"Comparative Study: Evaluating Factors Influencing Heart Disease Using Logistics and Covariate Analysis"

المؤلف:

ايمان شعيب عيسي
Eman S. Essa
قاسم عبدالنبي فرج
Kasem A. Farag

ابوبكر محمد جابر
Abobaker M. jaber

Abstract

Heart disease is one of the most dangerous and common diseases at the international and local levels, especially in recent years, and is considered one of the main causes of death. This study aims to identify the most important factors associated with the risk of heart disease through a methodology that relies on discriminant analysis and logistic analysis as two important and effective models in identifying risk factors resulting from heart disease and correct prediction of it, by subjecting a sample of individuals with heart disease. And healthy people, after obtaining their consent in a group of local clinics and hospitals, taking into account the inclusion criteria that include adults who are between the ages of 40 and 70 and have been diagnosed with heart disease and comorbidities such as high blood pressure, cholesterol, diabetes, smokers, and taking into account the exclusion criteria, which include excluding pregnant and breastfeeding wome patients who have infections or any bacterial infection, or suffer from chronic diseases, as well as patients who take Anticoagulants (such as warfarin) and some psychiatric medications. The results indicated that the two models were close, as the accuracy of the logistic analysis was 83% compared to 79% for the discriminant analysis, the sensitivity of the logistic analysis was 81% and the discriminant analysis was 78%, and the specificity was 84% compared to 80% for the discriminant analysis. As for the American University in Cairo analysis, the results of the logistic analysis reached 89% and 85% for the discriminant analysis. The results also indicated that logistic analysis has greater flexibility in interpreting and identifying infected cases compared to uninfected cases. The most influential variables were similar in the two models: age, smoking, diabetes, and physical activity.

Keywords:(logistic analysis, discriminant analysis, heart disease, accuracy, sensitivity, specificity)

Introduction

Heart disease is one of the most common and dangerous diseases at the local and international levels, especially in recent years, and it is considered one of the main causes of death in many countries. The total number of deaths from heart disease amounts to 32% of the total number of deaths in the world. Figure (1) shows a map of the world showing death rates [1].

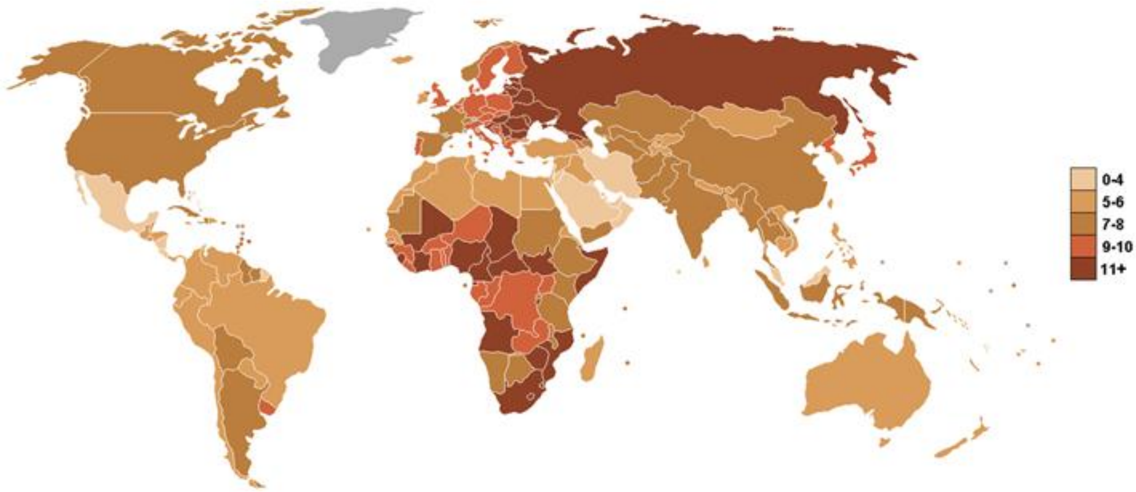

Figure 1: shows a map of the world showing death rates due to heart disease

Perhaps the reason for the spread of heart disease is the multiplicity of factors that affect the incidence of this disease, as it varies between biological factors such as age and gender, behavioral factors such as smoking and physical inactivity, and medical factors such as blood pressure, cholesterol, and diabetes. Therefore, early prediction of the risk of infection greatly helps in preventing and reducing complications resulting from heart disease and reducing deaths considering the effort to achieve sustainability in the medical sector like other fields [2].

The study aims to identify the most important factors associated with the risk of heart disease and use discriminant analysis to evaluate these factors, compare the results of the two models, and evaluate their ability to correctly predict heart disease, especially in light of the availability of many statistical models through which the results can be analyzed and the severity of heart disease predicted.[3]

Despite the development of therapeutic strategies and the availability of many statistical models to diagnose and predict the severity of heart disease, the difference in the performance of these models raises the question about the extent of their effectiveness in classifying patients and the extent of their effectiveness in accurately determining the factors affecting heart disease. Here the importance of the need for methods such as logistic analysis and discriminant analysis to determine the best models in terms of accuracy and predictive ability in light of the available data and in light of the development of programming sciences and the use of artificial intelligence techniques that rely on networks is highlighted. Artificial neural networks in assessing and predicting the risk of heart disease [4].

