ORIGINAL ARTICLE

Unveil the Features Influencing Hypertension Adults in Malaysia using Machine Learning Models

Ridwan Sanaudi[1], Zainul Azhar Zakaria[2], Aiman Adlan Khairulisam[3], Nurain Ibrahim[3,4], Ahmad Zia Ul-Saufie[3]

[1] Sector of Biostatistics and Data Repository, National Institutes of Health, Ministry of Health Malaysia, 40100 Shah Alam, Selangor, Malaysia
[2] Foundation in Medical Sciences, Faculty of Medicine, Royal College of Medicine Perak, Universiti Kuala Lumpur, 30450 Ipoh, Perak, Malaysia
[3] School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[4] Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

ABSTRACT

Introduction: The number of people affected by hypertension is staggering, with an estimated one billion people living with the disease worldwide. It has been shown that machine learning (ML) models surpass clinical risk; nevertheless, there isn't much research using ML to predict hypertension in Malaysia. **Materials and methods:** A study is being conducted using ML analyses to predict hypertension using secondary data from population-based surveys, such as the National Health & Morbidity Survey (NHMS) 2015. The dependent or target variable was hypertension status and 24 features. Three standard ML-based classifiers, which are logistic regression (LR), decision tree (DT) and artificial neural network (ANN), were used to predict hypertension and the associated factors that influence hypertension were obtained from filter-based feature selection, which are feature weight by information gain, feature weight by information gain ratio and feature weight by correlation. **Results:** Out of 11,520 respondents, 4,175 (36.24%) adults had hypertension. LR is the best model to predict hypertension since LR has the highest accuracy (76.73%) compared to DT and ANN (73.02%). In terms of odd ratio explanation, a person who does not have diabetes mellitus is 2.05 odds likely to have hypertension, and a person who does not have hypercholesterol has 1.67 odds of having hypertension, and with an increase in the age of adults, 6.0% are less likely to have hypertension. **Conclusion:** From LR model, the essential features that influence hypertension in adults were diabetes mellitus, hypercholesterolemia status, age, waist circumference, marital status, occupation, education, and total household income.
Malaysian Journal of Medicine and Health Sciences (2024) 20(6): 167-174. doi:10.47836/mjmhs20.6.22

Keywords: Artificial neural network, Decision tree, Hypertension, Logistic regression, Machine learning, Prediction

Corresponding Author:
Nurain Ibrahim, PhD
Email: nurain@tmsk.uitm.edu.my
Tel : +6011-39328310

INTRODUCTION

The disorder known as hypertension, or high blood pressure, is characterized by an excessively high blood pressure against the arterial walls. Two numbers are used to represent blood pressure. The first (systolic) number represents the blood vessel pressure during a heartbeat or contraction. The second (diastolic) number indicates the pressure in the arteries during the heart's resting heartbeat. If a person's diastolic blood pressure is 90 mmHg or higher on two different occasions, and their systolic blood pressure is 140 mmHg or higher, they are diagnosed with hypertension (1). Risk assessment for hypertension, or high blood pressure, can be quite complex and depends on a variety of factors. These can include genetics, lifestyle choices, and environmental conditions that can contribute to an increase in blood pressure readings. The World Health Organization (WHO) reported that high blood pressure is a leading cause of death worldwide, responsible for one in every eight deaths (2). It is considered the third leading cause of death globally. The number of people affected by hypertension is staggering, with an estimated one billion people worldwide living with the condition.

Malaysia's Ministry of Health (MOH) conducted the National Health and Morbidity Survey (NHMS). The survey found that the overall prevalence of hypertension, or high blood pressure, among adults aged 18 and above was 30.0% (3). The NHMS provides essential information about the population's health status and guides health policies and programs in the country. Hypertension is a burden and affects the health system in Malaysia, specifically in low-income households (4). A previous study suggested that web-based educational tools help spread hypertension-related information to increase hypertension awareness (5).

However, most studies employ various instruments to quantify risk variables, have small sample sizes, and lack community representation. One specific example of a risk prediction model is the Framingham Risk Score (FRS), commonly used to predict the risk of coronary heart disease. FRS has several drawbacks, including underrepresented ethnic diversity and non-representative populations (6–8). Many studies were conducted to develop the FRS and other models on populations that may not represent the general population. It makes it difficult to predict risk for different populations or ethnicities accurately. Besides that, many of the studies used to develop these models have limited ethnic diversity, making it difficult to accurately predict risk for people of different ethnicities.

