Original Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2024 (23 Oct);38.123. https://doi.org/10.47176/mjiri.38.123

A Study on Prevalence and Factors Affecting Hypertension in an Iranian Population: Results from the Fasa Cohort Study

Seyede Melika Taheri ghaleno[1], Abdollah Safari[2], Reza Homayounfar[3], Mojtaba Farjam[4], Mehdi Rezaeian[4], Fariba Asadi[1], Fatemeh Masaebi[1], Masoud Salehi[5], Maryam Heydarpour Meymeh[6], Farid Zayeri[7]*

Received: 13 Apr 2024 Published: 23 Oct 2024

Abstract

Background: In recent years, hypertension has been one of the most important noncommunicable diseases worldwide. In this context, identifying the predictors of this disease can help health policymakers to reduce its burden. This study aimed to identify some of the most important influential factors of hypertension and present a model to predict this disease in the data from a large sample cohort study.

Methods: The data set included 10,138 people from the baseline phase of the Fasa cohort study during 2014 and 2016. The main outcome under study was having hypertension in the baseline phase of the study according to self-reports or medical examinations. To identify the related factors of hypertension, logistic regression, classification tree, and random forest models were utilized. Statistical analyses were performed in R.

Results: Among the 10,138 people examined, 2819 (27.8%) had hypertension. In the initial screening, 39 variables were regarded as potential indicators of hypertension. After preliminary analysis, 11 variables were recognized as important predictors based on the importance index: history of cardiovascular disease, cardiac disease, waist circumference to height ratio, body mass index, sex, hypertension in a first-degree relative, weight, fatty liver, cardiac disease in a first-degree relative, diabetes in a first-degree relative, and energy intake. The area under the receiving operating characteristic (ROC) curve for predicting hypertension using logistic regression, classification tree, and random forest models was about 72.8%, 73%, and 87.6%, respectively. Also, the accuracy of these models was 65.2%, 67.4% and 77.8%, respectively.

Conclusion: In general, our findings showed that machine learning-based approaches, such as random forest models, outperformed classical methods, such as logistic regression in predicting hypertension. Regarding the rather high prevalence of hypertension in the population under study, there is an urgent need to pay more attention to its indicators for early diagnosis of the patients and reducing the burden of this silent disease in our country.

Keywords: Hypertension, Related Predictors, Random Forest Model, Fasa Cohort Study

Conflicts of Interest: *None declared*
Funding: *None*

This work has been published under CC BY-NC-SA 4.0 license.
Copyright© Iran University of Medical Sciences

Cite this article as: Taheri ghaleno SM, Safari A, Homayounfar R, Farjam M, Rezaeian M, Asadi F, Masaebi F, Salehi M, Heydarpour Meymeh M, Zayeri F. A Study on Prevalence and Factors Affecting Hypertension in an Iranian Population: Results from the Fasa Cohort Study. *Med J Islam Repub Iran.* 2024 (23 Oct);38:123. https://doi.org/10.47176/mjiri.38.123

Corresponding author: Dr Farid Zayeri, f_zayeri@sbmu.ac.ir

1. Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. School of Mathematics, Statistics, and Computer Science, Faculty of Science, University of Tehran, Iran
3. National Nutrition and Food Technology Research Institute, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4. Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran
5. Nutritional Sciences Research Center, Department of Biostatistics, School of Public Health, Iran University of Medical Sciences, Tehran, Iran
6. Department of English Language, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
7. Proteomics Research Center and Department of Biostatistics, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

↑What is "already known" in this topic:
The prevalence of hypertension in different parts of the world has been investigated by other researchers. Also, some of the important predictors affecting this disease have been previously reported in different populations. In this context, the performance of different discrimination and classification methods has been evaluated in the published literature.

→What this article adds:
The literature review shows limited number of studies on hypertension and conflicting results in identifying significant predictors of this disease in Iran, especially using more complex techniques of machine learning. In the present study, we aimed to identify the prevalence and some of the most important indicators of hypertension using machine learning methods in a large sample population.

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]
[DOI: 10.47176/mjiri.38.123]

Introduction

According to the World Health Organization (WHO) report, hypertension is the leading cause of death worldwide. About 1.28 billion adults aged 30 to 79 years have hypertension, and about two-thirds of them live in low- and middle-income countries. A global target related to noncommunicable diseases is to reduce hypertension by 33% from 2010 to 2030 (1). Many studies have been conducted in Iran to explore the prevalence of hypertension; according to reports, its prevalence in Iran is estimated at 31% for adult men and 27% for adult women (2). The global burden of hypertension is on the rise due to population aging and the increased prevalence of obesity. Hypertension is estimated to affect one-third of the world's population by 2025 (3). The impact of hypertension on young people could be more severe because there is a possibility of more prolonged disability for the rest of their lives (4). According to the American Society of Hypertension (ASH) and the International Society of Hypertension (ISH), people with a systolic blood pressure of ≥140 mm Hg or a diastolic blood pressure of ≥90 mm Hg have hypertension (5).

Although hypertension is asymptomatic in most cases and there is no specific reason for hypertension in 90% to 95% of patients (6), various studies have shown that increasing age, genetics, family history, inappropriate lifestyle (eg, inactivity and unhealthy diet), and overweight are among the principal factors affecting hypertension (1, 7-10). Other studies have also reported that smoking, excessive alcohol consumption, diabetes, stress, income level, central obesity, waist circumference, increased blood sugar, cardiovascular diseases, kidney failure, triglycerides, and blood urea nitrogen also affect hypertension (7-10).

