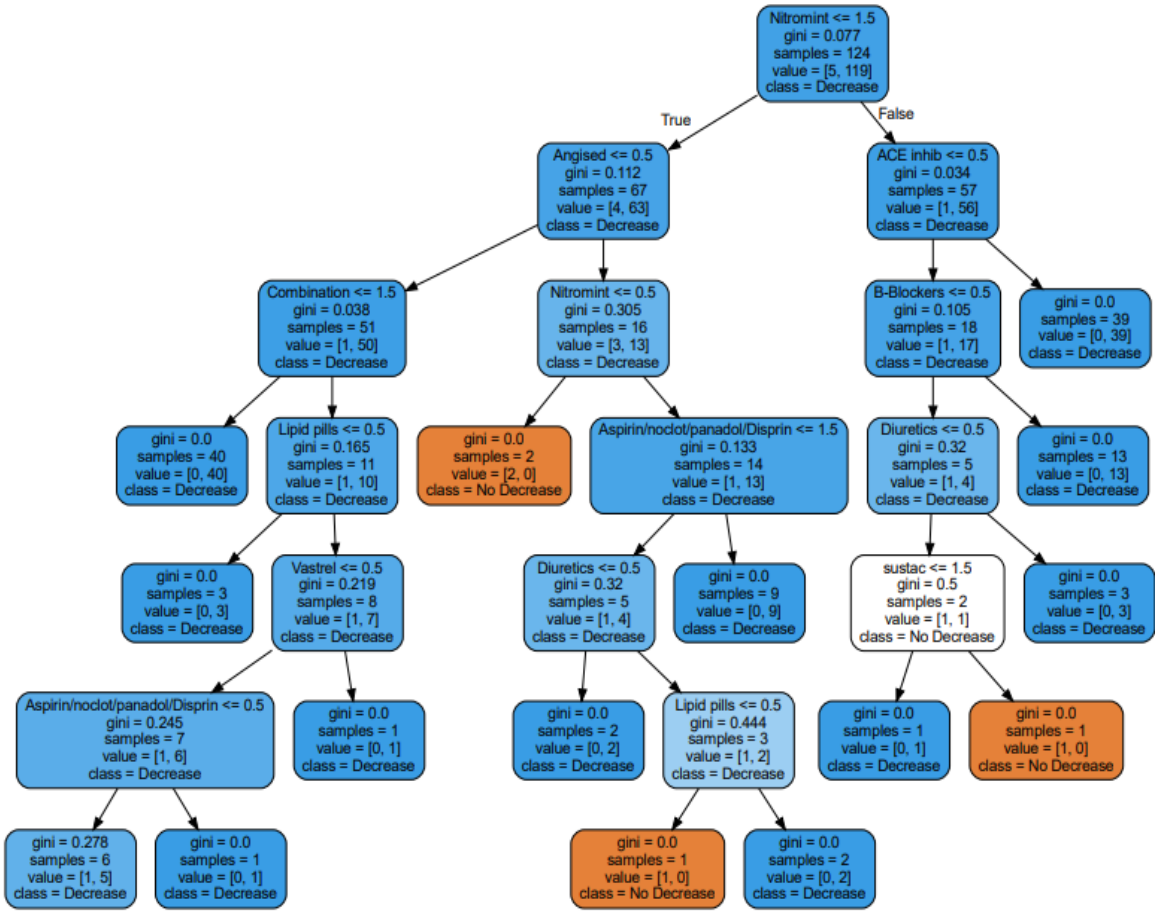


Figure 8 Decision Tree Decrease DBP information

In our study, using a decision tree to train the model, we found that four patients did not have lower SBP. Patient 1 does not consume sustac, β-blockers, Ascard Plus, ACE inhibitors, or diuretics. Patients 2 and 3 take Cardit, angised, and Nitromint but not diuretics. Additionally, patient 3, who also does not have a decrease their SBP, does not take lipid pills, cardnit or angised, but takes aspirin/panadol/disprin and diuretics. The decision tree visualisation for SBP prediction is shown in Figure 8, illustrating the hierarchical statistical importance and decision pathway. Regarding DBP, there were also four patients who did not lower their blood pressure. Patients 1 and 2 do not consume Nitromint or angised. Patient 3 does not consume lipid pills, angised, but consumes diuretics, aspirin/panadol/disprin, and nitromint. Patient 4 does not consume Sustac, diuretics, β-blockers, ACE inhibitors, or Nitromint.

