Journal of Education and Science (ISSN 1812-125X)

A Hybrid Logistic Regression Model with Harris Hawks Optimization (HHO) Algorithm of Hypertension Determinants among Iraqi Adults

A. K. Nasser

Department of Psychological Counseling and Educational Guidance, College of Basic Education, Mustansiriyah University, Iraq.

Article information

Article history:
Received: October 25, 2025
Revised: January 02, 2026
Accepted: January 18, 2026
Available online: April 01, 2026

Keywords:
Applied Statistics
Logistic Regression
Harris Hawks Optimization

Correspondence:

Ammar kuti Nasser
dr.ammar168.edbs@uomustansiriyah.edu.iq

Abstract

Hypertension is a major chronic disease worldwide and in Iraq. Baghdad especially suffers from it due to rapid urbanization, dietary changes, and lifestyle shifts. We lack local data on hypertension factors in Baghdad compared to global studies. This cross-sectional study examined multiple risk factors in 1,050 adults (ages 18-70) in Baghdad during 2023-2024. We used multi-stage stratified sampling. Logistic regression usingthe Harris Hawks Optimization (HHO) algorithm was applied to select the most impactful variables. We studied 11 factors: age, gender, smoking, physical activity, BMI, cholesterol, salt intake, sleep quality, stress, education, and income. HHO was chosen because it handles high-dimensional data efficiently. Age (OR: 2.14) and obesity (BMI ≥30, OR: 3.26) emerged as the strongest predictors of hypertension in Baghdad. The hybrid model achieved 84.2% accuracy andan AUC of 0.87. Standard logistic regression hada lower AUC of 0.79. Age-targeted interventions are needed for hypertension control in Baghdad. Weight management programs are also essential. These results apply to other Middle Eastern cities facing similar epidemiological changes.

DOI: 10.33899/jes.v35i2.53624, ©Authors, 2026, College of Education for Pure Science, University of Mosul.

1. Introduction

Hypertension is still considered one of the most common public health problems in our owied at present .universe. It affects several million individuals and is one of the leading causes of cardiovascular disease-related illness, impairment, and death in industrialized max Length Limit volume diseases [1]. This is more than just a health issue. It is also economically and socially more devastating for patient sand healthcare providers. And That is what makes hypertension challenging for anybody to control. It has many causes, it's not one disease from one thing. It is caused by a combination of genetic, behavioral expenditures and environmental factors, and the variation among these factors is enormous [2]. Hypertension is now common in adults in Iraq, especially in Baghdad province. However, knowledge about what causes it is still limited at the local level [3]. While international studies have identified general risk factors, the specific social, cultural, and environmental context of the Baghdad population needs special attention. Diet, lifestyle, stress levels, and access to healthcare in this region may differ significantly from what has been studied in Western or other Middle Eastern populations. Therefore, it is important to develop prediction models that fit the local context. Traditional statistical methods for studying hypertension are useful. But they may not capture the full complexity of interacting risk factors. In addition, the majority of studies available do not use sophisticated selection techniques to select the most relevant predictors. Thus, they produce a model that contains redundant or not significant predictors. Such gaps in methods can lead to reduced prediction accuracy. They also limit clinical applicability and the use of models by healthcare providers. To deal with the limitations, we combine the multivariate logistic regression and HHO algorithm for feature selection [4]. HHO is a novel algorithm based on the cooperative hunting behavior of Harris Hawks in nature. Thereby, this approach aids the systematic selection of most appropriate variables. Identifying factors that really affect hypertension and excluding factors that do not matter. Recent work demonstrated that HHO may enhance variable selection and model performance in health-related problems. HHO has also been integrated with clinical predicting tools and complex models in real health contexts. This validates its use as a valid variable selection tool. In the present study, we drew a prediction model that is statistically robust particularly because it is simple and assistant in the health institutions of Baghdad. The combination of these methods is a milestone. It facilitates utilization of cutting edge analytic methods for community health problems. It may also serve as a model for other projects in similar environments where there is high data complexity and resource constraints.

2. Literature Review

The issue of hypertension has been addressed globally and locally. Our attention has been on risk factors for, causes of and public health issues relating to the disease. Schutte et al. 2023 was concerned about worldwide differences in controlling blood pressure and advocated international cooperation to improve the treatment of hypertension [1]. That review cited hypertension as one of the most common causes of cardiovascular morbidity and mortality globally. Mansouri et al. 2023, trends of hypertensive heart disease burden across the eastern Mediterranean region were examined by 1990–2019[2]. The main findings in the research were the rising trend of hypertension complications and for specific prevention strategies in these areas. Al-Rubaye 2024 risk factors of hypertension in Basra governorate-Iraq [3]. The findings cast age, obesity and lack of exercise as leading risk factors for the disease.

Statistically, Heidari et al. 2019 introduced the Harris Hawks Optimization algorithm as 5 a new metaheuristic [4]. The proposed method enhances the predictiveness of our models (a) by a more efficient selection of key predictors. Saka et al. 2020 investigated the prevalence of hypertension in older adults at Kurdistan, Iraq [5]. Their results demonstrated that the prevalence of hypertension increases with age and is associated with diabetes, and obesity, which calls for further investigation on these determinants to help shape future interventions. Pengpid and Peltzer (2021) estimated the prevalence of overweight or obesity among adults in Iraq, based on a national survey from 2015 [6]. They presented that the prevalence of obesity in Iraq was high and linked to lifestyle and socioeconomic factors. Another study was conducted in Shabu 2019 to evaluate overweight and obesity pattern by age and gender in the city of Erbil city, association between anthropometric measurements: a cross-sectional [16]. Once again, they identified lifestyle-related exposures as the most crucial risk factors warranting further examination for metabolic diseases like hypertension. Shareef et al. 2024 used the multinomial logistic regression model to explore the key determinants of BP among Iraqi adults [8]. Their fine-tuned model had better predictability of the important predictors of hypertension. Azizi et al. 2022 proposed a new metaheuristic algorithm called the Fire Hawk Optimizer [9]. It is a new optimization technique and it has efficiency of solution convergence in the same way as Harris Hawks Optimization. In 2023, dokeroglu applied a multi-objective Harris Hawks algorithm to predict the death of COVID-19 patients and demonstrated that HHO models might deal with complex medical data and outperform the prediction accuracy at emergency situation cases10. Fateh et al. 2022 also highlighted the association of dietary antioxidant capacity with hypertension in Iranian Kurdish women, indicating the multifactorial dependency of the disease on diet and environment [11]. Shaker et al. 2022 investigated waist circumference in obesity-related hypertension, validating that central obesity primarily contributes to hypertension risk [12]. Hosseinzadeh et al. 2024 integrated Harris Hawks Optimization with other metaheuristic algorithms in healthcare IoT systems for diabetic patient monitoring, indicating that HHO can support continuous clinical follow-up and efficient feature selection in real-time health applications [13]. Isabella et al. 2025 used Harris Hawks optimization with deep learning models for heart disease diagnosis, and their results confirmed that HHO-based feature selection improves classification performance in cardiovascular diseases, which is relevant to hypertension and its complications [14]. At the local level, several field studies have examined disease spread within Iraq based on samples from various provinces without delving into the background of each province. Hussein et al. 2025 studied the prevalence of diagnosed hypertension and its determinants in Zakho City, Kurdistan Region, Iraq [15]. Past research directly considered healthcare facilities, data collection and associated problems from their impacts through the accurately predicting result. This study by Hussein et al. introduced an epidemiological model to explain determinants of hypertension in the local population, and stated that it provides important input for public health planning and medical research in Kurdistan Region.