Classification plays a key role in medical sciences and healthcare because it effectively improves the precision of diagnosis and chooses the right treatment (11, 12). Various statistical methods are employed for data analysis in classification problems. One of the oldest of these methods is Fisher's method, which does not require the assumption of normal distribution and makes linear combinations of predictors such that the populations are differentiated as much as possible (13). The model-based methods include logistic regression and machine learning models. Logistic regression is a generalized linear regression model and is a common and classic method for classifying binary outcomes (eg, sick and healthy) (14).

Machine learning models that have recently gained popularity for classification tasks are generally divided into supervised and unsupervised learning categories. In supervised learning methods, for example, decision trees, support vector machines, k-nearest neighbors, and random forests, the classifications are already known, and there is a predetermined target variable. However, in unsupervised learning methods, for example, clustering, there is no presupposition about what category or class the data belong to, and the algorithm performs classification merely based on the nature of the data. Support vector machine is a supervised learning method used for classification or regression. It is effective for solving complex problems with a high-dimensional feature space. Support vector machines create classifiers and, thus, maximize the margin between 2 sets of data. K-nearest neighbors is a supervised learning algorithm; it is nonparametric, that is, it does not require data distribution assumption. This algorithm ranks samples based on their proximity to the training samples in the feature space (15, 16). A classification tree is a supervised learning algorithm. Classification trees are a popular machine learning algorithm due to their ease of understanding and high interpretability, obtained by recursive partitioning and fitting a simple prediction model in each partition (17). In the

tree classification tree, 2 branches are formed from each root node, creating a binary tree. In this tree, the features can be quantitative or categorical. This algorithm is capable of handling missing data and at each node, it selects a variable that maximizes the difference in the distribution of the response variable between the 2 resulting subsets. Random forest is a nonparametric method belonging to the family of ensemble methods that combines several unpruned decision trees for prediction. Several bootstrap samples of data are involved in constructing the random forest, and some input variables are randomly involved in building each tree (15, 16).

Many recent studies have attempted to identify influential factors and predict hypertension. For example, Elshawi et al in 2018 evaluated and compared 5 machine learning methods: LogitBoost, the Bayesian network classifier, the artificial neural network, the support vector machine, and the random tree forest (18). Chang et al in 2019 utilized 4 machine learning methods—support vector machine, decision tree, random forest, and extreme gradient boosting—to predict hypertension (19). In 2020, Nour and Polat employed 4 machine learning methods (C4.5 decision tree classification, random forest, linear discriminant analysis, and linear support vector machine) to detect types of blood pressure based on individual characteristics and information and compared the results with one another (20). Montagna et al, in 2020, adopted 5 machine learning models—logistic regression, decision tree, support vector machine, random forest, and XGBoost—to diagnose hypertension (21). Asadullah et al, in 2023, used 4 machine learning methods—simple Bayes, logistic regression, support vector machine, and random forest—to predict hypertension (22).

Few studies have been conducted on hypertension and the determination of its influential factors in Iran. The results of studies on the identification of important predictors are also contradictory. Therefore, we decided to use the data of the Fasa cohort study to determine the influential factors and check the precision of hypertension prediction by these variables. The goal was to improve planning for preventing, controlling, and identifying hypertension in Iran.

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]
[DOI: 10.47176/mjiri.38.123]

Methods

Study Design and Participants

The Fasa cohort study is a population-based study that has been followed up for 15 years. This study was designed to explore the most common risk factors that expose the residents of the rural area of Sheshdeh (Fasa, Fars Province) to non-communicable diseases. The population comprised 10138 people between the ages of 35 and 70 years. This selection of the target population was appropriate due to the youth of the population and the fact that they had not reached the final stages of non-communicable diseases. In this paper, we used the baseline phase data of the Fasa cohort collected from October 2014 to September 2016. All information was gathered from the participants using written questionnaires and the study consent was obtained from all of the participants in the cohort (23).

Outcome Variable and Predictors

In this study, the response variable was considered as the presence of hypertension. The participants sat on a chair after resting for 30 minutes without engaging in any physical activity for the measurement of their blood pressure. They refrained from eating, drinking (especially tea and coffee), or smoking before the measurement. Since proper posture is crucial for accurate blood pressure readings, the participants comfortably sat on a chair without bending their bodies or tensing their muscles 5 minutes before the measurement, ensuring their feet were placed comfortably on the floor. The arm from which blood pressure was taken was positioned comfortably and without tension on the table, chair handle, or beside the person. The sphygmomanometer was placed on the arm or wrist at the level of the heart. If it was positioned above or below the level of the heart, blood pressure readings would not be accurate. Blood pressure was measured twice from each arm, with a 15-minute interval between readings. The average of these values was recorded as the systolic and diastolic blood pressure (24). When analyzing the data, the response variable was defined as a logical combination obtained from the person's self-reported hypertension diagnosis by a doctor or having a systolic blood pressure of ≥140 mmHg or a diastolic blood pressure of ≥90 mmHg (1).

According to the literature review, a list of 143 potential predictor variables was drawn up. Due to the large number of potential predictors, the univariate logistic regression model was used for their initial screening, and the variables with a *P* value of < 0.2 were included in the multivariate analysis. Finally, 39 predictors were deemed eligible to enter the multivariate analysis. Due to the high number of predictor variables, in the next step, 11 variables were included in the final analysis according to the importance index: a history of cardiovascular disease, cardiac disease, waist circumference to height ratio, body mass index (BMI), sex, hypertension in a first-degree relative, weight, fatty liver, cardiac disease in a first-degree relative, diabetes in a first-degree relative, and energy intake.