4.3 Discussion

This study evaluated tree-based machine learning models Decision Tree (DT), Random Forest (RF), and XGBoost to predict reductions in systolic (SBP) and diastolic blood pressure (DBP) in hypertensive patients, including those with cardiovascular disease and diabetes. A key innovation of our approach was the integration of 96 multi-domain features, revealing previously unrecognized predictive patterns. Among these features, the most predictive for SBP reduction included pre/post SBP readings, dry mouth, tingling sensation, and vegetable intake, while for DBP, the top predictors were pre/post DBP values, morning headache, shakiness, and fish consumption.

Our findings demonstrate superior interpretability compared to traditional approaches. DT achieved the best performance for SBP prediction (F1-score: 93%, accuracy: 90%), while RF excelled in DBP prediction (F1-score: 98%, accuracy: 97%).

XGBoost underperformed in both tasks. These results differ markedly from prior studies by Nuryunarsih et al. (2023), who found that logistic regression and neural networks performed better in Indonesian hypertensive patients (Nuryunarsih et al., 2023). This discrepancy reveals an important methodological insight: our study's inclusion of detailed symptom profiles (dry mouth, tingling sensations) and dietary factors (vegetable and fish consumption) provided tree-based models with interpretable decision pathways that traditional statistical methods could not effectively utilize.

The interpretability of our decision tree models revealed clinically meaningful medication hierarchies that extend beyond existing knowledge. For SBP reduction, diuretics emerged as the primary splitting feature, followed by ACE inhibitors and Nitromint. Notably, for DBP reduction, Nitromint ranked as the most important feature, followed by Angised and ACE inhibitors, suggesting distinct therapeutic pathways for diastolic versus systolic control a differentiation not clearly established in previous literature. These findings align with established treatment guidelines (Al-Makki et al., 2022) but provide quantitative evidence of medication importance rankings in our diverse patient population comprising hypertension-only, hypertension with cardiovascular complications, and hypertension with diabetes.

A unique contribution of our study is the identification of non-traditional predictive factors. Beyond conventional clinical parameters, dietary factors emerged as significant predictors: regular fish consumption ranked among the top correlates for DBP reduction, while regular vegetable consumption was significant for SBP reduction. This dietary-medication interaction pattern has not been systematically quantified in previous machine learning studies of hypertension management. This builds upon previous research showing associations between diet and hypertension prevalence (Yang et al., 2022). Additionally, specific symptom profiles (morning headaches for DBP, dry mouth for SBP) demonstrated predictive value, suggesting that patient-reported symptoms could enhance treatment personalization.

Our subgroup analysis across hypertension-only, hypertension with cardiovascular disease, and hypertension with diabetes revealed differential feature importance patterns, enabling more precise treatment targeting. These findings support the clinical utility of interpretable machine learning models for personalized antihypertensive therapy, providing actionable insights that traditional statistical approaches have not captured.

4.4 Limitations and Contributions

Our study's sample size of 160 participants with 96 predictor features aligns with established statistical guidelines for developing robust prediction models with binary outcomes. While our results are based on a single-centre, cross-sectional dataset which may limit external validity, this study makes several important contributions to understanding behavioural and lifestyle factors in hypertension management that extend beyond these constraints. First, this is one of the pioneering studies to use interpretable machine learning algorithms to predict blood pressure response based on a comprehensive set of 96 features, including dietary habits, socioeconomic factors, and health behaviours in a complex patient population with comorbidities. To address potential generalisability concerns, we deliberately included a heterogeneous patient population with equal proportions of hypertension-only, cardiovascular complications, and diabetes comorbidities, providing insights across diverse clinical presentations rather than a homogeneous cohort. While our single-centre design limits immediate generalizability, several factors support the broader applicability of our findings. Our study population represents a diverse socioeconomic spectrum typical of urban Pakistani healthcare settings, with patients from varying educational backgrounds, income levels, and geographical locations within the region. The inclusion of equal proportions of patients with hypertension-only, cardiovascular complications, and diabetes reflects the complex comorbidity patterns commonly encountered in clinical practice globally. Second, our correlation analysis revealed crucial behavioural insights - regular fish consumption, vegetable intake, and dietary patterns emerged as significant predictors alongside traditional clinical measures. The decision tree visualization provides a transparent framework for understanding how lifestyle factors interact with physiological parameters to influence blood pressure control, offering actionable behavioural modification strategies for clinicians and patients that can be adapted across different healthcare settings.