A recent study found and investigated machine learning (ML) strategies for detecting hypertension. It was noted that there was a shortage of research integrating clinical and sociodemographic data, which might improve model performance (9). XGboost method is one of the machine learning methods, and it was used to predict blood pressure control for hypertension patients. However, that study did not compare other ML methods (10). A study in China compared two ML methods, namely, the back-propagation neural network and random forest, to predict cardiovascular disease, and logistic regression was used to identify significant factors in cardiovascular disease (11). Therefore, this study aims to use ML approaches to probe features associated with hypertension diagnosis and compare the model's performances to predict hypertension in Malaysia. There is a strong connection between this research and Sustainable Development Goal No. 3: Good Health and Well-being. Hence, this study is important and necessary to reduce hypertension cases in Malaysia since it identifies the significant features so that Malaysians may increase their awareness of hypertension and avoid taking foods that can contribute to it. This statement is supported by the previous study findings, in which the risk dietary pattern with the most significant frequency of hypertension was found to be "seafood and alcohol" (12). Studies on predicting hypertension using ML methods focus on newborns (13) and pregnant women (14). The rest of the paper is organized as follows: Section 2 explains the material and method. Section 3 presents the results and the discussion of the experiment. Section 4 constitutes the conclusion.

MATERIALS AND METHODS

This study follows the six steps of Cross Industry Standard Process for Data Mining (CRISP-DM), including business understanding, data understanding, data preparation, modelling, evaluation and deployment. The first step is business understanding. The pace at which hypertension is occurring is concerning. Therefore, this research aims to investigate parameters related to the diagnosis of hypertension using machine learning techniques and then compare the models' output to forecast hypertension in Malaysia. The solution to this objective feature selection weights are ranked to determine the most relevant features that influence hypertension. Additionally, three data mining techniques (Logistic Regression, Decision Tree, and Artificial Neural Network) are compared based on their performance to select the best predictive model for hypertension prediction. The research flowchart is shown in Figure 1.

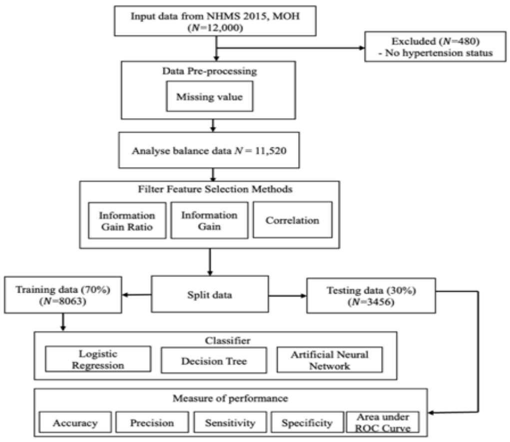

Figure 1: Research flowchart.

The second step is data understanding. Secondary data was used as it is easy to access and less time-consuming. The dataset was obtained from the National Health & Morbidity Survey (NHMS) 2015, released in June 2016. The original dataset is obtained from the official site of the Institute of Public Health under the Ministry of Health Malaysia. The survey has an exclusion criterion, with the respondents younger than 18 years old excluded from the data. This dataset consists of 12000 observations of 25 features. The variables can be categorized into two groups: continuous and categorical features. Of the 25 features, eight continuous features and 16 categorical features exist. The target variable, Γγ in this research, would be the Hypertension Status (Hyp_status=Yes, No), with the remaining 24 features will be the independent variables, xi. It is also to be noted that the target variable Hyp_status has 481 missing values. Table I displays the description of the hypertension data.