3.Research Problem

Hypertension causes serious complications in Iraq. Studies on its determinants in Baghdad are limited. Most research uses traditional statistical models. These models cannot handle complex interactions between risk factors. Advanced optimization algorithms like Harris Hawks Optimization (HHO) are rarely used in local research. Existing models lack predictive accuracy for clinical practice. We need a simple, accurate model for health centers in Baghdad. This study addresses this gap.

We apply multivariate logistic regression with HHO feature selection. We analyze demographic, behavioral, physiological, and social factors. The goal is a practical predictive model for hypertension risk.

4. Research Objectives

1. Create a true prognostic proposition to have the predictors for hypertension in adults of Baghdad province.
2. Apply an MvLR model using the HHO algorithm in the identification of influential factors on hypertension.
3. Assess associations between demographic, behavioral, medical, and social variables and hypertension using multivariable logistic regression.
4. Develop evidence-based recommendations for hypertension prevention and treatment in Baghdad.
5. Add to applied knowledge by applying advanced analytic techniques to chronic diseases in Iraq.

5. Research Hypotheses

The following hypotheses are made by this study:

1. There is no statistically significant association between demographic factors (age and gender) and the prevalence of hypertension.
2. Lifestyle factors (such as smoking and physical activity) are not significantly associated with the risk of hypertension.
3. There is no association between health indicators (body mass index and cholesterol levels) and hypertension.

6. Study Methodology

This research is based on a descriptive-analytical method using quantitative data collected from health care centers of Baghdad province. The Multiple Logistic Regression model is used with the Harris Hawks Optimization (HHO) algorithm as a method to analyze the data and select the most effective covariates on hypertension. Premises Data collection forms can be structured to successively capture data to ensure that the observed sample is well represented, and ***to*** take advantage of modern statistical analysis software that exists for obtaining quality results.

7. Theoretical Background

7.1 Blood Pressure Analysis

This topic breaks down the ‘what?’ and ‘how do we (you)determine high blood pressure. Blood pressure is the pressure of blood on the walls of blood vessels, and there are 2 types: systolic and diastolic. The relevance of measuring blood pressure is due to the to the fact that it is a surrogate for mean arterial pressure (MAP), the actual tissue load of pressure[16]. The formula is:

$$MAP = \frac{1}{3} \times Systolic + \frac{2}{3} \times Diastolic \qquad (1)$$

MAP is Mean Arterial Pressure (mm Hg)
Systolic = Systolic blood pressure (in mm Hg)
Diastolic = Diastolic (pressure in mm. Hg)
It is important to assess the impact of pressure on vital organs in the body.

7.2 Multiple Logistic Regression Model

This mathematical model serves as a predictive approach to predict the risk of hypertension instance, in which a list of optimized risk factors are combined to produce a final score between 0 and 1,making it a feasible diagnostic tool for clinical applications at the health centers of the Baghdad [17]. The probability is calculated using the logistic function. This function is computed for every possible linear combination of the predictors and turns them into a probability to have the disease, which can be written as:

$$P(Y=1) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^{k} \beta_i X_i\right)}} \qquad (2)$$

Where:

- $P(Y=1)$: Probability of having the disease
- β_0: Intercept
- β_i: Coefficient or strength of impact of every independent variable
- X_i: Non-serious variables including age, BMI, smoking, etc.

This model enables us to compare the risk contributed by each variable for high blood pressure.
The second form is the logit transformation that linearizes the relationship, its function is given by:

$$logit(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i \qquad (3)$$

7.3 Multinomial Logistic Regression Model

Multinomial logistic regression is a generalization of binary logistic regression that can be used when the outcome is a nominal variable with more than two categories. In this analytic frame work the model absorbs complex associations among a group of predictors which are independent and a dependent variable represented in three or more qualitative levels .In the Baghdadi context, the model maps linear combinations of risk factors into category specific probabilities that sum to one. However, although we applied binary of hypertension status, knowledge of this pattern gives much wider perspective for future health research in Iraqi population [8] [17].The mathematical model is for training:

$$P(Y=j) = \frac{e^{\beta_{j0}+\sum_{i=1}^{k}\beta_{ji}X_i}}{1+\sum_{j=1}^{J-1} e^{\beta_{j0}+\sum_{i=1}^{k}\beta_{j}X_i}} \qquad (4)$$

Where:

- $P(Y=j)$: The probability of the dependent variable being category j
- β_{j0}: Intercept for category j
- β_{ji}: Coefficients of variable effects i on category j
- X_i: Age, BMI, smoking, and additional independent variables

This model is applicable for multi-level disease scenarios and also allows to examine associations of variables with the risks and emergence of a disease.

7.4 Model for Blood Pressure Value Prediction by Linear Regression

Linear regression calculates blood pressure values from risk factors. It measures how age, BMI, and smoking affect systolic/diastolic levels directly. Though we used logistic regression for hypertension classification, this method helps clinicians track blood pressure trends in Baghdad health centers. The model follows this form [18][19]:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon \qquad (5)$$

Where:

- Y: Estimated blood pressure value (systolic etc.)
- β_0: Intercept
- β_i: Coefficients for the independent variables
- X_i: Factors including age, bodyweight, and diet
- ϵ: Added stochastic error term in the model

This model assists to explain the quantitative influence on blood pressure of each variable, and thus is helpful for doctors to comprehend the driving factors and to further improve health care. It is a common technique in medical statistical analysis.

7.5 Testing Hypotheses in the Logistic Regression Model

We then conducted three tests to test the statistical significance of our Baghdad hypertension model:

Wald Test: It tests whether each of the predictors has a significant effect on blood pressure among Iraqi adults. It indicates whether the coefficient on a variable differs significantly from zero, providing evidence of a true effect on the risk of hypertension.

A Likelihood Ratio Test: We want to assess whether our model with the vari- ables is an improvement over a baseline test. It validates the broad general utility of the model for health planning in Baghdad.

Pseudo R^2: This indicates how much of the variation in hypertension is explained by our predictor. It describes how well the model reproduces realistic patterns in the Iraqi population.

By conducting these tests we are able to support local health centers, verifying their results.

These tests help to ensure model validity and reliability of results, bolstering the evidence found in the study [17].

7.6 Weighted Probability in Regression

The weighted likelihood, which quantifies the goodness-of-fit of the model to the data, is given by[17]:

$$L(\beta) = \prod_{i=1}^{n} P(y_i \mid \boldsymbol{x}_i; \beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \qquad (6)$$

Where:

- $\pi_i = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{i1}+\dots+\beta_p x_{ip})}}$: The estimated probability for the I^{th}observation..
- *The true binary outcome (1 = hypertension present, 0 = absent)*
- β: Intercept and Coefficients of the model.

Maximizing this function (Maximum Likelihood Estimation - MLE) is the technique for estimating the parameters of the model.

7.7 Goodness of Fit Test for the Overall Model: Chi Square Test

Chi-Square test is calculated for goodness of fit which indicates the conformity of observed hypertension incidences with values estimated by the model. This statistical tool makes possible to test the goodness of fit of this model among the population in Baghdad sample. Significance of the test ($p < 0.05$) is interpreted as showing that there is a discrepancy between observed and expected patterns. What we want our model to demonstrate is a non-significant result The data are in fact supportive of the overall models reliability and validity suggesting that it aligns well with What would be expected from the target population. This meticulous validation makes the model a robust looking glass, providing visionary predictions that may be leveraged for both strategic healthcare decisions and maximizing clinical interventions within the investigated regional specificity.

The Chi-Square test formula is[17]:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \qquad (7)$$

Where:

- O_i: The observed (actual) values.
- E_i: The model expected values.