Third, this study demonstrates the feasibility of capturing complex behavioural patterns even in resource-limited, single-centre settings. Our high predictive accuracy (90% for SBP, 97% for DBP) suggests that behavioural and lifestyle data, when properly analyzed, can be as valuable as traditional

clinical parameters in predicting treatment outcomes. The robust cross-validation procedures employed help ensure that these results reflect genuine predictive relationships rather than overfitting to our specific dataset. Fourth, while cross-sectional, our study provides immediate insights into which behavioural modifications might yield the most significant blood pressure improvements. This can help prioritize lifestyle interventions and create more targeted, personalized behavioural counselling approaches rather than generic lifestyle advice. The interpretable nature of our tree-based models makes these insights readily translatable to clinical practice across different healthcare contexts.

Finally, by incorporating socioeconomic status, dietary habits, and lifestyle factors into our predictive models, we address a critical gap in hypertension research that often focuses solely on clinical parameters. This holistic approach establishes a framework for future studies to explore the complex interplay between behaviour, socioeconomic factors, and treatment response. To mitigate single-centre limitations, we recommend future multi-centre validation studies across diverse populations and healthcare settings to establish the broader applicability of our findings, ultimately supporting more comprehensive and equitable hypertension management strategies.

4.5 Implications

These findings have significant clinical implications. The decision tree visualization provides a practical tool for clinicians to select initial antihypertensive therapy based on patient characteristics, predict which patients are likely to achieve target BP reductions, identify patients who may require combination therapy upfront, and personalize treatment approaches for complex patients with comorbidities. For instance, our model suggests that patients not responding to diuretics alone may benefit from the early addition of ACE inhibitors or β-blockers.

5. Conclusion

The study demonstrates that tree-based machine learning algorithms offer valuable tools for predicting blood pressure response in complex hypertensive populations. Specifically, Decision Tree (DT) algorithms achieved 90% accuracy in predicting systolic blood pressure (SBP) reduction, making them particularly suitable for SBP management in patients with hypertension alone or complicated by cardiovascular disease and diabetes. The interpretability of DT models provides clinicians with clear decision pathways, revealing that diuretics, ACE inhibitors, and β-blockers play pivotal roles in SBP control.

Random Forest (RF) algorithms showed superior performance for diastolic blood pressure (DBP) prediction, with 97% accuracy, suggesting their utility in fine-tuning DBP management. The ensemble nature of RF captures complex interactions between multiple factors, including dietary habits (fish consumption), symptoms (morning headache, foot numbness), and medication combinations, offering a more nuanced approach to DBP control.

These findings have significant clinical implications. By leveraging patient-specific characteristics, including demographics, lifestyle factors, comorbidities, and symptom profiles, clinicians can move beyond one-size-fits-all approaches to truly personalized hypertension management. The visualization of decision pathways enables evidence-based treatment selection, potentially reducing the trial-and-error period often associated with antihypertensive therapy optimization.

Furthermore, this research establishes a framework for integrating machine learning into routine clinical practice, particularly in resource-limited settings. The high predictive accuracy achieved with a relatively modest sample size suggests that these approaches can be implemented effectively even in single-center environments. Future integration of these algorithms into clinical decision support systems could significantly improve hypertension management outcomes, reduce healthcare costs, and enhance patient quality of life through more targeted and effective treatment strategies.

6. CRediT Author Contribution Statement

All authors contributed to the manuscript. DN curated data, conducted analysis, interpreted data, wrote results, reviewed, and edited. SR collected data, curated data, wrote, reviewed, and edited. OO wrote drafts, interpreted data, reviewed, and edited. HW curated data, wrote drafts, interpreted data, reviewed, and edited. MZ: handled project administration, curated data, and edited manuscripts. LH, AA and SS: wrote, supervised, reviewed, and edited. All authors read and approved the final version of the manuscript.