Statistical Analysis

First, the logistic regression model was fitted to the data, and then the Classification And Regression Tree (CART) was applied using the *rpart* package of R software (25). The Ctree classification tree was fitted to the data using the *partykit* package of this software (26). Finally, the random forest model was fitted to the data. In the random forest model, the Gini index and accuracy index can be used to determine the importance of the variables. The mean accuracy reduction chart is a method to measure the importance of each variable in the model. This method expresses how much accuracy is reduced by removing a variable from the model. The greater the value of the precision reduction by removing the variable, the greater the importance of that variable for classification compared to the other variables. A decline in the average Gini index is another criterion for assessing the importance of a variable. This index measures how much each variable contributes to the homogeneity of nodes and leaves in the random forest. When the average value of the Gini index reduction is high, the variable is more important in the random forest model (27). All the analyses were performed in R Version 4.2.3.

Results

A total of 10,138 people participated in this research. Seven participants were excluded from the study because they did not complete the questionnaire. Moreover, 48 people were excluded from the statistical analysis because they were more than 70 years old. Finally, after removing the missing data, 10,039 people were included in the analysis, including 4494 (45%) men and 5509 (55%) women (Figure 1). The mean (standard deviation) age of the participants was 48.49 (9.35) years, and the age range was 35 to 70 years. The mean body mass index (BMI) (±SD) was 25.67 (4.83) kg/m^2 in the range of 13.46 to 55.43. The mean (±SD) of systolic and diastolic blood pressure was 111.3 (18.32) and 74.63 (11.91), respectively. The minimum and maximum systolic blood pressure were 51 and 225, while the minimum and maximum systolic blood pressure were 35 and 140, respectively. According to the definition of

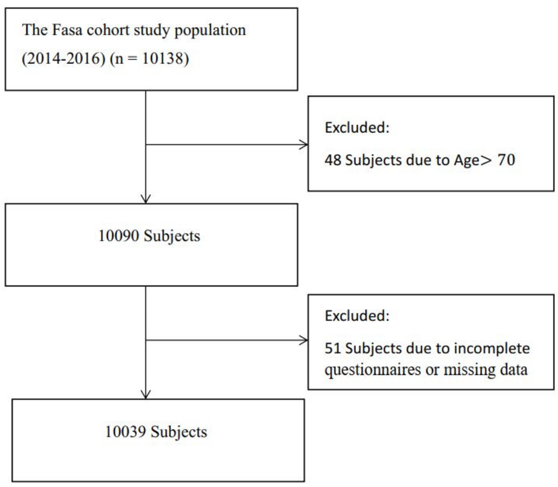


Figure 1. Flowchart of the study population

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]
[DOI: 10.47176/mjiri.38.123]

hypertension, 2819 (27.8%) people had hypertension. Table 1 presents the population's general characteristics according to the prevalence of hypertension.

Next, the logistic regression model was used to identify the variables affecting hypertension. Table 2 lists the results of this modeling for the significant variables. The variables of sex, age, current or previous cardiac disease, fatty liver, and a history of hypertension, and diabetes in a first-degree relative were identified as significant predictors of hypertension. The obtained *P* value in the applied logistic regression model was <0.001. In addition, the AIC and BIC criteria were 10490.9 and 10577.44, respectively.

In the next stage, classification trees with different algorithms and the random forest model were utilized to investigate the relationship between the predictors and hypertension. Among the different algorithms used in fitting the classification tree, the best results belonged to the Ctree algorithm; thus, the results are reported only for this algorithm. Table 3 presents the results of the predictive power indices for the logistic regression model, compared with the results of Ctree and random forest fitting. By comparing these indicators, it can be concluded that the random forest model had a greater prediction precision than logistic regression and Ctree.

Finally, Figure 2 depicts the importance indices for different predictors obtained from the random forest method. In the diagram of the average decline in accuracy, variables history of cardiovascular diseases, BMI, waist circumference to height ratio, and fasting blood sugar are

Table 1. General Characteristics of the Studied Population According to Hypertension

Variable	Category	Hypertension		*P*	Total
		Yes	No		
Sex	Male Female	960 (21.4)* 1854 (33.7)	3534 (78.6) 3655 (66.3)	<0.001	4494 5509
Age (years)	Male Female	49.99 ± 9.63** 51.32 ± 9.42	47.21 ± 9.34 47.92 ± 8.94	<0.001	48.5 ± 9.36 49.2 ± 9.12
Energy intake (kcal/day)	-	2907.36 ± 1157.27	2950.33 ± 1141.94	0.094	2938.21 ± 1146.1
Fasting blood sugar (mg/dL)	-	100.01 ± 37.47	89.68 ± 25.16	<0.001	92.6 ± 29.59
Body mass index (kg/m^2)	-	27.43 ± 4.78	24.97 ± 4.67	<0.001	25.66 ± 4.83
History of cardiac disease	Yes No	689 (59.3) 2125 (24.0)	472 (40.7) 6717 (76.0)	<0.001	1161 8842
Cardiac disease	Yes No	632 (58.8) 2182 (24.4)	443 (41.2) 6746 (75.6)	<0.001	1075 8928
Fatty liver	Yes No	484 (46.6) 2330 (26.0)	554 (53.4) 6635 (74.0)	<0.001	1038 8965
Diabetes in a first-degree relative	Yes No	952 (31.1) 1862 (26.8)	2110 (68.9) 5079 (73.2)	<0.001	3062 6941
Cardiac disease in a first-degree relative	Yes No	1232 (29.1) 1582 (27.4)	3004 (70.9) 4185 (72.6)	0.073	4236 5767
Hypertension in a first-degree relative	Yes No	1751 (33.6) 1063 (22.2)	3456 (66.4) 3733 (77.8)	<0.001	5207 4796

* Number (percentage)
** Mean±standard deviation

Table 2. Results of Fitting the Logistic Regression Model to Determine the Factors Related to Hypertension in the Fasa Cohort Study