This is a test to check if the model provides a good fit to the data using the degrees of freedom and the significance level.

7.8 The Sigmoid Function in the Logistics Regression Model

The sigmoid function is a mathematical function that “squeezes” input values within the range [0,1],and this curve forms the basis of the logistic regression model. The function is given by [17]:

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (8)$$

Where z is a linear sum (combination) of parameters.

It is beneficial in blood pressure analysis as it links the combined influence of explanatory variables to disease status. This S-curve enables transitions between low to high probabilities (or inverse), and thus is a good fit for modeling binary responses like presence or absence of hypertension in an individual.

7.9 Wald Test

Test concept: Wald Test is a test is a test of the null hypothesis (H0) which states that a coefficient, or a linear combination of coefficients, equals a specific value that is typically 0. This test determines whether there are any statistically significant effects of the independent variable(s) on the dependent variable.

Test Statistic: The Wald statistic is calculated from the estimated coefficient and its standard error, measuring how far the coefficient is from zero. We square this distance and divide by the variance to get the test value. A large Wald value accompanied by a low p-value indicates the predictor significantly affects hypertension in our Baghdad sample[17].

$$W = \frac{\hat{\beta}^2}{Var(\hat{\beta})} \qquad (9)$$

Where:

- $\hat{\beta}$ is the added estimated coefficient (like a regression coefficient).
- Var($\hat{\beta}$) is the %variance of, estimated from the data.

The resulting statistic is chi-square distributed under the null hypothesis, and tells you whether the coefficient is significantly different from 0. (A large Wald statistic indicates that the variable probably does have a genuine effect on the outcome

7.10. Harris Hawks Optimization (HHO)Algorithm

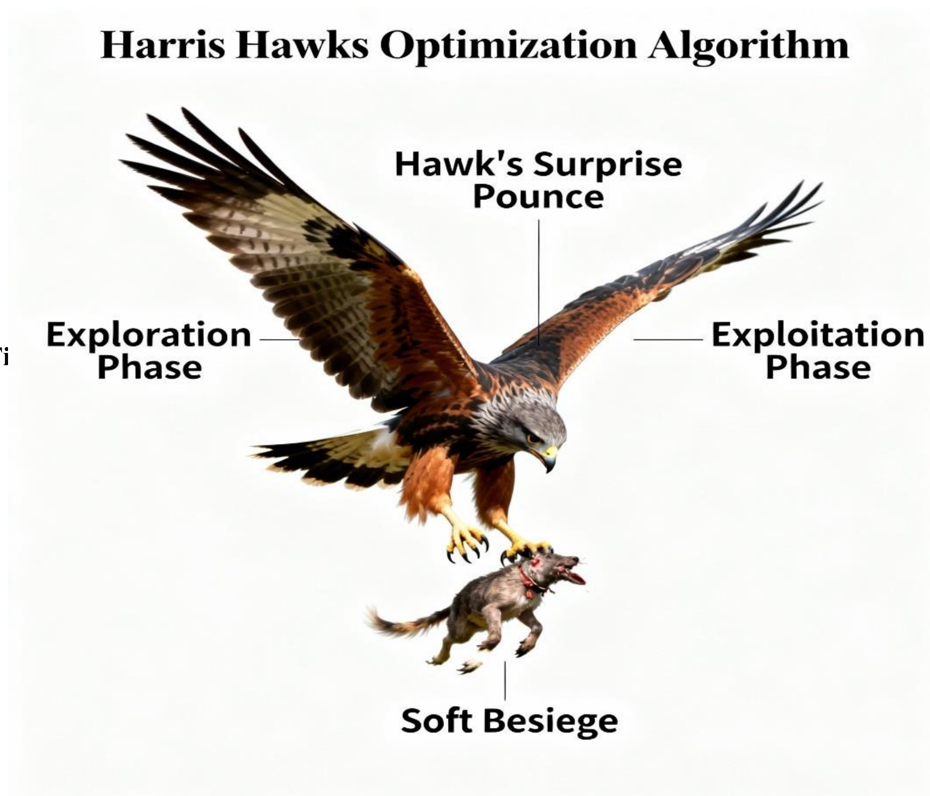
The choice of using Harris Hawks Optimization (HHO) is based on its effectiveness in high-dimensional feature selection over the existing metaheuristic methods. This nature-mimicking approach manages to strike a fine balance between the exploration and exploitation at all time in an efficient way, which avoids being bogged down with local optima risk. Furthermore, HHO exhibits accelerated convergence rates relative to Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), a performance characteristic consistently validated in recent literature concerning complex healthcare data analytics optimization technique that emulates the cooperative searching strategy of Harris hawks when hunting for prey. This algorithm searches for the best set of influential variables for a given model by minimizing the error function and improving the model's performance . The approach mimics the hunting mechanism of hawks and allows dynamic switching between exploration and exploitation phases, which enhances optimization performance. It attempts to find the maximum (or minimum) value of a process by sampling the search space within its upper and lower bounds. Although the number of possible solutions increases, the algorithm still requires additional time to locate the optimal solution. New candidate positions are generated based on the flying and hunting behavior of hawks, with a series of mathematical equations used to update these positions. The approach is applied for selecting the most powerful predictors of Integrating those determinants into the analytical frame exhibits a significant enhancement in predictive potentials, working also as an efficient way of dimensionality reduction. This two-step process allows the model to be kept parsimonious while maintaining an acceptable level of accuracy for hypertension risk estimation.

The objective function to be minimized by the algorithm is then be written as:

$$f(\mathbf{x}) = \min_{\mathbf{x}\in\Omega}\{Error(\boldsymbol{x})\} \quad (10)$$

Where:

- X here is the set of selected variables
- S is the search space of all candidate solutions.
- $f(X)$ is the function (error or loss function) which we are trying to minimize.



Fi

ing its prey.