Conflict of interest: Nothing to declare

Funding: No funding

7. References

Al-Makki, A., DiPette, D., Whelton, P. K., Murad, M. H., Mustafa, R. A., Acharya, S., ... & Khan, T. (2022). Hypertension pharmacological treatment in adults: A World Health Organization guideline executive summary. *Hypertension*, *79*(1), 293-301. https://doi.org/10.1161/HYPERTENSIONAHA.121.18192

Canoy, D., Nazarzadeh, M., Copland, E., Bidel, Z., Rao, S., Li, Y., & Rahimi, K. (2022). How much lowering of blood pressure is required to prevent cardiovascular disease in patients with and without previous cardiovascular disease?. *Current Cardiology Reports*, *24*(7), 851-860. https://doi.org/10.1007/s11886-022-01706-4

Chang, W., Liu, Y., Xiao, Y., Yuan, X., Xu, X., Zhang, S., & Zhou, S. (2019). A machine-learning-based prediction method for hypertension outcomes based on medical data. *Diagnostics*, *9*(4), Article 178. https://doi.org/10.3390/diagnostics9040178

Che, D., Liu, Q., Rasheed, K., & Tao, X. (2011). Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Software Tools and Algorithms for Biological Systems*, 191-199. https://doi.org/10.1007/978-1-4419-7046-6_19

Elendu, C., Amaechi, D. C., Elendu, T. C., Amaechi, E. C., & Elendu, I. D. (2024). Dependable approaches to hypertension management: A review. *Medicine*, *103*(24), Article e38560. https://doi.org/10.1097/MD.0000000000038560

Fernández, L., Mediano, P., García, R., Rodriguez, J. M., & Marin, M. (2016). Risk factors predicting infectious lactational mastitis: Decision tree approach versus logistic regression analysis. *Maternal and Child Health Journal*, *20*(9), 1895-1903. https://doi.org/10.1007/s10995-016-2000-6

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow* (1st ed.). California, US: O'Reilly Media.

Hong, K. S. (2017). Blood pressure management for stroke prevention and in acute stroke. *Journal of Stroke*, *19*(2), 152-165. https://doi.org/10.5853/jos.2017.00164

Kaur, K., Sagar, A. K., & Chakraborty, S. (2022). Accelerating the performance of sequence alignment using machine learning with RAPIDS enabled GPU. *Journal of Current Science and Technology*, *12*(3), 462-481. https://doi.org/10.14456/jcst.2022.36

Mills, K. T., Stefanescu, A., & He, J. (2020). The global epidemiology of hypertension. *Nature Reviews Nephrology*, *16*(4), 223-237. https://doi.org/10.1038/s41581-019-0244-2

Nuryunarsih, D., Herawati, L., Badi'ah, A., Donsu, J. D. T., & Okatiranti. (2023). Predicting changes in systolic and diastolic blood pressure of hypertensive patients in indonesia using machine learning. *Current Hypertension Reports*, *25*(11), 377-383. https://doi.org/10.1007/s11906-023-01261-5

Panyamit, T., Sukvivatn, P., Chanma, P., Kim, Y., Premratanachai, P., & Pechprasarn, S. (2022). Identification of factors in the survival rate of heart failure patients using machine learning models and principal component analysis. *Journal of Current Science and Technology*, *12*(2), 336-348.

Rahmati, O., Avand, M., Yariyan, P., Tiefenbacher, J. P., Azareh, A., & Bui, D. T. (2022). Assessment of Gini-, entropy-and ratio-based classification trees for groundwater potential modelling and prediction. *Geocarto International*, *37*(12), 3397-3415. https://doi.org/10.1080/10106049.2020.1861664

Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B.,… & Collins, G. S. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ*, *368*, Article m441. https://doi.org/10.1136/bmj.m441

Seni, G., & Elder, J. F. (2010). *Ensemble methods in data mining: Improving accuracy through combining predictions*. California, US: Morgan & Claypool Publishers.

Sheppard, C. (2019). *Tree-based machine learning algorithms*. Chicago, US: Independently published.

Treebupachatsakul, T., Boosamalee, A., Shinnakerdchoke, S., Pechprasarn, S., & Thongpance, N. (2022). Cuff-less blood pressure prediction from ECG and PPG signals using Fourier transformation and amplitude randomization preprocessing for context aggregation network training. *Biosensors*, *12*(3), Article 159. https://doi.org/10.3390/bios12030159

Yang, Y., Yu, D., Piao, W., Huang, K., & Zhao, L. (2022). Nutrient-derived beneficial for blood pressure dietary pattern associated with hypertension prevention and control: Based on China nutrition and health surveillance 2015–2017. *Nutrients*, *14*(15), Article 3108. https://doi.org/10.3390/nu14153108