Table I: Data Description

Feature	Description
Hyp_Status	Hypertension status 0=No, 1=Yes
Strata	Strata 1=Urban, 2=Rural
State	State origin 1=Johor, 2=Kelantan, 3=Melaka, 4=Negeri Sembilan, 5=Pahang, 6=Penang, 7=Perak, 8=Perlis, 9=Selangor, 10=Terengganu, 11=Sabah, 12=Sarawak, 13=WPKL, 14=Labuan, 15=WP Putrajaya
Sex	Gender 1=Male, 2=Female
Ethnicity	Ethnic group 1=Malay, 2=Chinese, 3=Indian, 4=Other
Occupation	Occupation 1=Government / Semi-Government, 2=Private Employee, 3=Self Employee, 4=Unpaid Worker /Homemaker, 5=Retiree
Education	Education 1=No formal education, 2=Primary, 3=Secondary, 4=Tertiary, 5=Unclassified
Marital_status	Marital status 1=Never married, 2=Married, 3=Widow / Widower /Divorce
Citizen	Citizenship 1=Malaysian, 2=non-Malaysian
Abd_obesity	Abdominal Obesity Status 0=No, 1=Yes
DM_status	Diabetes Melitus Status 0=No, 1=Yes
Hyperchol_status	Hypercholesterolaemia status 0=No, 1=Yes
PA_status	Physical activity status 1=Active, 2=Inactive
Smoking_status	Smoking status 1=Current tobacco smoker, 2=Current non-smoker
Anemia_Status	Anaemia status 0=non-anaemia, 1=anaemia
heavyedrinker_status	Heavy episodic drinking status 0=No, 1=Yes
Area	Klang Valley Area 1=Others, 2=Klang Valley
Age	Age
Weight	Weight
Height	Height
Total_Ind_Income	Total individual income
Total_HH_Income	Total household income
Waist_circumference	Waist circumference
BMI	Body Mass Index
Average_Intake_FV	Average intake per day of fruits and vegetables

The third step is data preparation. The dataset is explored in terms of data preparation, and the data quality is checked, such as missing values, outliers, and recording errors. The next step would be to run a descriptive analysis of the features to better understand the dataset's shape or distribution. Some of the most prepared graphical methods are clustered bar charts and pie charts. The data were split into two partitions: 70% for training data and 30% for testing data.

Once the data were well prepared, the feature selection procedures were applied to the dataset to find the best set of features. The feature weights used are filter features weight information gain ratio, filter features weight information gain and filter feature weight correlation. The fourth step is modelling. In the case of classification algorithms, the modelling techniques used are Logistic Regression, Decision Tree, and Artificial Neural Network. Thus, these three models will be compared to develop the best prediction model for predicting the features that influence hypertension in adults in Malaysia.

Logistic Regression

Various features may be used with logistic regression to help forecast the likelihood of reaching a certain level. Logistic regression was used to determine the essential features for categorizing hypertension status. In this study, strata, state, sex, ethnicity, occupation, education, marital status, citizen, abdominal obesity status, diabetes Miletus status, Hypercholesterolaemia status, physical activity status, smoking status, anaemia status, heavy drinker episodic drinking status, Klang Valley area, age, weight, height, total individual income, total household income, waist circumference, body mass index and average intake per day of fruits and vegetables will be the features and hypertension status will acts as target variable (15–17). The logistic regression equation is shown in Equation (1) as follows:

$$\ln\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_i X_i \qquad (1)$$

In this case, logistic regression will be used to classify the observation as having hypertension or no hypertension.

Decision Tree

Decision Tree is a classification algorithm that finds a hierarchical structure explaining the input space areas corresponding to different outcomes or classifications. A decision tree model comprises a collection of guidelines for separating large, heterogeneous populations according to specific splitting criteria or targets into smaller, more homogeneous groupings. There are three standard splitting criteria for decision tree gini index, gain ratio and information gain. The splitting characteristic with the lowest entropy value is selected via the information gain technique, yielding the most information gain. The Decision Tree's splitting feature is determined by calculating the information gain for each characteristic first then choosing the one that maximizes the information gain (18). Information Gain (IG) measures the reduction in entropy or impurity by partitioning the data according to a particular attribute X and is given by the formula displayed in Equation (2):

$$IG(D,X) = E(D) - \sum_{v \in Values(x)} \frac{|D_V|}{|D|} \cdot E(D_V), \qquad (2)$$

where, $\Gamma_{E(D)}$ is the entropy of the entire dataset, Γ_{DV} is the subset of D for which attribute X has value v, $\Gamma_{|DV|}$ is the number of elements in and *values (X)* is the set of all possible values that the attribute X can take. This equation subtracts the weighted entropies of each partition (based on an attribute Γ_X) from the original entropy of the dataset D, thus providing the amount of entropy reduced due to this partitioning.