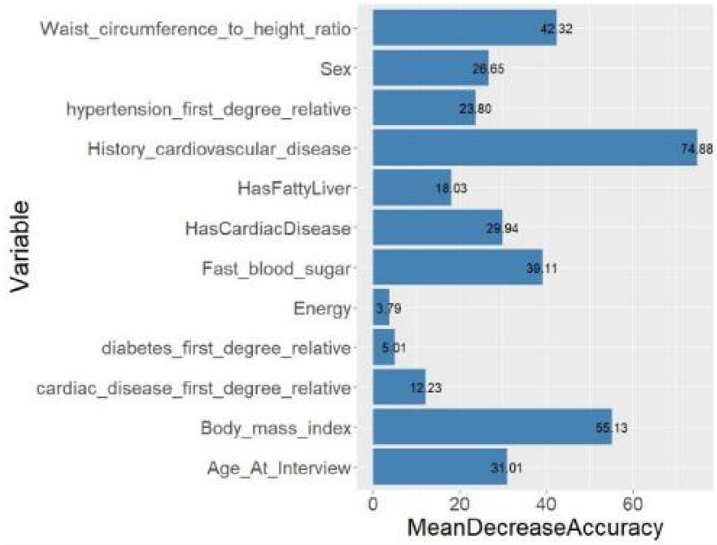
Variable	Category	Odds Ratio	95% confidence interval for the odds ratio
Sex	Male	0.54	(0.49,0.59)
	Female	Reference	
Age	-	1.04	(1.04,1.04)
Cardiac disease	Yes	4.41	(3.86,5.04)
	No	Reference	
Fatty liver	Yes	2.49	(2.18,2.84)
	No		
History of cardiac disease	Yes	4.61	(4.06,5.25)
	No	Reference	
History of hypertension in a first-degree relative	Yes	1.78	(1.63,1.95)
	No	Reference	
History of diabetes in a first-degree relative	Yes	1.23	(1.12,1.35)
	No	Reference	

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]

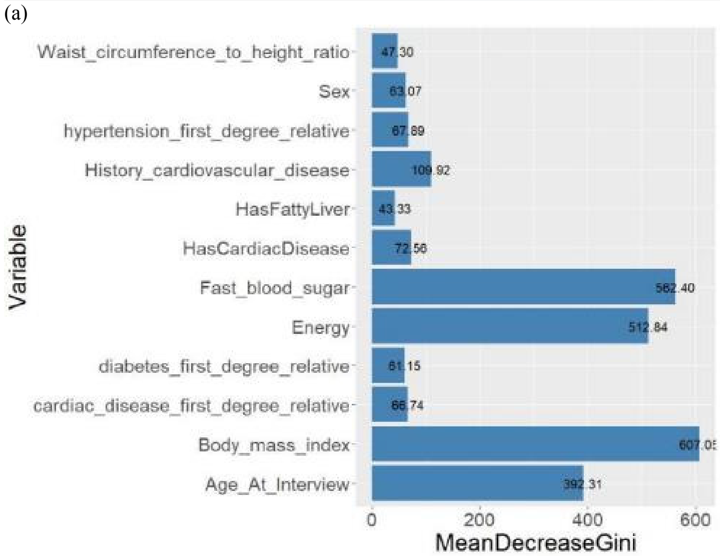
[DOI: 10.47176/mjiri.38.123]

Table 3. Comparison of Hypertension Prediction Precision by Logistic Regression, Ctree, and Random Forest Models

Model	Sensitivity	Specificity	Accuracy	AUC
Logistic regression	68.4	63.9	65.2	72. 8
Ctree	63.5	68.8	67.4	73.0
Random forest	88.9	73.4	77.8	87.6



(a)



(b)

Figure 2. Variable importance indices for potential predictors of hypertension; a: mean decrease accuracy, b: mean decrease Gini

among the most important indicators of hypertension. According to the diagram of the average decline in the Gini index, variables such as BMI, fasting blood sugar, and energy intake are more important than the other variables.

Discussion

Hypertension is a highly prevalent non-communicable disease and a leading cause of mortality worldwide. Despite advances in medical science, it remains a significant health concern (28). Given the high prevalence of this disease and its serious complications, the present study was designed to explore and identify the related risk factors for early diagnosis, and ultimately, prevention and treatment.

Today, machine learning methods are widely used for classification and prediction problems. Classification trees are popular due to their ease of understanding and high interpretability. In this study, various algorithms of classification trees were applied to the data. Due to the potential for a slight change in the data to result in significant changes in model prediction and to mitigate the risk of overfitting, the random forest model was chosen as the primary analytic approach. After identifying the potential predictors and fitting the mentioned models, the random forest model emerged as the best predictive model, demonstrating superior performance due to better indicators (15, 16).

The results of our study revealed that more than a quar-

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]

[DOI: 10.47176/mjiri.38.123]

ter of the studied population had hypertension. A systematic review and meta-analysis by Jafari Oori et al in 2019 in 31 Iranian provinces indicated a 25% prevalence of hypertension (29). Lu et al, in their 2019 study of 11,036 individuals in Nanjing, China, reported a prevalence of 18.5% for hypertension in the studied population (30). Similarly, Princewel et al's study in the rural population of Cameroon also demonstrated that the prevalence of hypertension is 19.8% (31). Furthermore, according to Singh et al's study in India, the prevalence of hypertension was found to be 32.9% (32). It is noteworthy that published studies in this field have reported a wide range of prevalence, varying between 11% and 44% for hypertension across different parts of the world (33). Apparently, the prevalence obtained in the present study also falls within this reported range.