Appendix

Table A1 Input Features X

No	Features (X)	Type	Information
1.	Age	Age participants during study Categorical	Less than 18 years = 1 18-30 = 2 31-40 = 3 41-50 = 4 51-60 = 5 More than 61=6
2.	Age of HTN	Categorical	Less than 18 years = 1 18-30 = 2 31-40 = 3 41-50 = 4 51-60 = 5 More than 61=6
3.	Age of CVD	Categorical	Less than 18 years = 1 18-30 = 2 31-40 = 3 41-50 = 4 51-60 = 5 More than 61=6
4.	Age of Diabetes	Categorical	Less than 18 years = 1 18-30 = 2 31-40 = 3 41-50 = 4 51-60 = 5 More than 61= 6
5.	Height	Numerical	5-6
6.	BMI(kg/m^2)	Categorical	<18.5 (Underweight) = 1 18.5-29.9 (Normal Weight) = 2 25-29.9 (Overweight) = 3 >30 (Obesity)= 4
7.	BP at that time	Categorical	Normal (130-85) = 1 High (140/90-160/100) = 2 Extremely hight (above 160/100) = 3
8.	Health	Categorical	Poor = 1 Fair = 2 Good = 3 Very good = 4 Excellent = 5
9.	Headache	Categorical	Yes = 1 No = 2 Sometimes = 3
10.	Dizziness	Categorical	Yes = 1 No = 2
11.	Blurred vision	Categorical	Yes = 1 No = 2
12.	Nausea	Categorical	Yes = 1 No = 2 Sometimes = 3
13.	Sleep apnea	Categorical	Yes = 1 No = 2
14.	Pain/Discomfort (neck,jaw,back)	Categorical	Yes = 1 No = 2
15.	Feeling weak,lightheaded/faint	Categorical	Yes = 1 No = 2

Table A1 Cont.

No	Features (X)	Type	Information
16.	Chest Pain	Categorical	Yes = 1 No = 2
17.	Shortness of Breath	Categorical	Yes = 1 No = 2
18.	Indigestion	Categorical	Yes = 1 No = 2
19.	Palpitations	Categorical	Yes = 1 No = 2
20.	Thirst	Categorical	Yes = 1 No = 2
21.	Dry mouth	Categorical	Yes = 1 No = 2
22.	Appetite	Categorical	Yes = 1 No = 2
23.	frequent Urination	Categorical	Yes = 1 No = 2
24.	Morning Headache	Categorical	Yes = 1 No = 2
25.	night sweats	Categorical	Yes = 1 No = 2
26.	Light-headedness	Categorical	Yes = 1 No = 2
27.	shakiness	Categorical	Yes = 1 No = 2
28.	foot numbness	Categorical	Yes = 1 No = 2
29.	Tungling	Categorical	Yes = 1 No = 2
30.	Foot sores	Categorical	Yes = 1 No = 2
31.	leg cramping	Categorical	Yes = 1 No = 2
32.	When were HTN Diagnose	Categorical	<= 1 year = 1 1-10 years = 2 11-20 years = 3 >= 20 years = 4
33.	BP Frequency check	Categorical	on visit to doctor = 1 Daily = 2 Weekly = 3 Monthly = 4
34.	When Diabetes Diagnose	Categorical	3 months = 1 1 year = 2 more than 1 year = 3 more than 3 years = 4
35.	Glucose that time	Categorical	< 99/normal = 1 100-125/prediabetes = 2 >126/higher = diabetes
36.	Frequency of checking DM	Categorical	On visit to doctor = 1 Daily = 2 Weekly = 3 Monthly = 4
37.	Family HTN	Categorical	Yes=1 No=2
38.	HTN family Specify	Categorical	1st-degree relatives=1 2nd relatives = 2

Table A1 Cont.