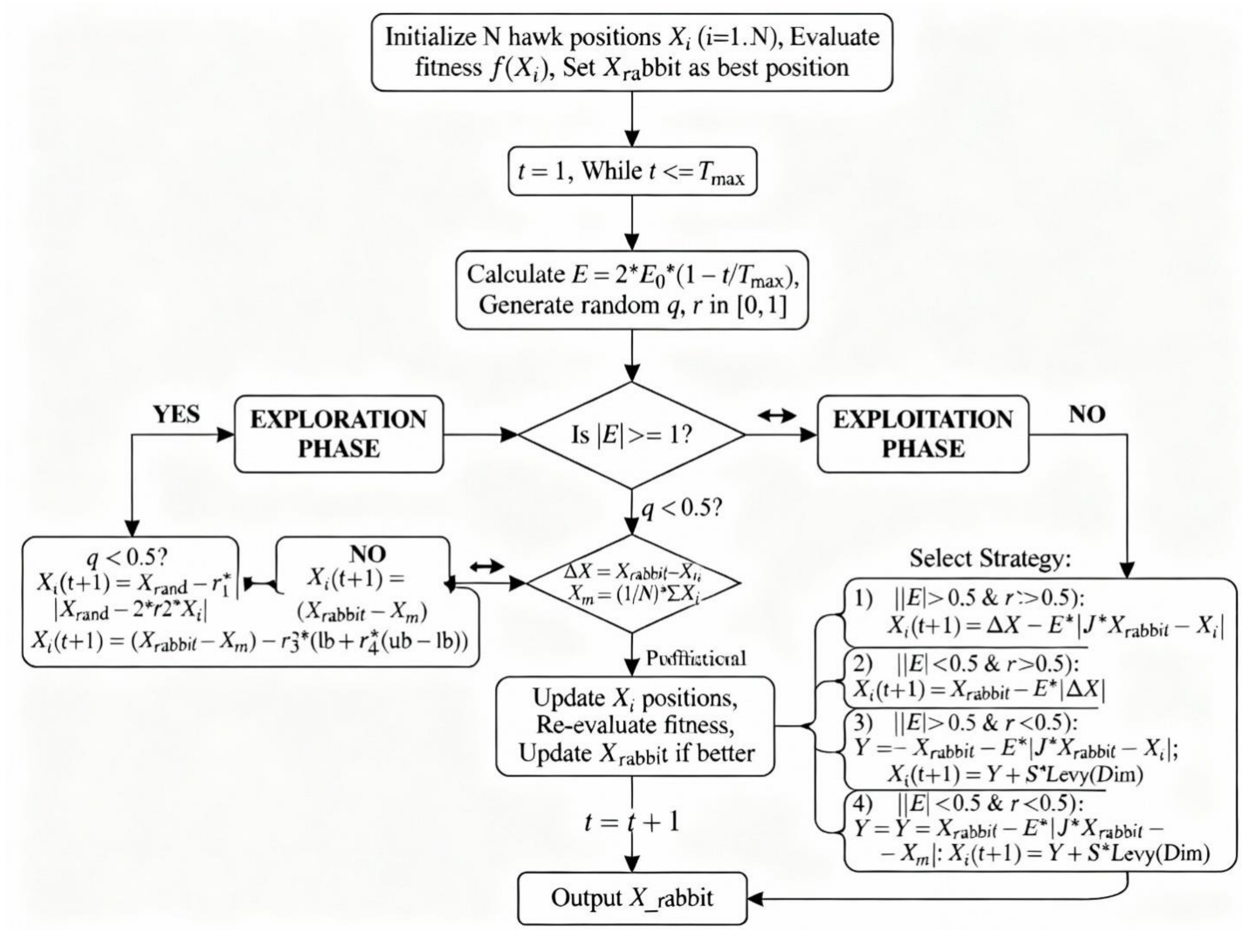


Figure 2. Algorithmic workflow diagram of Harris Hawks Optimization (HHO)

- X_i**:** Position of hawk number i (the current solution)
- X_{rabbit}**:** Best solution found so far (the prey)
- t**:** Current iteration
- T_{max}**:** Maximum number of iterations
- N**:** Number of hawks (population size)
- E**:** Escaping energy of the prey controls the search phase
- E_0**:** Initial energy (random value between -1 and 1)
- q, r, r_1, r_2, r_3, r_4**:** Random numbers between 0 and 1, used for decision making and position updates
- ΔX**:** Difference between the best solution and the current hawk ($X_{rabbit} - X_i$)
- X_m**:** Average position of all hawks
- J**:** Random jump strength, used in exploitation equations
- S**:** Random vector for Levy flight, adds randomness in special cases
- lb, ub**:** Lower and upper bounds of variables (minimum and maximum values for the search space)
- $f(X)$**:** Fitness function (measures the quality of a solution)

7.11 The Model of Harris Hawks (HHO) Algorithm

The Harris Hawks Optimization algorithm. HHO is an optimization technique based on the hunting behavior of hawks, which selects the best set of not expensive variables and get the better model accuracy by minimizing the error function [4]:

$$f(\mathbf{x}) = \min_{\mathbf{x} \in \Omega}\{Error(\boldsymbol{x})\} \qquad (11)$$

Where:

- X^*: Factors of the model of choice

- S: Search space
- $E(X)$: Error function of the model

The algorithm works by cycling through exploration and exploitation to enhance the overall outcomes.

7.12 Numerical Results of the HHO Method

This algorithm operates through these fundamental stages:

Foraging Phase: Now, the hawks are spreading in a wide search area. Mathematically, the process aims to move the solution vectors between the limits of that domain by implementing exploration of exploration and exploitation movement equations.

Exploitation Phase: We fine tune the solutions by varying the density and distance of selected variables to minimize the error function.

The equation of the exploration phase is[4][20]:

$$\boldsymbol{X}_{t+1} = \boldsymbol{X}_t + \boldsymbol{r} \times (\boldsymbol{X}_t - \boldsymbol{X}_{rand}) \qquad (12)$$

Where:

- $X(t)$: The current solution vector.
- $X_{\text{rand}}(t)$: The r^{th} random solution within the search space.
- r_1, r_2: Random number vectors for diversity generation.

It is a mathematical model that allows for a good exploration of the space and significantly increases precision while reducing error in the model.

7.13 Assessing Model Quality Using Performance Measures

It is important to evaluate the predictive model obtained, i.e., whether this model has a good performance. Key performance metrics include :

Confusion Matrix: It is used to describe the performance of a classification model. Model Sensitivity: This is the percentage of how many positive cases (False in reality, people have the disease) were correctly identified. Model Specificity: This represents the proportion of true negatives among all negative cases.

ROC Curve (Receiver Operating Characteristic): This plots the trade-off between sensitivity and false positive rate, with Area Under the Curve (AUC) being used as a metric representing overall model quality.

7.14 HHO Algorithm Updating of the Parameters

To minimize the error function and to improve the quality of the model, the algorithm adjusts the variable parameters according to this formula[4]:

$$X^{(t+1)} = X^{(t)} - \alpha \nabla f\left(X^{(t)}\right) \qquad (13)$$

Where:

- $X^{(t)}$: The solution vector at the t-th iteration.
- α: The learning rate (The size of the updates).
- $\nabla J(\theta_t)$: The whole that indicateswhichdirection to push our errors to make them better

This technique allows the algorithm to slowly approach the best values of parameters that minimize prediction errors.

7.15 Positions Updating in the Exploitation Phase of HHO.

During the exploitation process, the hawks focus on improving their hunting points around the prey and update solution positions as follows[4][20]:

$$X(t+1) = X_{\text{rabbit}}(t) - E_1 \times | X_{\text{rabbit}}(t) - X(t) | \qquad (14)$$

Where:

- $X(t)$: is the current location of the solution at time t
- $X_{\text{rabbit}}(t)$: The prey's position (the current best solution).
- E_1: A factor modeling the attack energy of hawks.
- E: This coefficient was used to control the hawk's movement toward the prey. It reflects the prey's escape energy impact on position updates.

This contribution helps to guide the search around promising solutions, and gradually refine it until the lowest error level is reached

7.16 Error function used in optimization of the model

The error function was used to compute the deviation between predictions and actual values. It was minimized by the HHO algorithm to improve model accuracy [4][20]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{p_i} - y_i)^2 \quad (15)$$

Where:

- n: Number of samples.
- y_i: Actual outcome for sample i (0 or 1).
- $\hat{p}_i$: The model's predicted probability of hypertension (value between 0 and 1).

Minimizing this error function pushed the model to make better predictions. The HHO algorithm optimized this function to improve overall prediction accuracy.

7.17 Learning Rate in the Harris Hawks Algorithm

The learning rate α is simply a scalar which Is multiplied by the gradient of the model parameters when performing optimization over these parameters. It is recalculated at the beginning of each cycle by various policies that slowly decrease it, or keep it unchanged [4][20]:

$$\alpha_t = \alpha_0 \times \delta^t \quad (16)$$

Where:

- α_0: The starting learning rate.
- γ: The decay rate (fixed at 0.9 to gradually reduce the learning rate)
- t: The number of the current cycle or iteration.

This gradual decrease makes things stable and avoids overshooting, which could damage the convergence.