After reviewing the related literature, 143 variables from the Fasa cohort study were considered for this investigation. Subsequently, fitting the classification tree models, 11 important variables were identified based on their importance index as the most significant indicators of hypertension. These variables include a history of cardiovascular disease, cardiac disease, waist circumference to height ratio, BMI, sex, hypertension in a first-degree relative, fatty liver, fasting blood sugar, cardiac disease in a first-degree relative, diabetes in a first-degree relative, and energy intake. Furthermore, in the logistic regression analysis, sex, fatty liver, history of cardiac disease, fasting blood sugar, waist circumference to height ratio, diabetes in a first-degree relative, hypertension in a first-degree relative, and cardiac disease in a first-degree relative were determined to be the significant predictors of hypertension. In the subsequent paragraphs, we will delve into a detailed discussion of these 11 significant variables.

The present study demonstrated a significant relationship between hypertension and a history of hypertension in a first-degree relative, such that the risk of hypertension in people who had a history of hypertension in a first-degree relative was significantly higher than those without such a familial history. These findings are consistent with the results of other studies. For example, a cross-sectional study by Ranasinghe et al in 2015 on 5000 people with an average age of 46 years in Sri Lanka showed that the presence of hypertension in at least 1 family member (parent, sibling, or grandparent) could significantly increase the risk of hypertension in the other family members (34). Furthermore, a cross-sectional study by Alkaabi et al in 2020 on 987 participants revealed that a history of hypertension in the mother can be an influential factor in developing hypertension (35). A systematic review study by Tesfa and Demeke in 2021 showed that people with a family history of hypertension are about 4 times more likely to develop this disease (36). Note that the WHO has also confirmed the effectiveness of genetic factors (10). Accordingly, the presence of hypertension in a first-degree relative is a serious warning for other members of the family.

The present study also found a significant relationship between a family history of diabetes and hypertension, such that people who had a family history of diabetes had a higher likelihood of developing hypertension. A review of similar articles shows that their results are often consistent with the findings of our study. Li et al's study on 684 people in Shanghai, China, found that having a family history of diabetes is one of the most important risk factors for hypertension (37). Since hypertension and diabetes are both aspects of the metabolic syndrome and have similar risk factors, they often co-occur. Accordingly, regular blood pressure monitoring in people with diabetes can lead to timely diagnosis and treatment of the complications of hypertension.

Our study also indicated a significant relationship between sex and hypertension, such that the chance of hypertension was higher in women than in men. Tauhidul Islam et al's study on the adult population of Bangladesh showed that the prevalence of hypertension was higher in women than in men (38). The study by Jafari Oori et al, in 2019, conducted in all regions of Iran, also revealed that the prevalence of hypertension in women is higher than in men (29). However, the results of the present study were contrary to Everett and Zajacova's study in 2015; this study showed that men have higher blood pressure levels than women and a lower level of awareness in this field (39). The study by Santosa et al, in 2020, also indicated that the prevalence of hypertension in men is higher than in women, and they mentioned the effect of self-awareness on controlling hypertension (40). Several studies have been conducted about sex differences and hypertension, reporting that the pathophysiology of hypertension is different between men and women. In women, factors such as reproduction and adverse pregnancy outcomes can be related to hypertension (41). On the other hand, factors such as occupational exposure, smoking, and alcohol consumption in men could affect hypertension in them (42). Overall, it seems that the association between sex and hypertension is still controversial and depends on factors such as the design of the study, the characteristics of the studied population, and the type of statistical analysis. These results highlight the importance of conducting more research on the relationship between hypertension and sex in different populations to correctly identify the effective factors separately in men and women. Based on the findings of this study, a significant relationship was observed between fasting blood sugar and hypertension, such that people with higher blood sugar levels had a higher risk of developing hypertension. According to a study by Liu et al, in 2021, in a rural Chinese population with a 15-year follow-up, high blood sugar levels increased the risk of hypertension in women. This study also reported no relationship between fasting blood sugar and hypertension in men (43). Ahn et al's study on a Korean population showed that the risk of hypertension increases as blood sugar levels rise (44). Accordingly, high blood sugar levels can be another alarm for hypertension. The current study also found a relationship between the waist circumference to height ratio and hypertension. More precisely, the chance of developing hypertension was higher in people with a waist circumference to height ratio of >0.5 (45). Choi et al studied 5000 people between the ages of 40 and 70 years. They found that people who had a higher

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]

[DOI: 10.47176/mjiri.38.123]

waist circumference to height ratio were 4.5 times more at risk of developing hypertension than others (46). A 2020 study by Kawamoto et al also showed that waist circumference to height ratio was a better predictor of hypertension than BMI. According to this study, people who have a higher waist circumference to height ratio are more prone to hypertension (47). Therefore, it seems that with a rise in the waist circumference to height ratio (as a marker for obesity), the chance of developing hypertension significantly increases.

A relationship between BMI and hypertension was also found in the current study; with raising BMI, the chance of developing hypertension also increased. Taleb et al, in 2020, conducted a cross-sectional study on 785 adults. They concluded that the prevalence of hypertension in obese people was higher than in overweight and normal-weight people (48). A study by Hossain et al, in 2019 in Bangladesh, India, and Nepal also showed that reducing the BMI in the studied populations can help decrease the burden of hypertension (49). Li et al examined 684 people and found that the risk of hypertension increases with raising the BMI (50). Pang et al also conducted a cross-sectional study of 45,925 people. They conclude that hypertension increases with a rise in BMI in men and women of all age groups (51). Colin Bell et al conducted a cross-sectional study of people in China, the Philippines, and the United States. They showed that ethnic differences can be associated with the strength of the relationship between BMI and hypertension (52). Landi et al also examined 7907 adults in Italy and found that the association between BMI and systolic and diastolic blood pressure was significant, such that the average systolic and diastolic blood pressure increased with an elevation in BMI (53). Therefore, increasing body mass index is effective in hypertension.

