

A Data Mining Approach to Identify Heart Disease Risk Factors in a Philippines Healthcare

Abrea, Kyla Trisha D¹, Cainoy, Angie I², Doquila, Marian Sandra S³,
Gonzales, Joyce T⁴

^{1,2,3,4}Student, College of Computing and Information Sciences (CCIS), Surigao Del Norte State University

Abstract

The study wants to explore data mining approach to identify heart disease risk factors in Philippines using a relevant healthcare dataset. Using classification algorithms containing J48, Naïve Bayes, Random Forest, and PART, evaluated a 13 patients attributes to determine their predictive power. The data mining used Random Forest found nine (9) primary factors that notable predictor of heart disease in the datasets. These factors include; (1) Thalassemia/Stress Test results (Thal), (2) Maximum Heart Rate (Thalach), (3) Chest Pain Type (CP), (4) Number of Major Vessels (CA), (5) ST Depression (Oldpeak), (6) Age, (7) Serum Cholesterol, (8) Resting Blood Pressure, and (9) Exercise-induced Angina. Within this factors, heart stress test results (Thal) and maximum heart rate (Thalach) found that it is the most critical clinical risk factors of heart disease. When it comes to model performance, Naïve Bayes reached the highest accuracy of 83.11%, followed by Random Forest that reached 78.75%. These results recommend that data mining is effectively to priority the clinical factors. to allow healthcare providers to focus on the most critical diagnoses makers for early prevention in the Philippines for healthcare context.

Keywords: Heart Disease Factors, Data Mining, Philippine Healthcare Datasets, WEKA (Waikato Environment for Knowledge Analysis), J48 Decision Tree, Naïve Bayes, Random Forest, PART (Partial Decision Tree)

1. Introduction

Heart disease is known for as one of the leading causes of death globally and alarming for public health concern within the Philippines. Several factors can contribute on this case such as unhealthy lifestyles, high cholesterol levels, hypertension and natural aging, this factor leads to cardiovascular diseases. Through early detection these specific risk factors is important to prevent for various complications and improved patient outcomes. This study aims to the healthcare datasets that contains patients' attributes for heart disease, using weka data mining tools pre-process the datasets for quality data analysis, apply the classification algorithms such as j48, Naïve Bayes, Random Forest and PART for building a predictive model, and lastly, determine the most effective algorithms for accurate predicting disease factors

2. Problem Statement

Heart disease is still a big problem in the Philippines, causing a lot of sickness and death. Early detection of heart disease risk factors is crucial, but traditional approaches in analyzing medical data might not be

powerful enough to detect patterns within large health-related datasets. With the ever-increasing volume of data in the healthcare sector, there is a demand for more efficient and accurate analytical methods. The main objective of this study is to overcome this problem by using data mining techniques to predict the risk factors for heart diseases with a healthcare data set. In particular, it aims to find out the classification algorithms (J48, Naïve Bayes, Random Forest, PART) to analyze patient's health attributes and make better prediction of heart disease risk. The study also tries to find out which algorithm gives the best prediction performance using WEKA data mining tool.

3. Objectives

The primary goal of this study is to predict the risk factors for heart disease with the use of data mining techniques on a healthcare database that is relevant to the Philippine population. In particular, the aim of this study is to:

1. To analyze the health care dataset with patient attributes about heart disease.
2. To preprocess the dataset using data mining tool called 'WEKA' to prepare it for data analysis.
3. To use classification algorithms like J48, Naïve Bayes, Random Forest, PART etc. to create predictive models.
4. To test the performance of the classification algorithms with the help of 10-fold cross validation and performance measures.
5. To find the best algorithm to predict factors that can lead to heart disease.

These research questions will be explored in the study:

1. What are the important patient characteristics that are linked to the risk for heart disease in this dataset?
2. How can the data be preprocessed to enhance the quality of the healthcare data?
3. What is the ability of J48, Naïve Bayes, Random Forest and PART to classify heart disease risk?
4. Which classification algorithm gives the best accuracy in predicting the risk factors for heart disease?

4 Scope of the Study

1. The aim of this study is to predict the risk factors of heart disease using data mining technique on a healthcare data relevant to the population of the Philippines.
2. Selected patient characteristics are studied like age, sex, cholesterol level, blood pressure, and other clinical factors associated with heart disease.
3. Use the WEKA data mining tool for data preprocessing, data classification and data evaluation.
4. classification algorithms such as J48, Naïve Bayes, Random Forest and PART are employed to create the predictive model.
5. These algorithms have been tested by 10-fold cross validation and using standard evaluation metrics.