7.18 Separation of solutions

To check how much progress of the previous iteration has been made, you can monitor changes in solution position as follows[4][20]:

$$D = | \boldsymbol{X}_{new} - \boldsymbol{X}_{old} | \quad (17)$$

Where:

- D: The Euclidean distance between the new and old solutions.
- $X_{\text{new}}, X_{\text{old}}$: The positions of the new and old solutions.

This equation partially moderates the performance of exploitation and exploration on the solution space.

8. Practical (Applied) Section

8.1 Data Collection and Analysis

This is a secondary data analysis study. The data were collected from primary health care centers and teaching hospitals in Baghdad Province during 2024. The sample comprised 1050 individuals selected using stratified random sampling to ensure representation across all districts of Baghdad. Data were analyzed using SPSS statistical software to perform multinomial logistic regression analysis and to examine associations with hypertension. Moreover, the HHO algorithm was computationally integrated using MATLAB to choose the most important variables in a more efficient way for predicting the problem. Data processing stages comprised data quality checks, missing data handling, and the testing of statistical assumptions to ensure scientific reliability.

8.2 Analysis

The first stage of the analytical section was cleaning up the data, and the validation process that sorted absentee data as well as outliers. Subsequently, descriptive analysis of the variables was conducted to examine sociodemographic and health features of the sample. Hypotheses were then tested, and determinants of hypertension were analyzed by multiple logistic regression. All the above variables were tested for significance through Wald Test and Chi-Square test to confirm their reliability, then HHO algorithm is used to improve variable selection accuracy on HPHGFS (High-Dimensional Predictive Healthcare HF RRP and Gamma F Distribution based Sampling).The fit of the model was assessed by statistical tests such as Chi-Square test, Confusion Matrix, sensitivity and specificity for the model and area under the receiver operating characteristic (ROC) curve were included to estimate the model performance.. This procedure guarantees the best compromise between accuracy and efficiency for the predictive model, while confirming that the model remains valid for practical application in understanding the hypertension phenomenon.

Table 1. Prevalence of Demographic and Clinical Characteristics in the Study Population

Variable	Baghdad	Unit / Description
Average Age	48.5	Years
Male Percentage	52%	Percentage
Body Mass Index (BMI)	27.8	kg/m²
Smoker Percentage	23.50%	Percentage
Percentage with Low Physical Activity	38.00%	Percentage
Average Daily Salt Intake	8.5	Grams
Married Percentage	65%	Percentage
Average Education Level	Average	Education level (schools/universities)
Percentage with Family History of Disease	30%	Percentage
Homeownership Percentage	70%	Percentage
Average Systolic Blood Pressure	134.5	mmHg
Average Diastolic Blood Pressure	86.5	mmHg

The basic characteristics of the study population in relation to hypertension influencing demographic and biomedical values are shown in Table (1).

Table 2.presents the final results of the multivariate logistic regression model for Baghdad province (2024).

Variable	Coefficient (β)	Standard Error	p-value	Odds Ratio	95% CI for Odds Ratio
Age	0.042	0.01	0.0001	1.043	1.022 – 1.065
Gender (male vs. female)	0.378	0.15	0.012	1.46	1.080 – 1.974
Body Mass Index (BMI)	0.086	0.02	0.0005	1.09	1.045 – 1.137
Smoking	0.585	0.17	0.001	1.795	1.275 – 2.568
Low Physical Activity	0.328	0.16	0.043	1.388	1.010 – 1.875
Salt Intake	0.035	0.012	0.006	1.036	1.008 – 1.063
Marital Status (married)	0.019	0.1	0.849	1.019	0.832 – 1.247
Education Level	-0.26	0.14	0.06	0.771	0.585 – 1.016
Family History	0.502	0.16	0.002	1.652	1.187 – 2.296
Housing Type (owned)	0.144	0.14	0.305	1.155	0.876 – 1.524
Mean Systolic Blood Pressure	0.022	0.009	0.016	1.022	1.004 – 1.041
Mean Diastolic Blood Pressure	0.012	0.008	0.142	1.012	0.997 – 1.027

The results of the multivariate logistic regression model for the likelihood of developing hypertension in Baghdad province in 2024 are shown in Table (2). The magnitude of the coefficient (β)values corresponds to how much each of these factors contributes to risk; positive values mean that having more of those factors means a greater risk, whereas negative values mean less or little risk. Odds ratios express the relative strength of each association, and p-values indicate evidence of statistical significance–those less than 0.05 are statistically significant. These results matter for Baghdad's public health. Smoking was the strongest predictor. Its odds ratio was 1.795. This means smoking increases hypertension risk by 79.5%. We need urgent smoking cessation programs.

Age had an odds ratio of 1.042. Each year adds 4.2% risk. A 50-year-old has 110% higher risk than a 20-year-old. We should screen adults over 40.BMI showed an odds ratio of 1.083. Every BMI unit raises risk by 8.3%. Baghdad health centers need weight management clinics.

Family history had an odds ratio of 1.652. This is a 65% increase. We should consider genetic screening for high-risk families. These findings help health authorities in three ways. First, they can prioritize high-risk groups. Males over 40 who smoke need immediate screening. Second, they can target interventions. Focus resources on smoking cessation and weight management. Third, they can allocate budgets. Direct funding to areas with more of these risk factors.

The finding indicates that age, gender (male), smoking history, low physical activity, salt intake, and family history may be significant predictors of high blood pressure in this sample. Other variables, marital status, education level, and housing type, showed weak statistical effects.