Artificial Neural Network

Artificial Neural Network (ANN) are among the oldest and most intensely studied Machine Learning approaches. The inspiration for ANN comes from the biological neural network and mimics the learning process of animals and humans. ANN is a classification algorithm that uses a feed-forward neural network trained by a backpropagation algorithm or a multi-layered perceptron. This operator's usual sigmoid function is used as the activation function. This perceptron will predict the classifications and adjust the weights and threshold values to improve the predictions. Three or more linked layers form an artificial neural network. The first layer is made up of input neurons. Deeper layers get input from these neurons, and these layers, in turn, provide the ultimate output layer of the data. The input layer aims to transmit input data comparably without changing it. The hidden layer uses the activation function to compute the linear combination of data supplied from the input layer. The hidden layer's data is processed by the output layer, which then outputs the results of its computations in a format that resembles the hidden layer's.

The fifth step is evaluation. The models were assessed using accuracy, specificity, sensitivity, precision, and area under ROC curve (AUROC). Accuracy is the ratio of correctly identified subjects in a pool of subjects (19,20). Sensitivity relates to the ratio of correctly identified subjects by test against all positive subjects. Specificity is the ratio of accurately negative identified subjects by a test against all negative subjects. Precision is a ratio of correctly positive identified subjects by a test against all positive subjects by test.

Deployment is the last step in the CRISP-DM. This step will present and use the knowledge or data models obtained for new data. The steps will be evaluated and reviewed to check whether the models have achieved the business objectives and solved the problem. The result is assessed, and recommendations will be provided if needed, with a model redoing done if necessary. Monitoring is essential upon deployment.

Ethical approval

This study was approved by the Ministry of Health, Malaysia in the data use agreement for scientific publication document effective as of 03 April 2023.

RESULTS

This section discusses the results of descriptive analysis, the results of ML computations, and the discussion on the findings of this project. First, the descriptive analysis was displayed. Several pre-processing techniques have been used for the dataset (Γ_N =12000) to examine the raw data. Only adult participants with documented hypertension status (18–114 years old) are included in the data. Among the 480 patients, those without a hypertension condition were excluded. After the records are excluded, the final dataset becomes Γ_N =11,520 records with one target variable and 24 input features. Records on other variables with missing values have been replaced using point average since there are about 862 (7.5%) maximum missing values from the total dataset, as shown in Table II.

Table II: Statistics of Missing Values in Dataset

Feature Name	Frequency of missing value (%)
Hyp_Status	0 (0.00)
Strata	459 (3.98)
State	459 (3.98)
Sex	459 (3.98)
Ethnicity	459 (3.98)
Occupation	459 (3.98)
Education	459 (3.98)
Marital_status	459 (3.98)
Citizen	459 (3.98)
Abd_obesity	862 (3.98)
DM_status	1 (0.01)
Hyperchol_status	1 (0.01)
PA_status	135 (1.17)
Smoking_status	10 (0.09)
Anemia_Status	756 (6.56)
heavyedrinker_status	459 (3.98)
Area	0 (0.00)
Age	0 (0.00)
Weight	795 (6.90)
Height	796 (6.91)
Total_Ind_Income	511 (4.44)
Total_HH_Income	459 (3.98)
Waist_circumference	862 (7.48)
BMI	812 (7.05)
Average_Intake_FV	25 (0.22)

The data also checked for the imbalance dataset. Imbalanced data is one of the potential problems in the field of data mining and machine learning. The target variable distribution, which is hypertension, was also checked, and it was found that the percentage difference between yes and no is 28%, which is less than 50%. It can be concluded that the dataset is balanc. Meanwhile, Table III displays the descriptive statistics for gender percentage according to hypertension status, diabetes mellitus status percentage according to hypertension status and marital status according to hypertension status.