The importance of the study is since it addressed the subject from multiple aspects that many studies have overlooked. It also presented the obstacles and challenges facing the use of such models, presented solutions, and presented a clear method for applying these models. The importance of this study is also due to the fact that it contributes to supporting medical decision-making by providing accurate analytical tools and helps direct preventive efforts towards the most influential factors. It also provides a practical comparison between two common statistical models in medical analysis, which benefits researchers and professionals in choosing the model. Most appropriate. In short, it can be considered a reference model for researchers and scholars in topics related to disease predictions using artificial intelligence or machine learning later.

Theoretical background and previous literature review (Literature Review)

In this section, the theoretical background of the study and the basic concepts and terminology will be presented so that the reader can form an insightful point of view about the study's procedures, objectives, methodology, and the most important results indicated by the study and the recommendations it presented. Also, some previous studies related to the use of heart disease analysis will be presented, as well as discriminant analysis, all through a literary review of previous studies and identifying the strengths and weaknesses of these studies and the points of difference and compatibility with The current study.

Logistic Regression

Logistic analysis is a statistical model that is used to estimate the relationship between a binary dependent variable, such as risk or non-risk, injury or non-injury, and other groups of independent variables. The logistic function can be expressed to determine the probability of an event occurring, such as injury, for example, through the following relationship [5]:

$P\ y1=1/e-(\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta n Xn)$

where:

- $b0u\ \beta 1\ b2, \ldots\ \beta n$: parameters
- $x1, x2, x3, \ldots, xn$: variables

- Odds Ratio: It shows how many times the probability of infection increases when the independent variable changes.
- Significance level (P-value): To determine how statistically significant a variable .
- ROC-AUC value: To measure the quality of the model in prediction.

Linear Discriminant Analysis - LDA

It is also a statistical model that is used to classify observations into categories, such as injured or not injured, such as dangerous or not dangerous, based on a set of variables and relying on a linear function that separates the categories, where the individual is classified based on the value of the discriminating function into the closest category. This model can be expressed by the following linear function [6]:

It is a model used to classify observations into categories (eg infected or uninfected) based on a set of variables.

It is based on finding a linear function that separates the categories:

$$D(x)=a+b1X1+b2X2+...+bnXn$$

where:

- X1, X2...Xn: is the number of variables
- a, b1, b2,...bn: parameters
- dx: the value of the discriminant function

One of the most important concepts associated with this model is that it assumes equal variance in the variance and change matrices between groups. The fundamental difference between logistic analysis and discriminant analysis can be identified in Table (1).

Table 1: Shows a comparison between logistic analysis and linear discrimination analysis.

ITEM	Logistic analysis	Discriminant analysis
Nature of the model:	Probabilistic	linear
Appropriate data type	Normal distribution not required	Normal distribution
Dealing with variance	Flexible	assumes equal variance
Primary objective	Quantitative estimation of the probability of infection	classification

3. heart **disease:** It can be defined as a group of disorders that affect the heart muscle, such as coronary artery disease and heart failure, and is affected by a group of important factors, as previously mentioned. The variables are biological factors such as age and gender, behavioral factors such as smoking, and genetic factors that may increase the probability of infection, in addition to accompanying diseases such as diabetes, blood pressure, and cholesterol [7].

4. Variables: This means the factors that affect heart disease, such as age, lifespan, incidence of concomitant diseases, and genetic and behavioral factors. The variables may be independent variables or dependent variables, where the independent variable is symbolized by the symbol

5. Model success indicators: which are a set of statistical indicators through which the extent of the model's success is evaluated and determined. Among these indicators are [9]:

1.Accuracy: Percentage of cases classified correctly.

2.Sensitivity: The model's ability to detect infected people.

3.Specificity: The model's ability to detect non-infected people.

Previous studies

Many previous studies have shown the importance of using statistical models in analyzing and evaluating the risk of heart disease and predicting it. Many of these studies have dealt with logistic analysis and discriminant analysis because of their ability to deal with binary variables and the independence of explanatory variables and considering discriminant analysis as an alternative performance that shows similar effectiveness in classification processes, especially in cases where basic statistical assumptions such as normal distribution and variance in models are met [10].

1. In a study he conducted in a study conducted by the Framingham Heart Study, the study aimed to determine the relationship between blood pressure, cholesterol, age, and smoking through methodology. It used logistic analysis and considered it a standard model for predicting infection. The results indicated the effectiveness of the analysis, as its accuracy reached 95%. [11].

2. In another study, another study was applied (Smith et al., 2015) that aimed to determine the relationship and classify patients into infected and uninfected through the use of analysis methodology and the use of discriminant analysis. The results indicated that the model achieves high accuracy, especially in the presence of cases of normal distribution of independent variables as well as independence of variance, but the effectiveness decreased slightly when the data were unbalanced and non-linear [12].