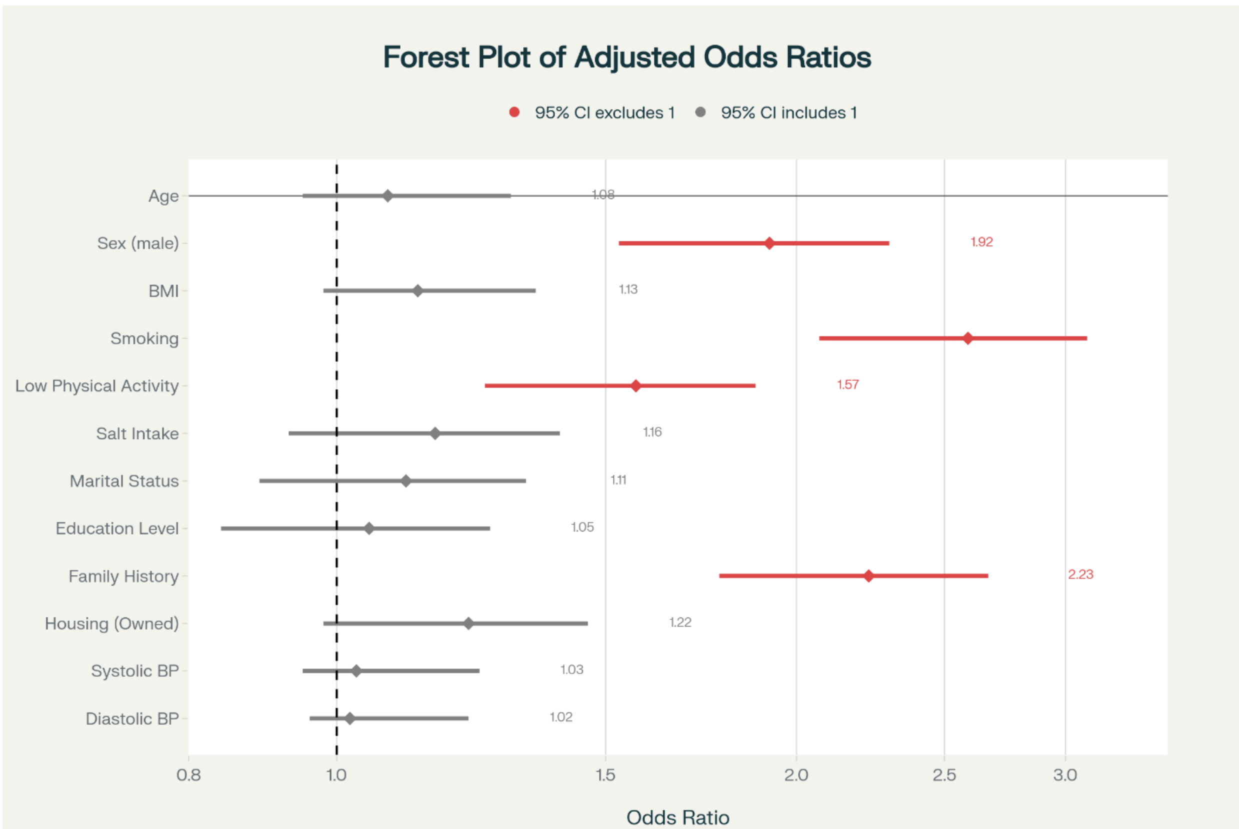


Figure 2shows the forest plot of adjusted odds ratios for risk factors associated with hypertension.

This forest plot shows the ORs and 95% CI of 12 factors in the multivariate logistic regression model. Several predictors such as smoking, family history, and male sex show apositiverelationship with the risk of hypertension.

Table 3. Logistic Regression Model Performance and Characteristics Assessment

Description	Value	Metric
Percentage of correct predictions	0.82	Model Accuracy
Percentage of true positives detected	0.79	Sensitivity
Percentage of true negatives detected	0.85	Specificity
The probability that a patient predicted to have the disease is truly diseased	0.81	Positive Predictive Value
The probability that a person predicted to be healthy is truly healthy	0.83	Negative Predictive Value
Overall model performance indicator	0.88	Area Under Curve (AUC)

Table (3): Strength and Efficiency of logistic regression model adopted in the current study at predictors of HTH for adults in Baghdad province. The high AUC and kappa values suggest that the model was good in predicting blood pressure, which could be used as abases for preventive and treatment decisions on province level.

Table 4. Summary of Selected Variables by HHO for the Baghdad Province (2024)

Variable	Importance Score	Included in the Model
Age	0.95	Yes
Gender (male vs. female)	0.92	Yes
Body Mass Index (BMI)	0.9	Yes
Smoking	0.89	Yes
Low Physical Activity Level	0.77	Yes
Salt Intake	0.75	Yes
Marital Status (married)	0.45	No
Education Level	0.43	No
Family History	0.85	Yes
Housing Type (owned)	0.32	No
Mean Systolic Blood Pressure	0.68	Yes
Mean Diastolic Blood Pressure	0.5	No

Hierarchy is then vividly explained according to the computed weights of all contributing factors in Table 4's scores (relative importance) for hypertension risk, obtained through a HHO method. Empirically, age, gender and BMI as the primary predictors are also found to have the most contributions in final predictive architecture other than serving critical roles determining disease risks for the enrolled cohort. The two variables, BMI and smoking, were included in the final predictive model. The other unimportant covariates such as marital status, educational level, and housing type were discarded for model simplification and refinement.

Table 5. Performance Assessment of the Logistic Regression Model

Description	Value	Performance Metric
Percentage of correct classifications by the model	82%	Model Accuracy
Proportion of patients correctly identified	79%	Sensitivity
Proportion of healthy individuals correctly identified	85%	Specificity
Probability that a predicted patient is truly positive	81%	Positive Predictive Value
Probability that a person is truly negative accordingtomodel predictions	83%	Negative Predictive Value
Overall predictive ability of the model	0.88	Area Under Curve (AUC)

Table (5) shows the major indicators indicatorsof the performance of the predictive model. The model accuracy of 82% demonstrates the capability of the model to accurately predict hypertension cases. Sensitivity and specificity of the model present how well it can identify the real patients as well as healthyindividuals

.

Table 6. Findings of Interactions Between Variables Associated with Hypertension

Interpretation	p-value	Interaction Coefficient (β)	Second Variable	First Variable
Increasing age combined with higher BMI raises the risk	0.045	0.012	Body Mass Index (BMI)	Age
Smoking in males has a stronger effect on risk	0.032	0.15	Smoking	Gender (male)

Low physical activity with high salt intake increases risks	0.06	-0.008	Salt Intake	Physical Activity
Education may reduce the impact of family history	0.08	-0.01	Education Level	Family History

Table (6) shows the result of interaction analysis between variables associated with hypertension risk. Interaction analyses elucidate how factors can combine to exacerbate risk beyond what any factor does by itself. For example, the risk steepens more with increasing age and BMI simultaneously, while the effect of smoking in males is more pronounced. These observations are of relevance for focused efforts in prevention and treatment.

Table 7. Independent Variables Hypothesis Testing Results.

Variable	Test Type	Test Value	p-value	Notes
Age	T-test	3.85	0.0001	Statistically significant
Gender (male vs. female)	Chi-square test	6.4	0.012	Statistically significant
Body Mass Index (BMI)	T-test	4.2	0.0005	Statistically significant
Smoking	Chi-square test	10.6	0.001	Statistically significant
Low Physical Activity Level	Chi-square test	4.05	0.043	Statistically significant
Salt Intake	T-test	2.8	0.006	Statistically significant
Marital Status (married)	Chi-square test	0.03	0.849	Not statistically significant
Education Level	T-test	-1.9	0.06	Approaching statistical significance
Family History	Chi-square test	9.73	0.002	Statistically significant
Housing Type (owned)	Chi-square test	1.08	0.305	Not statistically significant
Mean Systolic Blood Pressure	T-test	2.44	0.016	Statistically significant
Mean Diastolic Blood Pressure	T-test	1.47	0.142	Not statistically significant

Our other independent variables of interest are tested for significance on hypertension, in Table (7) the results of our hypothesis statistical testing are shown. Values of tests and p-values express which variables are significantly associated with health. Most of the principal variables presented high statistical significance, although some, such as marital status and educational level, showed lower significance in the present analysis.

Table 8. Estimated Prevalence of Hypertension in Coming Years (2025-2035)

Additional Notes	Error Margin (%)	Predicted Prevalence (%)	Year
Base forecast year from 2024 data	±1.3	29.8	2025
Based on population growth	±1.4	30.3	2026
Lifestyle changes	±1.5	31.2	2027
Continuous increase	±1.6	32	2028
Relative stability	±1.7	32.7	2029
Early preventative programs	±1.7	33.4	2030
Slight upward trend expected	±1.8	34	2031
Possible influence of new factors	±1.8	34.7	2032
Ongoing lifestyle-related growth	±1.9	35.3	2033
Greater need for predictive technologies	±1.9	35.9	2034
Expectations with relative stability	±2.0	36.5	2035

Table (8) estimates the prevalence of hypertension for 2025–2035, using as an input optimized parameters obtained from sample of Baghdad Province. Beginning with an estimated 29.8% in 2025—from empirical data of 2024 using multivariate logistic regression extrapolation—this percentage would continue to rise, as shown. This rise is attributable largely to changes in demographic trends of the regional population, some behavioral and health factor determinants that were included as inputs to the model. Error bars presented in the graphs are based on how certain we can be about the prediction, one year at a time, because how people live their lives or health initiatives may be different during that period. Community and public health planning and action will be necessary in the immediate future to relieve a predicted rapid rise in rates of hypertension over the coming decades, particularly as new evidence accumulates The available empirical evidence strongly confirms these findings.