No	Features (X)	Type	Information
39.	Family CVD	Categorical	Yes=1 No=2
40.	CVD family Specify	Categorical	1st-degree relatives=1 2nd degree relatives=2
41.	Family Diabetes	Categorical	Yes=01 No=02
42.	Specify family DM	Categorical	1st-degree relatives=1 2nd degree relatives=2
43.	Duration of HTN	Categorical	Less or equal to 1 year = 1 1-10 years = 2 11-20 years = 3 More than 20 years = 4
44.	Duration of CVD	Categorical	Less or equal to 1 year = 1 1-10 years = 2 11-20 years = 3 More than 20 years = 4
45.	Duration of Diabetes	Categorical	Less or equal to 1 year = 1 1-10 years = 2 11-20 years = 3 More than 20 years = 4
46.	ACE inhibitors	Categorical	'Zestril' 'Capoten' ‘renitec'=1 None = 2
47.	ARBs	Categorical	'eziday' 'Xavor' 'losartan’ ‘avsar’=1 None = 2
48.	Ca channels block	Categorical	'sofvasc', 'avsar', 'Norvasc','Adalat' = 1 None = 2
49.	Diuretics	Categorical	‘Spiromide' ‘Lasix' 'bepsar"xavor' ‘ditore', ‘carsel' = 1 None = 2
50.	B-Blockers	Categorical	'bisoprolol','monitor', 'carvedilol' ,'carsel', 'concor', 'merol or metroprolol', 'atenolol', 'Mepresor', 'carveda' = 1 None=2
51.	Combination	Categorical	'zestoretic' ,'CO-Eziday' = 1, None=2
52.	Aspirin/noclot/panadol/Disprin	Categorical	Yes=1 No=2
53.	Warfin	Categorical	Yes=1 No=2
54.	Lowplate	Categorical	Yes=1 No=2
55.	Loprin	Categorical	Yes=1 No=2
56.	Disprin	Categorical	Yes=1 No=2
57.	Angised	Categorical	Yes=1 No=2
58.	Ascard Plus	Categorical	Yes=1 No=2
59.	Cardnit	Categorical	Yes=1 No=2

Table A1 Cont.

No	Features (X)	Type	Information
60.	Nitromint	Categorical	Yes=1 No=2
61.	Vastrel	Categorical	Yes=1 No=2
62.	Naproxen	Categorical	Yes=1 No=2
63.	Digoxin	Categorical	Yes=1 No=2
64.	Rolip	Categorical	Yes=1 No=2
65.	Niglys	Categorical	Yes=1 No=2
66.	sustac	Categorical	Yes=1 No=2
67.	Nicorandil	Categorical	Yes=1 No=2
68.	Gastric pills	Categorical	Yes=1 No=2
69.	Lipid pills	Categorical	Yes=1 No=2
70.	Marital	Categorical	Single = 1 Married = 2 Married with cousin = 3 unmarried/divorce = 4
71.	Qualification	Categorical	Illiterate = 1 Primary = 2 High school graduate = 3 Bachelor or higher = 4
72.	Job	Categorical	Military/defence = 1 Transportation = 2 Homemaking = 3 Agriculture = 4 Retail = 5
73.	Income	Categorical	Unemployed = 1 500-10000 = 2 11000-20000 = 3 21000-30000 = 4 31000-40000 = 5 More than 50000 = 6
74.	Members	Categorical	Less than or equal to 5= 1 6-10 = 2 11-15 = 3
75.	CAR	Categorical	Yes = 01 No = 02
76.	House	Categorical	Yes = 01 No = 02
77.	Physical Activity	Categorical	Yes = 1 No = 2 sometime = 3
78.	Family Structure	Categorical	Single = 1 Both = 2
79.	Home Environment	Categorical	Pleasant = 01 Tense = 01

Table A1 Cont.

No	Features (X)	Type	Information
80.	Diet plan	Categorical	Yes = 1 No = 2 sometime = 3
81.	meals/day	Categorical	1 time/day = 1 2 times/day = 2 3 times/day = 3
82.	salt type	Categorical	Iodized =1 Non- iodized = 2
83.	fat diet	Categorical	Yes = 1 No = 2
84.	Milk	Categorical	N = 1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily = 5 Unsure = 6
85.	Eggs	Categorical	N = 1 Occasionally = 2 Sometime = 3 Mostly =4 Daily =5 Unsure= 6
86.	Meat	Categorical	N = 1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily = 5 Unsure = 6
87.	Chicken	Categorical	N = 1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily =5 Unsure = 6
88.	Fish	Categorical	N = 1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily = 5 Unsure = 6
89.	Pulses	Categorical	N =1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily = 5 Unsure = 6
90.	Vegetable	Categorical	N =1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily = 5 Unsure = 6

Table A1 Cont.