Table III: Descriptive Statistics for demographic factors

Feature Name	Percentage (%) hypertension	Percentage (%) non-hypertension
Hyp_Status distribution	64.00	36.00
Gender	Male-16.00 Female - 20.30	Male-30.20 Female - 33.60
DM status	No Diabetes Melitus - 23.20 Have Diabetes Melitus - 13.10	No Diabetes Melitus - 55.70 Have Diabetes Melitus - 8.10
Marital status	Never mrried- 3.10 Married - 27.50 Widow /Widower / Divorce - 5.70	Never manied-17.70 Married- 42.50 Widow /Widower / Divorce - 3.50

Next, the filter-based feature selection was applied to the hypertension dataset as this method is easy to use and has faster computation than other feature selection methods. Table IV shows the feature ranking. Age is the essential factor, followed by waist circumference, diabetes mellitus status, marital status, BMI, abdominal obesity status, hypercholesterolemia status, weight, height, occupation, total individual income, total household income, education, living area, and citizenship. As mentioned earlier, three filter feature selection were applied to the dataset and the features are ranked based the weight where the higher weight corresponds to the important features. Then, the total ranks were computed to identify the important features.

Third, the models' performance was assessed in this study. Decision trees are widely employed in various clinical scenarios because of their simplicity and prominence in machine learning. Age is the most significant feature in this research; features are arranged in branches based on their relative value. According to the decision tree model, the other essential features to predict hypertension were total individual income, BMI, age, height, waist circumference, total household income and weight. Meanwhile, this operator uses a feed-forward neural network trained via a multi-layer perceptron back propagation technique to enhance computing and develop a model. At the same time, the accuracy is the same as the decision tree result, which is 73.02%, as shown in Table IV.

Table IV: Feature ranking and total rank for filter-based feature selection

Feature Name	Rank-based on Filter feature weight information gain ratio	Rank-based on Filter feature weight information gain	Rank-based on Filter feature correlation	Total Rank
Age	1	1	1	3
Waist_ circumference	2	3	2	7
DM_status	5	4	3	12
Marital_status	8	2	7	17

CONTINUE

Table IV: Feature ranking and total rank for filter-based feature selection. (CONT.)

Feature Name	Rank-based on Filter feature weight information gain ratio	Rank-based on Filter feature weight information gain	Rank-based on Filter feature correlation	Total Rank
BMI	6	8	4	18
Abd_obesity	9	6	5	20
Hyperchol_status	10	7	6	23
Weight	7	10	8	25
Height	3	13	9	25
Occupation	13	9	10	32
Total_Ind_ Income	4	11	17	32
Total_HH_ Income	12	12	11	35
Education	11	5	20	36
Area	16	15	12	43
Citizen	14	17	14	45
Smoking_status	17	16	13	46
Strata	18	19	15	52
PA_Status	19	20	16	55
State	20	14	22	56
Average_Intake_ FV	15	23	21	59
Sex	22	21	18	61
Ethnicity	21	18	24	63
Anemia_Status	23	22	19	64
Heavydrinker_ status	24	24	24	71

One of the objectives of this study is to identify the best model for predicting hypertension among adults in Malaysia. The best model must be selected to provide the best compromise between sensitivity, specificity, and accuracy. Table V presents the assessment metrics of the various models for each of the assessed schemes. From the point of view of accuracy and sensitivity, the logistic regression model is better than the decision tree model, with an accuracy of 76.73% and a sensitivity of 86.47%. In contrast, on the other side, the decision tree is better for specificity and precision than different MLs, with 71.49% and 82.02%, respectively.

Table V: Performance of machine learning model

Performance indicator	Logistic Regression	Decision Tree	Artificial Neural Network
Accuracy	76.73	73.02	73.02
Specificity	59.58	71.49	67.57
Sensitivity	86.47	73.90	76.12
Precision	79.01	82.02	80.21

DISCUSSION

According to the result, logistic regression is the best ML model for predicting hypertension in adults in Malaysia. Based on this expected model shown in Table VI, a person who does not have diabetes mellitus is 2.05 odds likely to have hypertension. J. Yang et al. (2023)

supported these findings and used logistic regression to predict hypertension, and the model made accurate predictions. Meanwhile, a person who does not have hypercholesterol has a 1.67 odds of having hypertension, and with an increase in the age of adults, 6.0% are less likely to have hypertension. Age is also identified as an essential risk for predicting hypertension in adults based on previous studies (21). Another study identified the same important features in predicting hypertension: marital status, age, household income, diabetes mellitus status and education (22).