Also, many other studies that used logistic analysis or discriminant analysis have shown good results, whether in terms of classification or in terms of accuracy and prediction. In this study, the two analyzes were used and a comparison was made between them, and this is the most important thing that distinguishes the study, as it is estimated that it is new in integrating more than one technique to predict heart disease and benefiting from the advantages of these techniques [13].

Material and Methods:

Different methodologies were adopted in the study. A descriptive methodology was used to describe the data, a quantitative methodology was used to collect data, an analytical methodology was used to analyze and evaluate the results, and a comparative methodology was used to assess the relationship between a group of factors affecting heart disease. Figure 2 illustrates the applied framework of the study, starting from the stage of defining the objective and formulating the research problem, through collecting and processing data, then determining the tools and work plan. 250 individuals were selected for testing, 137 were excluded according to the exclusion criteria, and 113 individuals were selected according to the inclusion criteria, including those with and without heart disease. A comparative analysis was then conducted to evaluate the relationship between a group of factors (age, gender, blood pressure, cholesterol, etc.) and the likelihood of developing heart disease. This was done using two statistical models: logistic analysis and linear discriminant analysis. A sample of 113 individuals were used, including those with and without heart disease. The results of the logistic analysis and linear discrimination analysis were then compared to assess the compatibility of the results and evaluating the performance of the two models in terms of accuracy, sensitivity, and specificity [14].

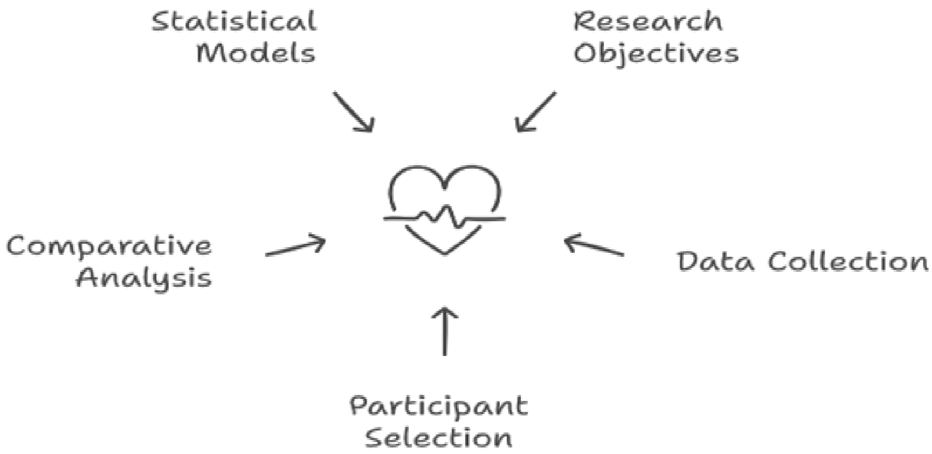

Figure 2: shows study methodology.

procedures

1. Determining the goal of the study.

Determine the goal and formulate the research problem, The main objective was to compare logistic analysis and linear discriminant analysis to determine the risks associated with heart disease based on a set of variables such as age,

gender, smoking, body weight, region of residence, and some comorbidities such as diabetes, blood pressure, and cholesterol. The secondary objective was to evaluate the accuracy of the models in predicting heart disease and identify the most discriminating variables in distinguishing between heart patients[15].

2. Study design

The type of study was determined as an applied study using data extracted from 250 people, of whom 113 were selected according to exclusion and inclusion criteria and through a comparison methodology between logistic analysis and discriminant and linear analysis using local data in a group of local hospitals and the databases associated with them.

3. Sample selection

A group of visitors to hospitals was selected, including data from those with and without heart disease. The sample number was 250 people, who were filtered to 113.

Sample number: 250 people selected from the relevant hospital or clinic.

- 113 people were selected based on comprehensive criteria including age, gender and health status.
- Exclusion criteria: 137 people were excluded based on certain criteria such as:
- Lack of complete data available.
- The presence of advanced medical conditions (such as cancer).
- Too old or too young.

4. Data collection

Data were collected related to demographic data such as age and gender, behavioral data such as smoking and physical activity, health variables such as cholesterol levels, blood pressure, and family medical history, in addition to physical indicators such as body mass index.

5. Tools and techniques

Medical records and laboratory tests recorded in the databases of a group of hospitals were relied upon, and the necessary permissions were taken from the participating individuals to agree to their participation in the results of the analysis and to rely on their data.

The logistic analysis method and the linear discriminant analysis method were chosen, in addition to Anova tests to analyze the extent of the importance of the data and the extent of its variance and its suitability for statistical analysis. A program was used. SPSS and Python

the data was cleaned and processed, then the missing and invalid values were deleted, the behavioral and health variables were classified correctly, after that the means and standard deviations for the independent variables were calculated, and finally, the age distribution and percentages for gender, age, percentages of smokers, health conditions, and comorbidities such as diabetes, cholesterol, and blood pressure were determined[16].