As another finding of this study, a significant relationship was found between the history of cardiac disease and hypertension, such that the risk of developing hypertension was significantly higher in people who had cardiac disease or a history of cardiac disease. Mancia reported a strong relationship between blood pressure and cardiac disease, such that lowering blood pressure was effective in improving cardiovascular disease (54). The study by Fuchs and Whelton in 2020 also showed that controlling hypertension at younger ages will alleviate the consequences of cardiovascular disease at older ages (55). Therefore, there is undoubtedly a significant relationship between hypertension and cardiac disease, but the nature of this association and the cause-effect relationship are highly dependent on the type of disease or outcome being studied.

According to our findings, there was a significant relationship between fatty liver disease and hypertension, such that the presence of fatty liver disease increased the risk of developing hypertension. In 2022, Li et al's study examined the relationship between fatty liver and hypertension. They showed that fatty liver disease is significantly associated with a rise in the incidence of hypertension (56). A study by Michilot et al in 2023 on 234 people found that the prevalence of hypertension is significantly higher in patients with fatty liver (57). Consequently, it seems that fatty liver disease can increase the risk of hypertension.

The current study employed logistic regression, classification trees, and random forest models. The random forest model was recognized as the best model for prediction because it could correctly predict almost 9 out of 10 people. In addition to classical models, other papers used machine learning-based models. For instance, Merajul Islam et al in 2023 utilized 4 models (logistic regression, neural network, random forest, and extreme gradient boosting) to predict hypertension; extreme gradient boosting was recognized as the best model for prediction, with a precision of 88.8% (58). Liu et al, in 2020, also used support vector machine, decision trees, random forest, and extreme gradient boosting models, and extreme gradient boosting was identified as the best model with a precision of 84.9% (59). In 2005, Ture et al compared 3 decision tree models, 4 statistical algorithmic models, and 2 neural network models, all of which investigated the risk of hypertension. Based on the analysis of sensitivity and specificity, the neural network model was identified as the best model (60). Nour and Polat in 2020 adopted the C4.5 decision tree, random forest, linear discriminant analysis (LDA), and linear support vector machine models; the decision tree and random forest were recognized as the best models for hypertension classification with a 99.5 precision (20). Asadullah et al, in 2013, used logistic regression, simple Bayes, support vector machine, and random forest methods to predict hypertension (22). In addition, Montagna et al, in 2023, used the logistic regression model, decision tree, random forest, support vector machine, and XGBoost. The logistic regression model had a sensitivity of 0.66, specificity of 0.63, precision of 0.64, and area under the curve (AUC) of 0.71. The decision tree model had a sensitivity of 0.52, specificity of 0.65, precision of 0.62, and AUC of 0.61. The random forest model had a sensitivity of 0.58, specificity of 0.72, precision of 0.69, and AUC of 0.71 (21). In general, by summarizing the results of the present study and other studies on this subject, it seems that machine learning-based methods—for example, random forest, support vector machine, and boosting—outperform other methods.

Conclusion

In medical literature, hypertension is regarded as one of the most important non-communicable and asymptomatic diseases in many patients. Due to inactivity and an unhealthy urban lifestyle, the burden of hypertension is on the rise annually. This study identified a series of risk factors as a serious warning for hypertension. Given the relatively high prevalence of hypertension in the studied population, identifying effective factors can be significantly beneficial in controlling hypertension and reducing its burden at the national level. The results revealed that machine learning methods, such as classification tree and random forest methods, can detect this disease through its potential indicators (eg, sex, hypertension in a first-degree relative, fasting blood sugar, history of cardiac disease,

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]

[DOI: 10.47176/mjiri.38.123]

BMI, and energy intake) with higher accuracy than classical methods.

Authors' Contributions

The roles of the authors are as follows: Farid Zayeri, Abdollah Safari, and Seyede Melika Taheri Ghaleno presented the ideas of the paper, statistical analyses, and interpretation of results. Reza Homayounfar, Mojtaba Farjam, and Mehdi Rezaeian collected the data. Farid Zayeri, Abdollah Safari, Seyede Melika Taheri Ghaleno, Reza Homayounfar, and Masoud Salehi reviewed and edited the manuscript. Fatemeh Masaebi and Fariba Asadi edited the manuscript. Maryam Heydarpour Meymeh was the translator.

Ethical Considerations

The study was approved by the Ethics Committee of Shahid Beheshti University of Medical Sciences (Ethics code: IR.SBMU.RETECH.REC.1402.508).

Acknowledgment

The proposal of this research project was approved by the Faculty of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran (Project Code: 43007792). The authors would like to express their gratitude to the Research Deputy of this faculty for the facilitation of the process. The authors are also grateful to the Fasa University of Medical Sciences for providing the data.

Conflict of Interests

The authors declare that they have no competing interests.