Table VI: Regression coefficient, odds ratio, and p-value for logistic regression model

Features	Regression coefficient Γ(b)	Odds ratio Γ(e^b)	p-value
DM_status.No	0.719	2.05	<0.001
Hyperchol_status.No	0.512	1.67	<0.001
Age	-0.059	0.94	<0.001
Waist_circumference	-0.019	0.98	<0.001
Marital_status.Never married	-0.304	0.74	0.001
Occupation.Unpaid worker/homemaker	-0.214	0.81	0.021
Education.Primary	-0.157	0.85	0.033
Total_HH_Income	0.000	1.00	0.050

Meanwhile, Luo et al. (23) revealed that marital status, age, diabetes mellitus status and education were essential features when they applied the logistic regression model to the hypertension dataset. Specifically, for every unit increase in waist circumference of adults, 2.0% less likely to have hypertension. This finding is paralleled to the previous study where the ratio of waist circumference to the body mass index significantly acts as a predictor for hypertension (24,25). Adults who never married are 26.0% less likely to have hypertension. Similarly, a previous study also found that single-marital status patients have a lower risk of hypertension (26). In addition, unpaid workers/homemakers are 19% less likely to have hypertension. Inversely, the previous study contradicts the finding that the employed worker has a lesser risk towards hypertension compared to homemakers (27). For adults with Primary education as their highest qualification, 15.0% less likely to have hypertension. It is supported by the previous study that lower education leads to a lower risk of hypertension (28). The final logistic regression model for predicting the hypertension in Malaysia is shown as follows:

Logit (Hypertension)
= 0.719DMstatus. No + 0.512Hyperchol_status.No - 0.059Age
- 0.019Waist_circumference - 0.304Marital_status. Nevermarried
- 0.214Occupation.Unpaid worker/homemaker
- 0.157Education.Primary +0.00Total_HH_Income + 4.599

CONCLUSION

The optimal model for predicting hypertension using data from the Malaysian Ministry of Health (MOH) was created using machine learning. The initial dataset has 24 features. Using filter feature selection, which includes information gain, information gain ratio, and correlation, can eliminate unimportant variables and improve accuracy and classification performance. It reduced to 15 from 24 features. The result showed that the logistic regression model is the best machine learning for predicting factors that influence hypertension in adults in Malaysia and achieved 76.73% accuracy, 86.47% sensitivity, 79.01% precision and 59.58% specificity. From this logistic regression model, the essential features that influence hypertension in adults in Malaysia were diabetes mellitus, hypercholesterolemia status, age, waist circumference, marital status, occupation, education, and total household income. The result of the predictive models can help healthcare professionals and patients be aware of their health. The recommendations would be that the government and private sectors may create more awareness or campaigns of hypertension, focusing more on the essential and significant risk factors. It is also advisable for the next researcher to study the other factors associated with hypertension more in-depth and expand the comparison on different ML model algorithms.

ACKNOWLEDGEMENT

Authors acknowledge the Universiti Teknologi MARA for funding under the Malaysian Research Assessment (MyRA) Grant Scheme 600-RMC/GPM LPHD 5/3 (055/2021). The authors also acknowledge Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Research Nexus UiTM (ReNeU), and College of Computing, Informatics and Mathematics UiTM.

REFERENCES

1. Ramezankhani A, Kabir A, Pournik O, Azizi F, Hadaegh F. Classification-based data mining for identification of risk patterns associated with hypertension in Middle Eastern population. Medicine (Baltimore) [Internet]. 2016 Aug 1;95(35):e4143. Available from: https://journals.lww.com/00005792-201608300-00005
2. Nowbar AN, Gitto M, Howard JP, Francis DP, Al-Lamee R. Mortality From Ischemic Heart Disease. Circ Cardiovasc Qual Outcomes [Internet]. 2019 Jun;12(6). Available from: https://www.ahajournals.org/doi/10.1161/CIRCOUTCOMES.118.005375
3. Alias N, Ying Ying C, Kuang Kuay L, Ahmad A, Mat Rifin H, Shahein NA, et al. Physical Inactivity and Its Associated Factors among Adults in Malaysia: Findings from National Health and Morbidity Survey (NHMS) 2019. Int J Public Heal Res [Internet]. 2022 Feb 17;12. Available from: https://spaj.ukm.my/ijphr/index.php/ijphr/article/