Limitations of the Study

1. The study is based on the data available and may not be generalizable to the entire Philippine population.
2. Only certain patient characteristics within the data set are analyzed; there may be other relevant patient medical or lifestyle characteristics that are not included.

3. Only four classification algorithms, namely J48, Naïve Bayes, Random Forest and PART, are studied.
4. The analysis is then conducted via the WEKA data mining tool and results may differ in other data mining platforms or methods.

5. Related Work

Healthcare data analysis has gained vital importance because medical records take up most of the available data. Researchers use data mining and machine learning techniques to discover information that helps with early diagnosis and disease prevention (Saleh B.J., et al., 2020)[6]. Classification algorithms use blood pressure and cholesterol and age data to predict cardiovascular diseases according to Detrano R[1]. and his fellow researchers who studied this topic in 1989. Researchers prefer decision tree algorithms and Naïve Bayes models because these methods offer simple implementation with straightforward results interpretation but recent studies have shown that medical record processing needs Random Forest ensemble techniques to achieve better results (Haider M., et al., 2025)[5].

The Philippines records diabetes and cardiovascular diseases (CVDs) as its two most common causes of death. Prior research has examined the application of predictive analytics for diabetes and cardiovascular disease risk assessment as well as strategies and obstacles for obtaining or creating datasets that are available in the nation (Sy M.P., et al., 2020)[7]. Recent research shows that disease risk assessment needs data collection which requires access to specific datasets. The Philippines has yet to develop any official plans that will implement predictive analytics for diabetes and CVD risk assessment despite government support for the Philippine National eHealth solution (Sy M.P., et al., 2020)[7].

The Philippines has developed localized solutions such as CAREdio which operates as a mobile-connected screening system for cardiovascular disease prediction that uses machine learning algorithms and hardware sensors to provide accurate results for low-income rural areas in the Philippines (Galarosa M.B., et al., 2020)[8].

Gaps in Current Literature

1. Fewer research studies explore heart disease risk factors in Philippine healthcare systems through data mining methods, while most existing research focuses on predicting general cardiovascular diseases through worldwide databases.
2. The current research on heart disease prediction is limited because most studies evaluate only one classification algorithm performance.
3. The assessment of different predictive model effectiveness in healthcare datasets becomes complicated because some studies fail to compare multiple machine learning methods.
4. The development of healthcare statistics into effective predictive tools for clinical decision support and early patient detection and healthcare management needs additional research.

5. Methodology

This study applies a quantitative data mining methodology to determine heart disease risk factors using a healthcare dataset that relevant to the Philippine citizens. This used a weka tools for open-source platforms for machine learning and data analysis. The methodology followed a step-by-step procedure for collecting a healthcare dataset including 297 patients' instances, it was codified in Microsoft Excel and strict filtration process, including data cleaning, outlier handling, and normalization using a Min-

Max scaling formula. Four machine learning algorithms—J48, Naïve Bayes, Random Forest, and PART— this used to produce predictive models based on thirteen attributes. The models used was assessed through 10-fold cross-validation, using metrics like accuracy and ROC area to pinpoint the most effective method for early heart disease detection.



Figure 1: Data Mining Process Flow of the Study

6.1 Data Collection

Source of Data

The research used data from a publicly available healthcare dataset which contains information about heart disease. The collection includes patient medical data, including characteristics like age, sex, blood pressure, cholesterol, kind of chest discomfort, and other clinical markers linked to the risk of heart disease. The researchers used these characteristics to study medical research which investigates and predicts cardiovascular disease.

Data Collection Methods and Tools

The dataset was downloaded in a structured format appropriate for data mining research after being gathered from an online data source.

The dataset was put into the WEKA data mining program for additional processing once it was acquired. Before being used for preprocessing, classification, and model evaluation in the study, the data was checked for completeness and relevance.

6.2 Data Pre-Processing

Data Cleaning and Preparations

The data mining process requires data preparation because it ensures that data remains clean and stays

consistent while meeting analysis requirements. The research utilized the WEKA data mining tool to prepare the healthcare dataset which researchers used to implement classification algorithms. The dataset preparation process involved three steps which included data cleaning and outlier management and data normalization to improve dataset quality and reliability used for heart disease risk factor prediction. The dataset cleaning process aimed to remove all missing and inconsistent and incomplete information from the dataset. The analysis team assessed attributes to verify their data types and values conformed to legitimate analysis requirements. The team evaluated all records to identify redundant or irrelevant content which they removed when necessary. The dataset underwent cleaning and proper formatting which resulted in ready-to-use data for classification models.