6.Apply statistical analysis

1) Logistical analysis:

- Build a logistic regression model from independent variables (age, gender, smoking, BMI …).
- Use p-value to determine statistical significance of variables.
- Estimate probabilities for each case.

2) Linear discriminant analysis:

- Discriminate Function: Training a Discrimination model
- Variance differences across groups were analyzed using Wilks' Lambda test.
- Use Standardized Coefficients (B) and proceed to find the length of each one to determine the discriminatory weight of each variable.

7.Compare models

1) Accuracy comparison:

You have been analyzing the performance of both models using common metrics to compare the fit of the two models, such as the percentage of correct classification, the area under the ROC curve (AUC), and the p-value.

2) Compare important variables:

Determine influential features per model and their impact to the output.

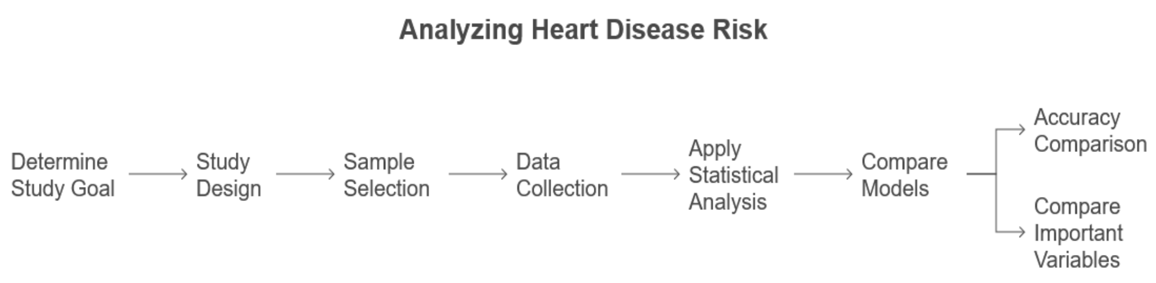

Figure 3: shows Analyzing Heart Disease Risk procedures

Results and discussion

In this section, we will present the results of the descriptive and statistical analysis for the participating sample, in addition to presenting the logistical analysis, displaying the results of the discriminant analysis, and finally the comparison between the results of the logistical analysis and the discriminant analysis.

Statistical Description of the Variables

Table2: the descriptive and statistical analysis for the participating sample

Characteristics	total (n = 113) No.	%	f	p-value
Sex				
Male	60	53		
Female	53	47		
Age (years)			7.8	<0.001
Min. – Max.	40.0 – 76.0			
Mean ± SD.	60.40 ± 10.20			
Median (IQR)	60.50 (50.0 – 66.0)			
BMI (kg/m^2)			4.04	<0.001
Min. – Max.	25.0 – 29.0			
Mean ± SD.	26.50 ± 1.65			
Median (IQR)	26.0 (25.0 – 28.0)			
smoking			5.2	<0.001
Male	69	60		
Female	2	1.7		
Residence			3.14	<0.001
Urban	80	70		
Rural	32	28		
Comorbidities			6.36	<0.001
Diabetes	43	38		
High blood pressure	41	36		
Cholesterol	35	31		

The table shows that the females and males are nearly evenly distributed, with a slightly higher proportion of males (53%). The age ranges between 40 and 76 years, with a mean of 60.40 years and a standard deviation of 10.20, reflecting wide age variability. The p-value for age is <0.001, reflecting a significant age difference. BMI ranges from 25.0 to 29.0 kg/m², with a mean of 26.50 kg/m² and a standard deviation of 1.65 kg/m², and the p-value is less than 0.001, indicating a significant difference in BMI. Smoking is more common in males (60%), with p-value less than 0.001, indicating a significant gender difference in smoking. Most of the participants live in urban areas (70%), p-value < 0.001, indicating a significant difference in residence. Regarding comorbidities, a high proportion of the participants have diabetes (38%), high blood pressure (36%), and high cholesterol (31%), where the p-value for high blood pressure is < 0.001, indicating a significant difference. The coefficient of variation for the data was large, which means that the data is suitable for statistical analysis as it has differences and statistical significances, and the p-value value was less than the limit value, which is equal to 5%, which means that the data is of a large degree[17]