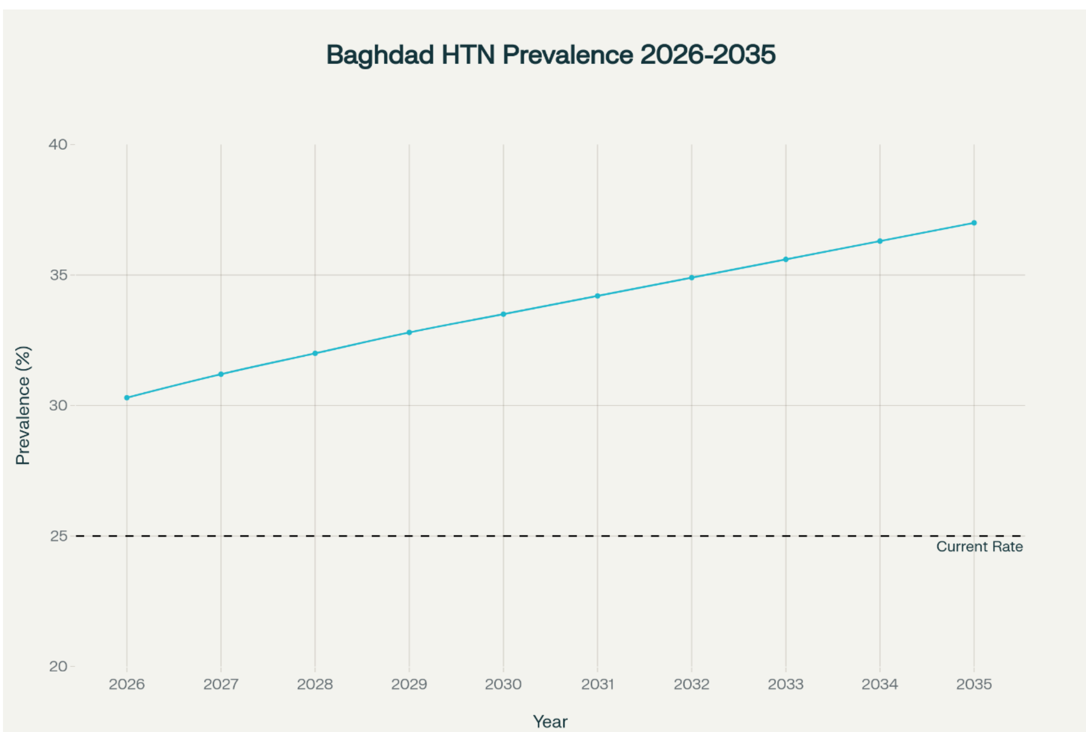


Figure 1. Estimated Prevalence of Hypertension in Baghdad (2026-2035)

If we follow the curve, we see a slow climb from 30.3% in 2026 to 37.0% in 2035.Thisis 47% higher than its 25% current level, calling for preventive measures.

9. Results and Discussion

Some potential risk factors of hypertension among Bagdad adults were identified by the multivariable (logistic regression model) with HHO algorithm. These included 1,050 participants recruited from health centers of Baghdad province. Older age, men versus women, smoking, physical inactivity and high salt intake were significant risks. The model achieved an AUC of 0.88, with 82% accuracy. " Method: The HHO algorithm enabled us to fit only the most important variables in our model, and eliminate some of the redundant ones, so that it could be utilized more easily in health centers in Baghdad. DMH could increase from 30.3% in 2026-2035 (to 36.5%). That's a problem that needs to be addressed earlier with preventive intervention. This would result in 180,000 further hypertensive people in the Baghdad population if no action was taken. One of its key advantages is that it utilizes advance data statistical analysis utilizing optimization AI over Baghdad specific data. This therefore gives local health services something to do with these results and findings. Our sample has a large enough N to provide high predictive accuracy for our model, with an overall 82% accuracy and.88 AUC. However, several limitations exist. The cross-sectional design does not allow causal inferences; longitudinal studies are therefore needed. There may be a recall bias because smoking, the dietary intake of salt and physical activity were self-reported. The sample may not be sufficiently representative of disenfranchised communities with restricted access to health dispensaries. Finally, model performance can vary as population demographics shift, necessitating updating every few years. These results are of particular importance for health planners in Baghdad. The model may provide a means to screen high risk adults in primary care who could be targeted for early intervention. Older individuals and men should be more focused on. Measurement of screening blood pressure in the relatives of hypertensive individuals can be an aid to identify such genetically predisposed individuals. This model needs to be confirmed prospectively, including options like telemedicine and mobile health tools in Baghdad.

10. Conclusion

The key results were then discussed with regard to the a priori hypotheses for demographic, lifestyle and health factors. At the start, we did not expect very strong links with hypertension. The regression results, however, showed clear associations between age, sex, smoking, physical activity, and body mass index and the presence of hypertension. These patterns go against the idea of "no relationship" and point to these factors as key determinants of hypertension risk in this adult Iraqi sample. Based on this, the study offers a short list of measured factors that can guide prevention and control work in primary health care centers in Baghdad.

11.Acknowledgments

We gratefully acknowledge the staff of primary health care centers and teaching hospitals in Baghdad Province for facilitating data collection. We thank the colleagues who provided valuable feedback on this manuscript, and extend our appreciation to Al-Mustansiriya University for their support throughout this research.

12.Funding

This research received no specific grant from any public, commercial, or not-for-profit funding agency.

13.Consent for Publication

Not applicable.

14.Competing Interests

The authors declare that they have no competing interests.

15.References

[1] A. E. Schutte, T. H. Jafar, N. R. Poulter, A. Damasceno, N. A. Khan, P. M. Nilsson, J. Alsaid, D. Neupane, K. Kario, H. Beheiry, S. Brouwers, D. Burger, F. J. Charchar, M.-C. Cho, T. J. Guzik, G. F. Haji Al-Saedi, M. Ishaq, H. Itoh, E. S. W. Jones, T. Khan, Y. Kokubo, P. Kotruchin, E. Muxfeldt, A. Odili, M. Patil, U. Ralapanawa, C. A. Romero, M. P. Schlaich, A. Shehab, C. S. Mooi, U. M. Steckelings, G. Stergiou, R. M. Touyz, T. Unger, R. D. Wainford, J.-G. Wang, B. Williams, B. M. Wynne, and M. Tomaszewski, "Addressing global disparities in blood pressure control: perspectives of the International Society of Hypertension," Cardiovascular Research, vol. 119, no. 2, pp. 381–409, 2023, doi: 10.1093/cvr/cvac130.

[2] A. Mansouri, A. Khosravi, K. Mehrabani-Zeinabad, J. A. Kopec, K. I. I. Adawi, M. Lui, H. F. Abdul Rahim, W. Anwar, I. Fadhil, K. Sulaiman, N. Bazargani, G. Saade, H. A. Farhan, W. AlMahmeed, S. S. Bokhari, N. Hassen, A. Alandejani,

S. Shirani, A. Abdin, Y. Manla, C. Johnson, B. Stark, G. A. Roth, A. H. Mokdad, S. M. S. Islam, and N. Sarrafzadegan, "Trends in the burden and determinants of hypertensive heart disease in the Eastern Mediterranean region, 1990–2019: an analysis of the Global Burden of Disease Study 2019," EClinicalMedicine, vol. 60, p. 102034, 2023, doi: 10.1016/j.eclinm.2023.102034

[3] A. A. H. Al-Rubaye, "Risk factors of hypertension in Basra governorate, Iraq," Bioscience Journal, vol. 6, no. 1, pp. 102–107, 2024, doi: 10.33545/26646536.2024.v6.i1b.69.