References

1. https://www.who.int/news-room/fact-sheets/detail/hypertension.
2. Katibeh M, Moghaddam A, Yaseri M, Neupane D, Kallestrup P, Ahmadieh H. Hypertension and associated factors in the Islamic Republic of Iran: a population-based study. East Mediterr Health J. 2020;26(3):304-14.
3. Oliveros E, Patel H, Kyung S, Fugar S, Goldberg A, Madan N, Williams KA. Hypertension in older adults: Assessment, management, and challenges. Clin Cardiol. 2020;43(2):99-107.
4. Hird TR, Zomer E, Owen AJ, Magliano DJ, Liew D, Ademi Z. Productivity Burden of Hypertension in Australia. Hypertension. 2019;73(4):777-84.
5. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J. 2018;39(33):3021-104.
6. Entezari MH, Hadi A, Kafeshani M. Effects of dietary approaches to stop hypertension diet versus usual dietary advice on glycemic indices in women at risk for cardiovascular disease; a randomized controlled clinical trial. Journal of Renal Injury Prevention. 2017;6:205-9.
7. Mirzaei M, Mirzaei M, Gholami S, Abolhosseini H. Prevalence of hypertension and related risk factors in central Iran: Results from Yazd Health Study. ARYA Atheroscler. 2021;17(1):1-9.
8. Esteghamati A, Abbasi M, Alikhani S, Gouya MM, Delavari A, Shishehbor MH, et al. Prevalence, awareness, treatment, and risk factors associated with hypertension in the Iranian population: the national survey of risk factors for noncommunicable diseases of Iran. Am J Hypertens. 2008;21(6):620-6.
9. Bijani M, Parvizi S, Dehghan A, Sedigh-Rahimabadi M, Rostami-Chijan M, Kazemi M, et al. Investigating the prevalence of hypertension and its associated risk factors in a population-based study: Fasa PERSIAN COHORT data. BMC Cardiovasc Disord. 2020;20(1):503.
10. [https://www.who.int/news-room/fact-sheets/detail/hypertension.
11. Balogh EP, Miller BT, Ball JR. Improving diagnosis in health care. 2015.
12. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform. 2002;35(5-6):352-9.
13. Johnson RA, Wichern DW. Applied multivariate statistical analysis. 2002.
14. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. Applied linear statistical models. 1996.
15. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction: Springer; 2009.
16. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: Springer; 2013.
17. Kim A, Kim H. A new classification tree method with interaction detection capability. Computational Statistics & Data Analysis. 2022;165:107324.
18. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. BMC Med Inform Decis Mak. 2019;19(1):146.
19. Chang W, Liu Y, Xiao Y, Yuan X, Xu X, Zhang S, Zhou S. A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. Diagnostics (Basel). 2019;9(4).
20. Nour M, Polat K. Automatic Classification of Hypertension Types Based on Personal Features by Machine Learning Algorithms. Mathematical Problems in Engineering. 2020;2020:1-13.
21. Montagna S, Pengo MF, Ferretti S, Borghi C, Ferri C, Grassi G, et al. Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data. J Med Syst. 2022;47(1):1.
22. Asadullah M, Hossain M, Rahaman S, Amin M, Sharmin M, Sumy A, et al. Evaluation of machine learning techniques for hypertension risk prediction based on medical data in Bangladesh. Indonesian Journal of Electrical Engineering and Computer Science. 2023;31:1794-802.
23. Homayounfar R, Farjam M, Bahramali E, Sharafi M, Poustchi H, Malekzadeh R, et al. Cohort Profile: The Fasa Adults Cohort Study (FACS): a prospective study of non-communicable diseases risks. Int J Epidemiol. 2023;52(3):e172-e8.
24. Dionne JM, Abitbol CL, Flynn JT. Hypertension in infancy: diagnosis, management and outcome. Pediatr Nephrol. 2012;27(1):17-32.
25. Atkinson TTaB. rpart: Recursive Partitioning and Regression Trees. 2023.
26. Zeileis THaA. partykit: A Modular Toolkit for Recursive Partytioning in R. Journal of Machine Learning Research. 2015;16:3905-9.
27. Martinez-Taboada F, Redondo JI. The SIESTA (SEAAV Integrated evaluation sedation tool for anaesthesia) project: Initial development of a multifactorial sedation assessment tool for dogs. PLoS One. 2020;15(4):e0230799.
28. Mirkarimi A, Khoddam H, Vakili MA, Sadeghi MB, Modanloo M. Assessment of dietary adherence in hypertensive patients referred to Shahid Sayyad Shirazi teaching hospital in Gorgan. Journal of Research Development in Nursing and Midwifery. 2016;13(1):39-46.
29. Oori MJ, Mohammadi F, Norozi K, Fallahi-Khoshknab M, Ebadi A, Gheshlagh RG. Prevalence of HTN in Iran: meta-analysis of published studies in 2004-2018. Current hypertension reviews. 2019;15(2):113-22.
30. Lu S, Bao MY, Miao SM, Zhang X, Jia QQ, Jing SQ, et al. Prevalence of hypertension, diabetes, and dyslipidemia, and their additive effects on myocardial infarction and stroke: a cross-sectional study in Nanjing, China. Ann Transl Med. 2019;7(18):436.
31. Princewel F, Cumber SN, Kimbi JA, Nkfusai CN, Keka EI, Viyoff VZ, et al. Prevalence and risk factors associated with hypertension among adults in a rural setting: the case of Ombe, Cameroon. Pan Afr Med J. 2019;34:147.
32. Singh S, Shankar R, Singh GP. Prevalence and Associated Risk Factors of Hypertension: A Cross-Sectional Study in Urban Varanasi. Int J Hypertens. 2017;2017:5491838.
33. Aghaei Meybodi HR, Khashayar P, Rezai Homami M, Heshmat R, Larijani B. Prevalence of hypertension in an Iranian population. Ren Fail. 2014;36(1):87-91.
34. Ranasinghe P, Cooray DN, Jayawardena R, Katulanda P. The influence of family history of hypertension on disease prevalence and associated metabolic risk factors among Sri Lankan adults. BMC Public Health. 2015;15:576.
35. AlKaabi LA, Ahmed LS, Al Attiyah MF, Abdel-Rahman ME. Predicting hypertension using machine learning: Findings from Qatar Biobank Study. PLoS One. 2020;15(10):e0240370.