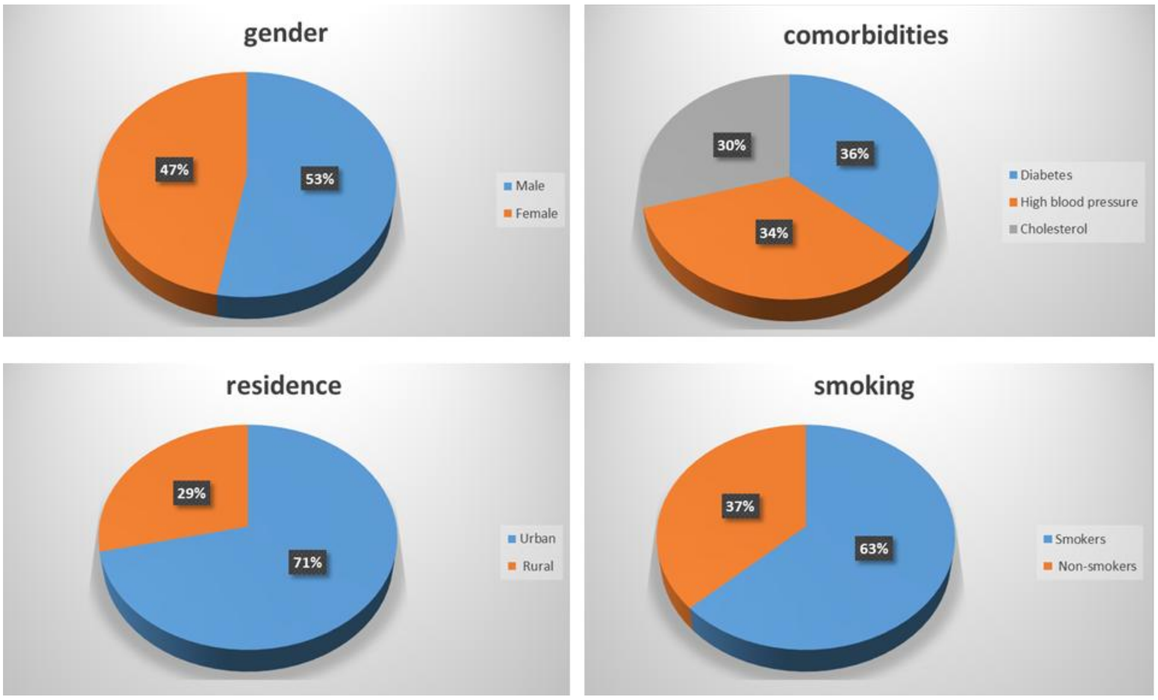

Figure 4: the descriptive and statistical analysis for the participating sample.

The previous figure shows the demographic characteristics of the individuals who participated in the experiment, where the percentage of females was 47% and the percentage of males was 53%, the number of smokers was 63% and the number of non-smokers was 37%. As for the place of residence, the percentage of residents in urban areas was 71% and in rural areas 29%. As for the comorbidities, out of a total of 113 people, 36% were suffering from diabetes, 34% from high blood pressure, and 30% from cholesterol.

Logistic Analysis Results

Table 3: Influential Variables and their Statistical Coefficients.

Variable	Coefficient β	Probability Value (p)	Odds Ratio (OR)
Age	0.045	0.001	1.05
Smoking	0.82	0	2.27
Diabetes	64	0.008	1.68
Blood Pressure	0.03	0.021	1.03
Cholesterol	0.04	0.03	1.21

Physical Activity	-0.55	0.014	0.58
BMI	0.041	0.028	1.11

The table outlines the logistic regression analysis output to examine the impact of various independent variables on a binary dependent variable. The key findings are that age, smoking, diabetes, blood pressure, cholesterol, physical activity, and BMI significantly impact the event probability. Specifically, smoking (OR = 2.27) and diabetes (OR = 1.68) increase the event likelihood, while physical activity (OR = 0.58) decreases it. With each unit rise in age, cholesterol, blood pressure, and BMI, the risk of the event is also rising, with maximum rise being of cholesterol at 21%. The p-values for all the variables are less than 0.05 and therefore are statistically significant. The diabetes coefficient (β = 64) appears exceedingly high and possibly needs to be checked[18].

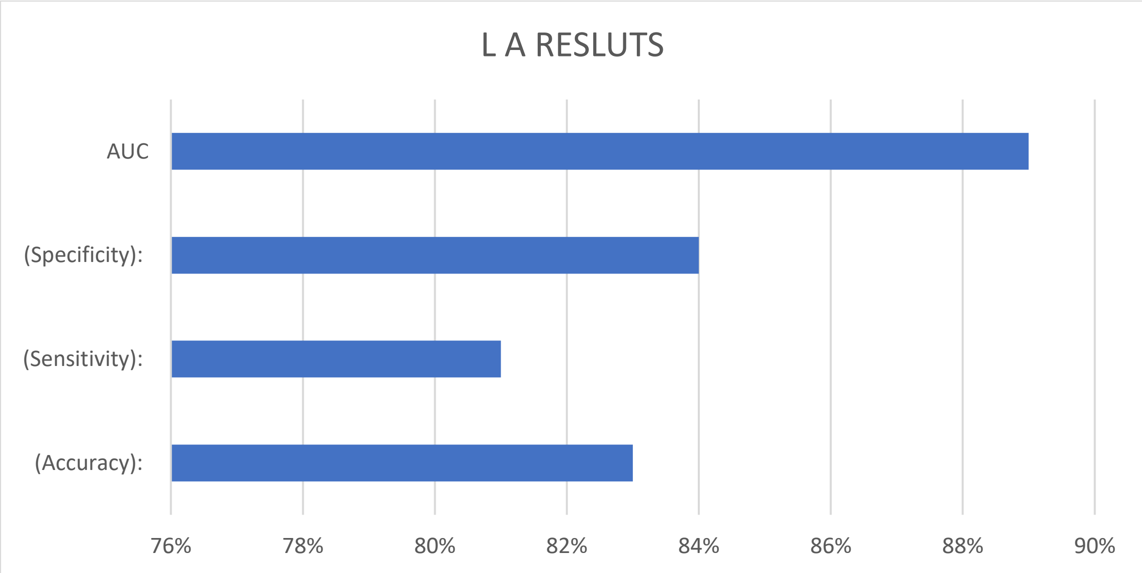

Figure 5: Logistic Analysis Results.