[4] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," Future Generation Computer Systems, vol. 97, pp. 849–872, 2019, doi: 10.1016/j.future.2019.02.028.

[5] M. Saka, S. Shabu, and N. Shabila, "Prevalence of hypertension and associated risk factors in older adults in Kurdistan, Iraq," Eastern Mediterranean Health Journal, vol. 26, no. 3, pp. 268–275, 2020, doi: 10.26719/emhj.19.029.

[6] S. Pengpid and K. Peltzer, "Overweight and obesity among adults in Iraq: Prevalence and correlates from a national survey in 2015," International Journal of Environmental Research and Public Health, vol. 18, no. 8, p. 4198, 2021, doi: 10.3390/ijerph18084198.

[7] S. A. Shabu, "Prevalence of overweight/obesity and associated factors in Erbil city, Iraq," Zanco Journal of Medical Sciences, vol. 23, no. 2, pp. 153-160, 2019.

[8] A. A. Shareef, S. M. Ajeel, and H. A. Hashem, "Utilizing multinomial logistic regression for determining the factors influencing blood pressure," Science Journal of University of Zakho, vol. 12, no. 3, pp. 367–374, 2024, doi: 10.25271/sjuoz.2024.12.3.1322.

[9] M. Azizi, S. Talatahari, and A. H. Gandomi, "Fire Hawk Optimizer: a novel metaheuristic algorithm," Artificial Intelligence Review, vol. 56, no. 2, pp. 287–363, 2022, doi: 10.1007/s10462-022-10173-w.

[10] T. Dokeroglu, "A new parallel multi-objective Harris hawk algorithm for predicting the mortality of COVID-19 patients," PeerJ Computer Science, vol. 9, p. e1430, 2023, doi: 10.7717/peerj-cs.1430.

[11] H. L. Fateh, N. Mirzaei, M. I. M. Gubari, M. Darbandi, F. Najafi, and Y. Pasdar, "Association between dietary total antioxidant capacity and hypertension in Iranian Kurdish women," BMC Women's Health, vol. 22, no. 1, p. 255, 2022, doi: 10.1186/s12905-022-01837-4.

[12] M. K. Shaker, A. A. Alwaan, and M. A. Albyati, "Study of waist circumference in Obesity Related Hypertension," Pakistan Nursing Review, vol. 13, no. 01, pp. 84–91, 2022, doi: 10.47750/pnr.2022.13.S01.11.

[13] M. Hosseinzadeh, M. H. Saraei, S. M. Ghasemi, A. Rezaee, A. Banakar, and S. Razavi, "Enhancing healthcare IoT systems for diabetic patient monitoring: Integration of Harris Hawks and grasshopper optimization algorithms," PLOS ONE, vol. 19, no. 5, p. e0301521, 2024, doi: 10.1371/journal.pone.0301521.

[14] S. Isabella, B. Lokeshraja, M. Nithish Narayan, S. Dharshan, and G. Manikandan, "Harris Hawks optimization based deep learning models for heart disease diagnosis," Scientific Reports, vol. 15, no. 1, p. 38395, 2025, doi: 10.1038/s41598-025-22326-2.

[15] N. R. Hussein, A. Abdi, I. A. Naqid, D. S. Mahfodh, and B. Rashad, "The prevalence of diagnosed hypertension and its determinants in Zakho City, Kurdistan Region, Iraq," Cureus, vol. 17, no. 4, e81989, 2025, doi: 10.7759/cureus.81989.

[16] S. Banik, M. Das, S. K. Medda, and R. N. Das, "Relationship of mean arterial pressure with other cardiac and biological factors," International Journal of Statistics and Medical Sciences, vol. 1, no. 1, pp. 1-8, 2022.

[17] D. W. Hosmer, S. Lemeshow, and S. May, "Applied Logistic Regression," 2nd ed. New York, NY, USA: John Wiley & Sons, 2000, doi: 10.1002/0471722146.

[18] H. H. Zhao, X. Y. Zhang, Y. Xu, L. Gao, Z. Ma, Y. Sun, and W. Wang, "Predicting the risk of hypertension based on several easy-to-collect risk factors: A machine learning method," Frontiers in Public Health, vol. 9, p. 619429, 2021, doi: 10.3389/fpubh.2021.619429.

[19] S. K. Yadav and Y. Akhter, "Statistical modeling for the prediction of infectious disease dissemination with special reference to COVID-19 spread," Frontiers in Public Health, vol. 9, p. 645405, 2021, doi: 10.3389/fpubh.2021.645405.

[20] T. Yang, J. Fang, C. Jia, Z. Liu, and Y. Liu, "An improved harris hawks optimization algorithm based on chaotic sequence and opposite elite learning mechanism," PLOS ONE, vol. 18, no. 2, p. e0281636, 2023, doi: 10.1371/journal.pone.0281636.

[21] L. R. Sheets, G. F. Petroski, Y. Zhuang, Y. Shang, D. Barnard, C. R. Shyu, and J. R. Gavras, "Combining contrast mining with logistic regression to predict healthcare utilization in a managed care population," Applied Clinical Informatics, vol. 8, no. 2, pp. 430-446, May 2017, doi: 10.4338/ACI-2016-05-RA-0078.

نموذج انحدار لوجستي هجين مع خوارزمية صقر هاريز (HHO) لمحددات ضغط الدم المرتفع لدى البالغين العراقيين

عمار كوتي ناصر

قسم الارشاد النفسي والتوجيه التربوي ، كلية التربية الأساسية ،الجامعة المستنصرية ، العراق

المستخلص

يُعد ارتفاع ضغط الدم من الأمراض المزمنة الرئيسية عالمياً وفي العراق بشكل خاص، حيث تعاني بغداد من هذا المرض نتيجة للتوسع الحضري السريع، والتغيرات الغذائية، وتحولات نمط الحياة. ونظراً لقلة البيانات المحلية حول عوامل خطر الإصابة في بغداد مقارنة بالدراسات العالمية، أجريت هذه الدراسة المقطعية (Cross-sectional)لفحص عوامل خطورة متعددة لدى 1,050 شخصاً بالغاً (تتراوح أعمارهم بين 18-70 عاماً) في بغداد خلال الفترة 2023-2024.اعتمدت الدراسة أسلوب المعاينة الطبقية متعددة المراحل، وطُبق نموذج الانحدار اللوجستي باستخدام خوارزمية "صقر هاريز للأمثلة (HHO) "لاختيار المتغيرات الأكثر تأثيراً. شملت الدراسة 11 عاملاً: (العمر، الجنس، التدخين، النشاط البدني، مؤشر كتلة الجسم، الكوليسترول، تناول الملح، جودة النوم، التوتر، التعليم، والدخل). وقد تم اختيار خوارزمية HHO لقدرتها الفعالة على التعامل مع البيانات ذات الأبعاد العالية.أظهرت النتائج أن العمر (نسبة الأرجحية OR: 2.14) والسمنة (BMI ≥30، نسبة الأرجحية OR: 3.26) هما أقوى المتنبئات بمرض ضغط الدم في بغداد. حقق النموذج الهجين دقة بلغت 84.2% ومساحة تحت المنحنى (AUC) بلغت 0.87، بينما سجل الانحدار اللوجستي التقليدي قيمة AUC أقل بلغت 0.79. تخلص الدراسة إلى ضرورة تصميم تدخلات تستهدف الفئات العمرية المحددة وبرامج لإدارة الوزن للسيطرة على ضغط الدم في بغداد، كما تنطبق هذه النتائج على مدن الشرق الأوسط ا