[Downloaded from mjiri.iums.ac.ir on 2025-06-07]
[DOI: 10.47176/mjiri.38.123]

36. Tesfa E, Demeke D. Prevalence of and risk factors for hypertension in Ethiopia: A systematic review and meta-analysis. Health Sci Rep. 2021;4(3):e372.
37. Li A-l, Peng Q, Shao Y-q, Fang X, Zhang Y-y. The interaction on hypertension between family history and diabetes and other risk factors. Scientific Reports. 2021;11(1):4716.
38. Islam MT, Siraj MS, Hassan MZ, Nayem M, Chandra Nag D, Islam MA, et al. Influence of height on blood pressure and hypertension among Bangladeshi adults. Int J Cardiol Hypertens. 2020;5:100028.
39. Everett B, Zajacova A. Gender differences in hypertension and hypertension awareness among young adults. Biodemography Soc Biol. 2015;61(1):1-17.
40. Santosa A, Zhang Y, Weinehall L, Zhao G, Wang N, Zhao Q, et al. Gender differences and determinants of prevalence, awareness, treatment and control of hypertension among adults in China and Sweden. BMC Public Health. 2020;20(1):1763.
41. Chapman N, Ching SM, Konradi AO, Nuyt AM, Khan T, Twumasi-Ankrah B, et al. Arterial Hypertension in Women: State of the Art and Knowledge Gaps. Hypertension. 2023;80(6):1140-9.
42. Primatesta P, Falaschetti E, Gupta S, Marmot MG, Poulter NR. Association between smoking and blood pressure: evidence from the health survey for England. Hypertension. 2001;37(2):187-93.
43. Liu J, Cheng NN, Zhou ZY, Zhang Y, Yang J, Liu LS, et al. Effect of fasting blood glucose on risk of new-onset hypertension in rural Chinese population: a 15-year follow-up cohort. BMC Cardiovasc Disord. 2021;21(1):531.
44. Ahn SK, Lee JM, Ji SM, Kim KH, Park JH, Hyun MK. Incidence Hypertension and Fasting Blood Glucose from Real-World Data: Retrospective Cohort for 7-Years Follow-Up. Int J Environ Res Public Health. 2021;18(4).
45. Dereje R, Hassen K, Gizaw G. Evaluation of Anthropometric Indices for Screening Hypertension Among Employees of Mizan Tepi University, Southwestern Ethiopia. Integr Blood Press Control. 2021;14:99-111.
46. Choi JR, Koh SB, Choi E. Waist-to-height ratio index for predicting incidences of hypertension: the ARIRANG study. BMC Public Health. 2018;18(1):767.
47. Kawamoto R, Kikuchi A, Akase T, Ninomiya D, Kumagi T. Usefulness of waist-to-height ratio in screening incident hypertension among Japanese community-dwelling middle-aged and elderly individuals. Clin Hypertens. 2020;26:9.
48. Taleb S, Boulaba K, Yousfi A, Taleb N, Difallah B, Negrichi S. Associations between body mass index, waist circumference, waist circumference to-height ratio, and hypertension in an Algerian adult population. Environ Sci Pollut Res Int. 2021;28(34):46514-22.
49. Hossain FB, Adhikary G, Chowdhury AB, Shawon MSR. Association between body mass index (BMI) and hypertension in south Asian population: evidence from nationally-representative surveys. Clin Hypertens. 2019;25:28.
50. Li AL, Peng Q, Shao YQ, Fang X, Zhang YY. The effect of body mass index and its interaction with family history on hypertension: a case-control study. Clin Hypertens. 2019;25:6.
51. Pang W, Sun Z, Zheng L, Li J, Zhang X, Liu S, et al. Body mass index and the prevalence of prehypertension and hypertension in a Chinese rural population. Intern Med. 2008;47(10):893-7.
52. Colin Bell A, Adair LS, Popkin BM. Ethnic differences in the association between body mass index and hypertension. Am J Epidemiol. 2002;155(4):346-53.
53. Landi F, Calvani R, Picca A, Tosato M, Martone AM, Ortolani E, et al. Body Mass Index is Strongly Associated with Hypertension: Results from the Longevity Check-up 7+ Study. Nutrients. 2018;10(12).
54. Mancia G. Blood pressure reduction and cardiovascular outcomes: past, present, and future. Am J Cardiol. 2007;100(3a):3j-9j.
55. Fuchs FD, Whelton PK. High Blood Pressure and Cardiovascular Disease. Hypertension. 2020;75(2):285-92.
56. Li G, Peng Y, Chen Z, Li H, Liu D, Ye X. Bidirectional Association between Hypertension and NAFLD: A Systematic Review and Meta-Analysis of Observational Studies. Int J Endocrinol. 2022;2022:8463640.
57. Chambergo-Michilot D, Rodrigo-Gallardo PK, Huaman MR, Vasquez-Chavesta AZ, Salinas-Sedo G, Toro-Huamanchumo CJ. Hypertension and Histopathology Severity of Non-Alcoholic Fatty Liver Disease Among Adults with Obesity: A Cross-Sectional Study. Clin Exp Gastroenterol. 2023;16:129-36.
58. Islam MM, Alam MJ, Maniruzzaman M, Ahmed N, Ali MS, Rahman MJ, Roy DC. Predicting the risk of hypertension using machine learning algorithms: A cross sectional study in Ethiopia. PLoS One. 2023;18(8):e0289613.
59. Liu Y, Li S, Jiang H, Wang J. Exploring the Relationship between Hypertension and Nutritional Ingredients Intake with Machine Learning. Healthcare Technology Letters. 2020;7.
60. Ture M, Kurt I, Turhan Kurum A, Ozdamar K. Comparing classification techniques for predicting essential hypertension. Expert Systems with Applications. 2005;29(3):583-8.

[DOI: 10.47176/mjiri.38.123] [Downloaded from mjiri.iums.ac.ir on 2025-06-07]