No	Features (X)	Type	Information
91.	Fruits	Categorical	N = 1 Occasionally = 2 Sometime = 3 Mostly = 4 Daily = 5 Unsure = 6
92.	Chest pain	Categorical	Yes = 1 No = 2
93.	SBP when diagnose or before medication	Continues (mmHg)	100-234
94.	DBP when diagnose or before medication	Continues (mmHg)	70-120
95.	SBP after prescribed medication	Continues (mmHg)	100-200
96.	DBP after prescribed medication	Continues (mmHg)	60-110

Table A2 Output features y1 and y2

Feature	Type	Information
Decrease of Systolic blood pressure after medication	Nominal	Systolic blood pressure decreased after taking medicines: 1. Yes, decreased 2. No, not decrease
Decrease of Diastolic blood pressure after medication	Nominal	Diastolic blood pressure decreased after taking medicines: 1. Yes, decreased 2. No, not decrease

Table A3 Pearson correlation between dependent and independent variable for decrease of diastolic blood pressure

No	Feature	Correlation	Rank
1.	DBP after prescribed medication	0.490	1
2.	DBP when diagnose or before medication	0.433	2
3.	Morning Headache	0.229	3
4.	foot numbness	0.228	4
5.	shakiness	0.222	5
6.	Fish	0.222	6
7.	SBP	0.220	7
8.	Lightheadedness	0.218	8
9.	Sleep apnea	0.210	9
10.	Income	0.208	10
11.	night sweats	0.206	11
12.	Palpitations	0.201	12
13.	Foot sores	0.196	13
14.	Height	0.192	14
15.	leg cramping	0.186	15
16.	Pain/Discomfort (neck,jaw,back)	0.186	16
17.	Vegetable	0.175	17
18.	Age	0.174	18
19.	Members	0.165	19
20.	Specify family DM	0.161	20
21.	fat diet	0.157	21
22.	When were HTN Diagnose	0.157	22
23.	Frequency of checking DM	0.157	23

Table A3 Cont.

No	Feature	Correlation	Rank
24.	frequent Urination	0.156	24
25.	Blurred vision	0.153	25
26.	When Diabete Diagnose	0.151	26
27.	Age of Diabetes	0.148	27
28.	BP at that time	0.147	28
29.	Family Diabetes	0.145	29
30.	Qualification	0.144	30
31.	Duration of HTN	0.137	31
32.	Home Env	0.133	32
33.	Thirst	0.133	33
34.	Dry mouth	0.131	34
35.	Tungling	0.130	35
36.	Warfin	0.128	36
37.	CVD family Specify	0.127	37
38.	Eggs	0.125	38
39.	HTN family Specify	0.121	39
40.	meals/day	0.120	40
41.	Health	0.118	41
42.	Dizziness	0.114	42
43.	Family CVD	0.113	43
44.	Apetite	0.109	44
45.	Fruits	0.108	45
46.	Glucose that time	0.105	46
47.	Duration of Diabetes	0.100	47
48.	Digoxin	0.100	48
49.	Indigestion	0.099	49
50.	Family HTN	0.097	50
51.	Aspirin/noclot/panadol/Disprin	0.097	51
52.	Headache	0.091	52
53.	SBP_Before	0.090	53
54.	Diet plan	0.086	54
55.	Loprin	0.084	55
56.	Nicorandil	0.083	56
57.	Disprin	0.082	57
58.	Naproxen	0.082	58
59.	Niglys	0.082	59
60.	salt type	0.081	60
61.	House	0.079	61
62.	AGE of HTN	0.078	62
63.	Vastrel	0.076	63
64.	Nitromint	0.076	65
65.	Cardnit	0.075	66
66.	B-Blockers	0.074	67
67.	Angised	0.073	68
68.	Ascard Plus	0.072	69
69.	Chiken	0.071	70
70.	Breath	0.071	71
71.	Family Struct	0.068	72
72.	Combination	0.065	73
73.	Milk	0.061	74
74.	Lowplate	0.061	75

Table A3 Cont.