The previous figure shows the accuracy, sensitivity and specificity measures (AUC) for logistic analysis. The accuracy of logistic analysis was 83%, the specificity of logistic analysis was 81%, while the sensitivity of logistic analysis was 84% and AUC 89%.

Discriminant Analysis Results

Table 4: Discriminant Function and Associated Statistical Values.

Comparison Group	Important Variables in Discrimination	Discriminant function coefficients	Classification accuracy (%)	f	p-value

Males vs. Females	Smoking, BMI	0.55	75	15.2	<0.001
Urban vs. Rural	Age, BMI	0.3	70	8.7	0.004
Diabetic vs. Non-Diabetic	Age, BMI, Smoking	0.51	73	12.5	0.001
Hypertensive vs. Non-Hypertensive	Age, BMI	0.5	77	18.9	<0.001
Hypercholesterol vs. Non-Hypercholesterol		0.5	71	10.3	0.002

The table presents the discriminant analysis between different groups, revealing significant variables, discriminant function coefficients, classification accuracy, and statistical significance. In comparison between males and females, smoking and BMI were significant discriminators with 75% classification accuracy and high statistical significance ($p < 0.001$). For urban vs. rural comparisons, age and BMI were significant contributors with 70% classification accuracy and statistical significance ($p = 0.004$). For diabetic vs. non-diabetic comparisons, age, BMI, and smoking were significant contributors with 73% classification accuracy and statistical significance ($p = 0.001$). For hypertensive vs. non-hypertensive comparisons, age and BMI were the significant discriminators with 77% classification accuracy and high significance ($p < 0.001$). For hypercholesterolemia, while significant variables were not reported, the model was 71% accurate in classification with significant results ($p = 0.002$). Overall, the analysis indicates that age and BMI are useful for distinguishing between groups, with good classification performance and statistical significance in comparisons[19].

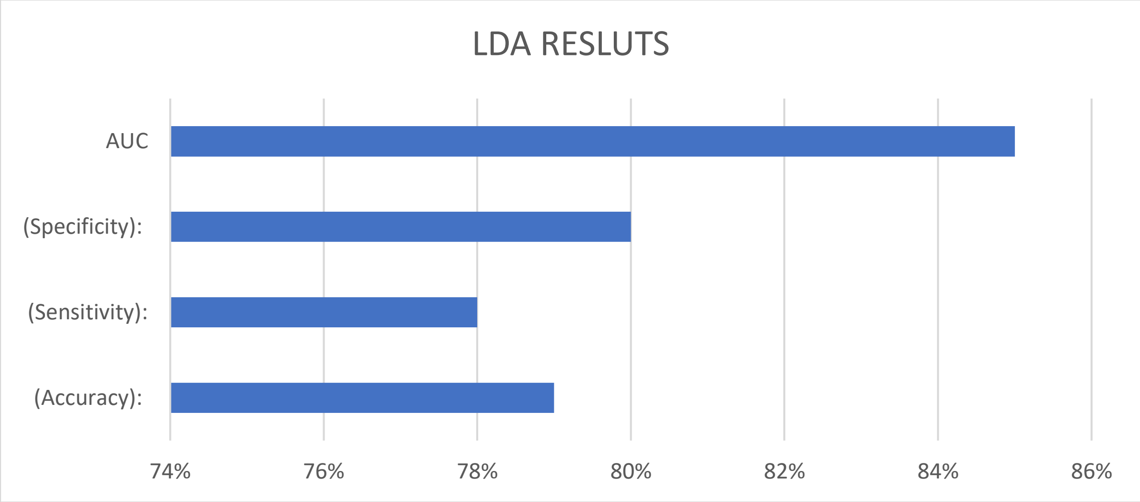

Figure 6: Discriminant Analysis Results

The previous figure shows the accuracy, sensitivity and specificity measures (AUC) for logistic analysis. The accuracy of Discriminant Analysis was 79%, the specificity of logistic analysis was 78%, while the sensitivity of logistic analysis was 80% and AUC 85%.

Comparison of the Two Models

Table 5: Comparison of the Classification Accuracy of Each Model.