view/366
4. Isamail N, Said R, Ismail NW, Haron SA. Non-Communicable Diseases Impact Low-Income Households in Malaysia. Malaysian J Med Sci [Internet]. 2024 Feb 28;31(1):124–39. Available from: http://www.mjms.usm.my/MJMS31012024/MJMS31012024_11.pdf
5. Ab Hamid MR, Buhari SS, Noor HM, Azizan N 'Ain, Malek KA, Mohd Asmawi UM, et al. Development and validation of D-PATH website to improve hypertension management among hypertensive patients in Malaysia. Digit Heal [Internet]. 2024 Jan 1;10:20552076241242660. Available from: https://doi.org/10.1177/20552076241242661
6. Islam SMS, Talukder A, Awal MA, Siddiqui MMU, Ahamad MM, Ahammed B, et al. Machine Learning Approaches for Predicting Hypertension and Its Associated Factors Using Population-Level Data From Three South Asian Countries. Front Cardiovasc Med [Internet]. 2022 Mar 31;9(March). Available from: https://www.frontiersin.org/articles/10.3389/fcvm.2022.839379/full
7. AlQuaiz AM, Siddiqui AR, Kazi A, Batais MA, Al-Hazmi AM. Sedentary lifestyle and Framingham risk scores: a population-based study in Riyadh city, Saudi Arabia. BMC Cardiovasc Disord [Internet]. 2019;19(1):88. Available from: https://doi.org/10.1186/s12872-019-1048-9
8. Jahangiry L, Farhangi MA, Rezaei F. Framingham risk score for estimation of 10-years of cardiovascular diseases risk in patients with metabolic syndrome. J Heal Popul Nutr [Internet]. 2017 Dec 13;36(1):36. Available from: https://jhpn.biomedcentral.com/articles/10.1186/s41043-017-0114-0
9. Martinez-Rнos E, Montesinos L, Alfaro-Ponce M, Pecchia L. A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. Biomed Signal Process Control [Internet]. 2021;68. Available from: https://www.sciencedirect.com/science/article/pii/S1746809421004109
10. Mroz T, Griffin M, Cartabuke R, Laffin L, Russo-Alvarez G, Thomas G, et al. Predicting hypertension control using machine learning. PLoS One [Internet]. 2024 Mar 20;19(3):e0299932. Available from: https://doi.org/10.1371/journal.pone.0299932
11. Wu Y, Xin B, Wan Q, Ren Y, Jiang W. Risk factors and prediction models for cardiovascular complications of hypertension in older adults with machine learning: A cross-sectional study. Heliyon [Internet]. 2024 Mar 30;10(6). Available from: https://doi.org/10.1016/j.heliyon.2024.e27941
12. Li L, Momma H, Chen H, Nawrin SS, Xu Y, Inada H, et al. Dietary patterns associated with the incidence of hypertension among adult Japanese males: application of machine learning to a cohort study. Eur J Nutr [Internet]. 2024; Available from: https://doi.org/10.1007/s00394-024-03342-w
13. Ragnarsdottir H, Ozkan E, Michel H, Chin-Cheong K, Manduchi L, Wellmann S, et al. Deep Learning Based Prediction of Pulmonary Hypertension in Newborns Using Echocardiograms. Int J Comput Vis [Internet]. 2024; Available from: https://doi.org/10.1007/s11263-024-01996-x
14. Edvinsson C, Bjurnsson O, Erlandsson L, Hansson SR. Predicting intensive care need in women with preeclampsia using machine learning–a pilot study. Hypertens Pregnancy [Internet]. 2024;43(1). Available from: https://doi.org/10.1080/10641955.2024.2312165
15. Srimaneekarn N, Hayter A, Liu W, Tantipoj C. Binary Response Analysis Using Logistic Regression in Dentistry. Khurshid Z, editor. Int J Dent [Internet]. 2022 Mar 8;2022(1):1–7. Available from: https://www.hindawi.com/journals/ijd/2022/5358602/
16. Md Noh SS, Ibrahim N, Mansor MM, Yusoff M. Hybrid filtering methods for feature selection in high-dimensional cancer data. Int J Electr Comput Eng [Internet]. 2023 Dec 1;13(6):6862. Available from: https://ijece.iaescore.com/index.php/IJECE/article/view/31694
17. Ibrahim N, Abd Rahman HA, Azran AA, Mohd Fadillah MA, Qamarudin M, Muhammad Adhwa' Qayyum. Prediction of Water Quality for the Selangor Rivers using Data Mining Approach. J Sustain Sci Manag [Internet]. 2023 Sep 30;18(9):171–83. Available from: https://jssm.umt.edu.my/wp-content/uploads/2024/02/12-JSSM-Volume-18-Number-9-September-2023-FINAL.pdf
18. Ibrahim N, Ishak UM, Ali NNA, Shaadan N. Machine Learning-Based Approaches for Credit Card Debt Prediction. Malaysian J Comput [Internet]. 2024 Apr 1;9(1):1722–33. Available from: https://ir.uitm.edu.my/id/eprint/61998/
19. Rosly R, Makhtar M, Awang MK, Rahman MNA, Deris MM. The Study on the Accuracy of Classifiers for Water Quality Application. Int J u- e-Service, Sci Technol [Internet]. 2015 Mar 31;8(3):145–54. Available from: http://article.nadiapub.com/IJUNESST/vol8_no3/13.pdf
20. Sweeney C, Ennis E, Mulvenna M, Bond R, O'Neill S. How Machine Learning Classification Accuracy Changes in a Happiness Dataset with Different Demographic Groups. Computers [Internet]. 2022 May 23;11(5):83. Available from: https://www.mdpi.com/2073-431X/11/5/83
21. Yang J, Wang X, Jiang S. Development and validation of a nomogram model for individualized prediction of hypertension risk in patients with type 2 diabetes mellitus. Sci Rep [Internet]. 2023;13(1):1–10. Available from: https://doi.org/10.1038/s41598-023-28059-4
22. Deng D, Chen C, Wang J, Luo S, Feng Y. Association between triglyceride glucose-body mass index and hypertension in Chinese adults: A cross-sectional study. J Clin Hypertens [Internet].