No	Feature	Correlation	Rank
75.	sustac	0.056	76
76.	Treatment	0.054	77
77.	Chest pain	0.052	78
78.	Age of CVD	0.052	79
79.	Rolip	0.052	80
80.	BP Frequency check	0.051	81
81.	Meat	0.050	82
82.	Job	0.049	83
83.	ARBs	0.048	84
84.	Pulses	0.046	85
85.	Gastric pills	0.044	86
86.	Feeling weak,lightheaded/faint	0.042	87
87.	Duration of CVD	0.040	88
88.	CAR	0.036	89
89.	Diuretics	0.035	90
90.	ACE inhib	0.024	92
91.	Physical Activity	0.019	93
92.	Marital	0.017	94
93.	Nausea	0.011	95
94.	BMI(kg/m^2)	0.007	96
95.	Lipid pills	0.003	97
96.	Ca channels block	0.001	98

Table A4 Pearson correlation between dependent and independent variable for decrease of systolic blood pressure in patients

No	Feature	Correlation	Rank
1.	SBP after prescribed medication	0.477	1
2.	SBP when diagnose or before medication	0.430	2
3.	DBP	0.267	3
4.	Dry mouth	0.208	4
5.	Tungling	0.180	5
6.	Morning Headache	0.177	6
7.	Vegetable	0.176	7
8.	Members	0.165	8
9.	Headache	0.159	9
10.	Thirst	0.157	10
11.	night sweats	0.155	11
12.	shakiness	0.150	12
13.	B-Blockers	0.140	14
14.	Foot sores	0.139	15
15.	Lightheadedness	0.135	16
16.	foot numbness	0.132	17
17.	frequent Urination	0.132	18
18.	Lipid pills	0.128	19
19.	leg cramping	0.121	20
20.	CAR	0.119	21
21.	ARBs	0.118	22
22.	Lowplate	0.113	23
23.	Aspirin/noclot/panadol/Disprin	0.112	24
24.	When Diabete Diagnose	0.112	25
25.	Age	0.111	26
26.	Sleep apnea	0.104	27

Table A4 Cont.

No	Feature	Correlation	Rank
27.	AGE of HTN	0.104	28
28.	Ca channels block	0.103	29
29.	Home Env	0.100	30
30.	Milk	0.099	31
31.	Nitromint	0.098	32
32.	Duration of Diabetes	0.094	33
33.	BMI(kg/m^2)	0.091	34
34.	Diet plan	0.089	35
35.	Physical Activity	0.089	36
36.	Combination	0.087	37
37.	BP at that time	0.085	38
38.	Blurred vision	0.083	39
39.	Apetite	0.082	40
40.	Fish	0.081	41
41.	Job	0.079	42
42.	Glucose that time	0.078	43
43.	Family Struct	0.077	44
44.	Age of CVD	0.075	45
45.	Cardnit	0.075	46
46.	Gastric pills	0.074	47
47.	Age of Diabetes	0.074	48
48.	Nausea	0.073	49
49.	Height	0.072	50
50.	Rolip	0.070	51
51.	Frequency of checking DM	0.068	52
52.	Disprin	0.066	53
53.	Naproxen	0.066	54
54.	Niglys	0.066	55
55.	Dizziness	0.065	56
56.	Marital	0.063	57
57.	Health	0.063	58
58.	Angised	0.060	60
59.	ACE inhib	0.059	61
60.	DBP_Before	0.058	62
61.	Palpitations	0.058	63
62.	Loprin	0.055	64
63.	Digoxin	0.055	65
64.	Warfin	0.054	66
65.	Nicorandil	0.051	67
66.	Vastrel	0.049	68
67.	Chiken	0.048	69
68.	meals/day	0.046	70
69.	Breath	0.046	71
70.	Pain/Discomfort (neck,jaw,back)	0.044	72
71.	sustac	0.041	73
72.	Ascard Plus	0.040	74
73.	Meat	0.038	75
74.	Qualification	0.036	76
75.	Income	0.036	77
76.	fat diet	0.034	78
77.	Specify family DM	0.029	79

Table A4 Cont.

No	Feature	Correlation	Rank
78.	salt type	0.028	80
79.	When were HTN Diagnose	0.021	81
80.	Eggs	0.020	82
81.	Family CVD	0.019	83
82.	Feeling weak,lightheaded/faint	0.017	84
83.	Fruits	0.016	85
84.	Chest Pain	0.015	86
85.	HTN family Specify	0.014	87
86.	Family HTN	0.014	88
87.	Family Diabetes	0.013	89
88.	House	0.011	90
89.	Duration of HTN	0.011	91
90.	BP Frequency check	0.009	92
91.	Diuretics	0.007	93
92.	Duration of CVD	0.007	94
93.	Chest pain	0.006	95
94.	Pulses	0.005	96
95.	Indigestion	0.003	97
96.	CVD family Specify	0.001	98