Standard	Logistic Analysis	Discriminant Analysis
Accuracy	83%	79%
Sensitivity	81%	78%
Specificity	84%	80%
AUC	89%	85%

The previous table shows a comparison between logistic analysis and discriminant analysis for measures of accuracy, sensitivity, and specificity, where accuracy is the percentage of correct classifications, while sensitivity means the ability of a model to correctly identify positive cases, and specificity means the model's ability to correctly identify negative cases. (AUC) is an indicator of the general performance measure, and it is clear from the table that the logistic analysis was superior to the discriminant analysis in overall performance, as the general performance percentage for the logistic analysis reached 89% and for the discriminant analysis 85%.

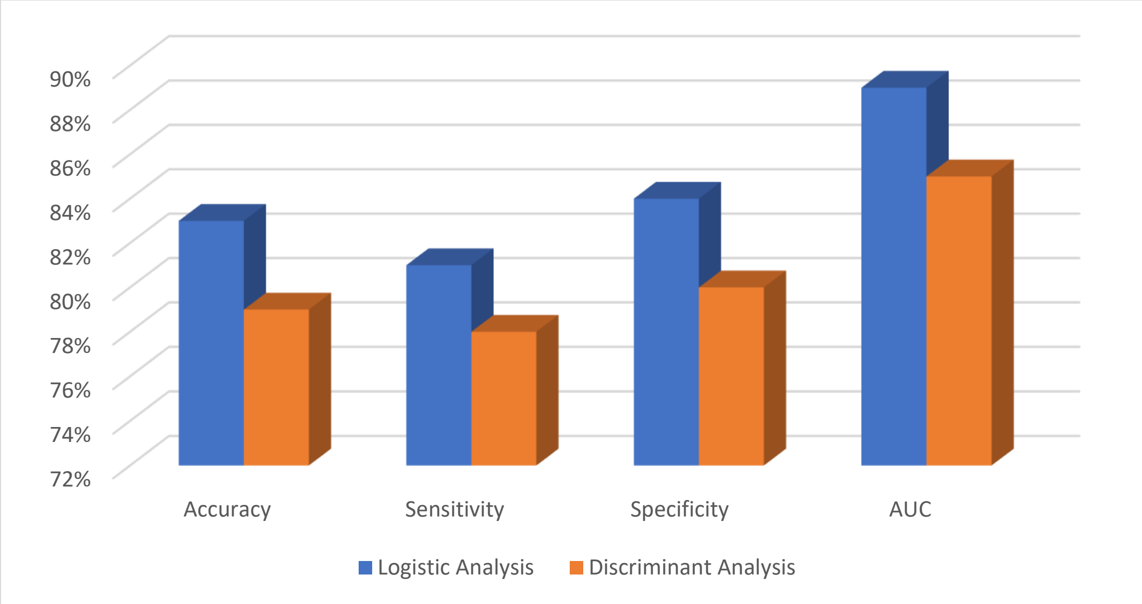

figure7: Comparison of the Classification Accuracy of Each Model.

The previous graph shows a comparison between logistic analysis and discriminant analysis through four measures: accuracy, sensitivity, and specificity (AUC). It is clear from the figure that logistic analysis is more accurate than discriminant analysis and that it was better in identifying positive cases and negative cases and has a better overall performance than discriminant analysis. Despite the superiority of logistic analysis, discriminant analysis provided a good performance[19]. The accuracy of logistic analysis was 83%, while the accuracy of discrimination analysis was 79%. The sensitivity of logistic analysis was 81%, the sensitivity of discriminant analysis was 78%, while (AUC) for logistic analysis was 89%. For discriminant analysis, 85%

Conclusions

Some of the most significant takeaways that were derived from this study are as follows:

- The accuracy of results in classifying the cases especially in the infected cases was higher with the use of logistic analysis [20].
- These two models are relatively similar, but the logistic model has a bit more wiggle room in terms of how it interprets a medical outcome.
- The accuracies as well as the specificities for logistic analysis (83%, 84%) were slightly better than for discriminant analysis (79%, 79%), while the sensitivities were comparable (81%, 78%). For the AUC analysis, the AUC value of logistic analysis 89% and discriminant analysis were 85%. The findings also demonstrated that logistic analysis is much more adaptable than Fuzzy logic to interpretation and cases identification. The second method gets infected from non-infected cases. The

most significant variables were nearly identical across the two models: age, smoking, diabetes and physical activity [21].

- The most significant factors are similar between the two models: age, smoking, diabetes, and physical activity.
- Age: The older someone is, the more likely he is to have it, a finding consistent with most medical studies, since with age come changes to blood vessels and increased calcification.
- And here are more destructive components: Smoking has doubled the risk of infection, speaking to the direct impact of smoking in petrifying and corroding the channels.
- Diabetes and high blood pressure: stressors that add a burden to the heart and impair its function, and which were strong predictors in both models.

Recommendations

The most important recommendations that can be presented through this study are the following:

- Logistic analysis and discriminant analysis are important tools in identifying risk factors, and they must be used with some expansion in studies related to heart disease. In addition to these models in patient records, they can quickly assess the level of risk and take the necessary measures.
- Preventive campaigns must be launched for smokers, the elderly, and diabetics to raise awareness about heart disease and how to take preventive measures to maintain heart health, in addition to supporting physical activity programs at the community level, especially for those at risk.
- It is necessary to repeat the study on larger samples and in different countries and environments, and to use other models, such as deep machine learning models, to evaluate risk factors, especially in light of the development of computer technology and computer vision.
- Trying to search for new techniques and methods to evaluate risk factors in heart disease by integrating techniques and using hybrid techniques to take advantage of every advantage of each technique.