2023 Apr 17;25(4):370–9. Available from: https://onlinelibrary.wiley.com/doi/10.1111/jch.14652
23. Luo J, Zhang T, Yang L, Cai Y, Yang Y. Association between relative muscle strength and hypertension in middle-aged and older Chinese adults. BMC Public Health [Internet]. 2023 Oct 25;23(1):2087. Available from: https://doi.org/10.1186/s12889-023-17007-6
24. Wang Y, Hua Y, Zhang H, Liang S, Cao Z, Chen L, et al. Ratio of waist circumference to body mass index: A novel predictor of clinical outcome in hypertension patients. J Clin Hypertens [Internet]. 2024 Jan 21;26(1):24–35. Available from: https://onlinelibrary.wiley.com/doi/10.1111/jch.14739
25. Yang Y, Sun Q, Ma S, Li X, Lang X, Zhang Q. Association of serum creatinine to cystatin C to waist circumference ratios and hypertension: evidence from China health and retirement longitudinal study. Front Endocrinol (Lausanne) [Internet]. 2024 May 1;15(May):1–10. Available from: https://www.frontiersin.org/articles/10.3389/fendo.2024.1375232/full
26. Ramezankhani A, Azizi F, Hadaegh F. Associations of marital status with diabetes, hypertension, cardiovascular disease and all-cause mortality: A long term follow-up study. Shimosawa T, editor. PLoS One [Internet]. 2019 Apr 22;14(4):e0215593. Available from: https://dx.plos.org/10.1371/journal.pone.0215593
27. Rose KM, Newman B, Tyroler HA, Szklo M, Arnett D, Srivastava N. Women, Employment Status, and Hypertension: Cross-Sectional and Prospective Findings from the Atherosclerosis Risk in Communities (ARIC) Study. Ann Epidemiol [Internet]. 1999;9(6):374–82. Available from: https://www.sciencedirect.com/science/article/pii/S1047279799000150
28. Zacher M. Educational Disparities in Hypertension Prevalence and Blood Pressure Percentiles in the Health and Retirement Study. Journals Gerontol Ser B [Internet]. 2023 Sep 1;78(9):1535–44. Available from: https://doi.org/10.1093/geronb/gbad084