References

[1] Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V., & Roth, G. A. (2022). The global burden of cardiovascular diseases and risk: a compass for future health. Journal of the American College of Cardiology, 80(25), 2361-2371.

[2] Mijwil, M. M., Shukur, B. S., & Mahmood, E. S. (2022). The most common heart diseases and their influence on human life: A Mini-review. Journal of Advances in Medicine and Medical Research, 34(15), 26-36.

[3] Gadó, K., Szabo, A., Markovics, D., & Virág, A. (2022). Most common cardiovascular diseases of the elderly–A review article. Developments in Health Sciences, 4(2), 27-32.

[4] Waigi, D., Choudhary, D. S., Fulzele, D. P., & Mishra, D. (2020). Predicting the risk of heart disease using advanced machine learning approach. Eur. J. Mol. Clin. Med, 7(7), 1638-1645.

[5] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. Journal of clinical epidemiology, 122, 56-69.

[6] Zhao, S., Zhang, B., Yang, J., Zhou, J., & Xu, Y. (2024). Linear discriminant analysis. Nature Reviews Methods Primers, 4(1), 70.

[7] Steyerberg, E. W. (2019). Clinical Prediction Models. Springer.

[8] World Health Organization (WHO). (2023). Cardiovascular Diseases (CVDs). Retrieved from https://www.who.int

[9] Zheng, J., Heidenreich, P. A., Kohsaka, S., Fearon, W. F., & Sandhu, A. T. (2022). Variability in coronary artery disease testing for patients with new-onset heart failure. Journal of the American College of Cardiology, 79(9), 849-860.

[10]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An Introduction to Statistical Learning. Springer.

[11] Battisti, N. M. L., Welch, C. A., Sweeting, M., de Belder, M., Deanfield, J., Weston, C., ... & Ring, A. (2022). Prevalence of cardiovascular disease in patients with potentially curable malignancies: a national registry dataset analysis. Cardio Oncology, 4(2), 238-253.

[12] Hosmer, D. W., & Lemeshow, S. (2013). Applied Logistic Regression. Wiley.

[13] Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics, 2, 100016.

[14] Xu, L., Yan, X., Tang, Z., & Feng, B. (2022). Association between circulating oxidized OxLDL/LDL-C ratio and the severity of coronary atherosclerosis, along with other emerging biomarkers of cardiovascular disease in patients with type 2 diabetes. Diabetes Research and Clinical Practice, 191, 110040.

[15] Rossello, X., & González-Del-Hoyo, M. (2022). Survival analyses in cardiovascular research, part II: statistical methods in challenging situations. Revista Española de Cardiología (English Edition), 75(1), 77-85.

[16] Nkonde-Price, C., Reynolds, K., Najem, M., Yang, S. J., Batiste, C., Cotter, T., ... & Funahashi, T. (2022). Comparison of home-based vs center-based cardiac rehabilitation in hospitalization, medication

adherence, and risk factor control among patients with cardiovascular disease. JAMA network open, 5(8), e2228720-e2228720.

[17] Dang, K., Wang, X., Hu, J., Zhang, Y., Cheng, L., Qi, X., ... & Li, Y. (2024). The association between triglyceride-glucose index and its combination with obesity indicators and cardiovascular disease: NHANES 2003–2018. Cardiovascular diabetology, 23(1), 8.

[18] Austin, A. M., Ramkumar, N., Gladders, B., Barnes, J. A., Eid, M. A., Moore, K. O., ... & Goodney, P. P. (2022). Using a cohort study of diabetes and peripheral artery disease to compare logistic regression and machine learning via random forest modeling. BMC Medical Research Methodology, 22(1), 300.

[19] Han, M., Qie, R., Shi, X., Yang, Y., Lu, J., Hu, F., ... & Zhao, Y. (2022). Cardiorespiratory fitness and mortality from all causes, cardiovascular disease and cancer: dose–response meta-analysis of cohort studies. British journal of sports medicine, 56(13), 733-739.

[20] Ray, K. K., Ference, B. A., Séverin, T., Blom, D., Nicholls, S. J., Shiba, M. H., ... & Santos, R. D. (2022). World heart federation cholesterol roadmap 2022. Global heart, 17(1), 75.

[21] Writing Committee, Lloyd-Jones, D. M., Morris, P. B., Ballantyne, C. M., Birtcher, K. K., Covington, A. M., ... & Wilkins, J. T. (2022). 2022 ACC expert consensus decision pathway on the role of nonstatin therapies for LDL-cholesterol lowering in the management of atherosclerotic cardiovascular disease risk: a report of the American College of Cardiology Solution Set Oversight Committee. Journal of the American College of Cardiology, 80(14), 1366-1418.