Handling Outliers

Outliers are very large or very small values, which could influence predictive models. In this research, observations that might be considered outliers were checked in the data for cholesterol, blood pressure and age. These values were checked when detected to see if they were valid medical readings or an error. Data quality was ensured by the use of appropriate handling methods like the removal of incorrect values, or the adjustment of values when necessary's

Data Normalization

Data normalization was applied to scale numerical attributes into a consistent range to improve the performance of the machine.

$$X_{norms} = + \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where:

- X represents the original value of the attribute
- X_{min} represents the minimum value of the attribute
- X_{max} represents the maximum value of the attribute
- X_{norm} represents the normalization value

6.3 Data Mining Techniques

The study utilized patient healthcare information to identify heart disease risk factors through data mining classification methods. The supervised learning method of classification establishes predefined categories which it uses to sort incoming data based on its analysis of past records. The research team tested multiple classification techniques on the healthcare dataset through the WEKA data mining software.

The researchers applied four classification algorithms through the WEKA system:

1. J48 Decision Tree: J48 is a tree-based classification technique that classifies data according to attribute values by building a decision tree structure. The method produces straightforward rules which help identify essential elements that contribute to heart disease risk assessment therefore it sees common usage.
2. Naïve Bayes: Naïve Bayes operates as a probabilistic classification system which uses Bayes' theorem as its foundation. The system provides straightforward operation for users while delivering high performance with extensive datasets because it assumes that all attributes function independently.
3. Random Forest: Random Forest employs ensemble learning methods to build multiple decision trees which it combines for better predictive performance. The system effectively handles complex data

patterns while it decreases the chance of developing model overfitting problems.

- The rule-based assessment system known as PART (Partial Decision Tree) uses partial decision trees to develop decision-making guidelines. The system combines rule-based learning with decision trees to produce accurate and understandable classification outcomes.

Table 1: Data Mining Techniques Used

Algorithms	Description	Purpose in the Study
Naïve Bayes	This probabilistic classifier is based on Bayes' theorem	This is used as baseline for comparison
J48 Decision Tree	This implements a java of C4.5 algorithms to generate decision tree	It is used to build a predictive model foe disease factor
Random Forest	Its ensemble learning method that operates by constructing a multiple decision tree	This identified as the most effective model for the clinical dataset's complexity
PART (Partial Decision Tree)	It is rule-based classifier to avoid full free construction by using partial trees	This applied to produce predictive model based on clinical attributes

Table 2: Comparison of Classification Techniques Used in the Study

Algorithms	Accuracy	Interpretability	Handling of Missing Data	Computational Efficiency
Naïve Bayes	Moderate	Moderate	High	Very Fast
Random Forest	Very High	Low	High	Moderate
J48 Decision Tree	High	High	Moderate	Fast
PART Classifier	High	High	Moderate	Moderate

6.4 Tools and Technologies Used

The research utilized various techniques and technologies for its data analysis which required the development of heart disease risk factor predictive models. The research used WEKA because it is an open-source data mining software that enables machine learning and data analysis research according to its creators from Waikato University. WEKA was used to conduct data preprocessing and testing of classification algorithms which included J48 and Naïve Bayes and Random Forest and PART while evaluating the model performance results. The dataset needed to be organized and analyzed through Microsoft Excel before the data could be transferred to WEKA. Excel helped to assess the data structure and verify that the dataset had been properly arranged. A laptop or personal computer with sufficient processing capabilities was needed to operate the WEKA software and perform data mining research. The study used these technologies and techniques to conduct research while creating heart disease risk prediction models which required assessment.

7. Data Analysis

The study employed the WEKA data mining tool to analyze healthcare data which allowed researchers to discover patterns associated with heart disease risk factors. The research team used classification algorithms which included J48 Naïve Bayes Random Forest and PART to analyze patient data that included age and sex and blood pressure and cholesterol and other clinical indicators. The modeling assessment used 10-fold cross-validation to evaluate how well the models performed. The research team evaluated heart disease prediction accuracy for different classification algorithms through assessment metrics which included accuracy and precision and recall and confusion matrix.

Key Insights

1. Age as a Major Risk Factor: The analysis shows that older patients from the dataset have increased heart disease risks.
2. Cholesterol Levels Influence Risk: The study found that heart disease patients had higher cholesterol levels which demonstrated a strong connection between cholesterol and cardiovascular risk.
3. Blood Pressure Impact: The study discovered that high blood pressure or hypertension served as a crucial risk factor for developing heart disease.
4. Chest Pain Type as a Significant attribute.
5. Effectiveness of Classification Models: The data mining methods successfully identified heart disease patterns from the dataset through their various algorithms which included J48 Naïve Bayes Random Forest and PART

8. Result

Datasets of Age

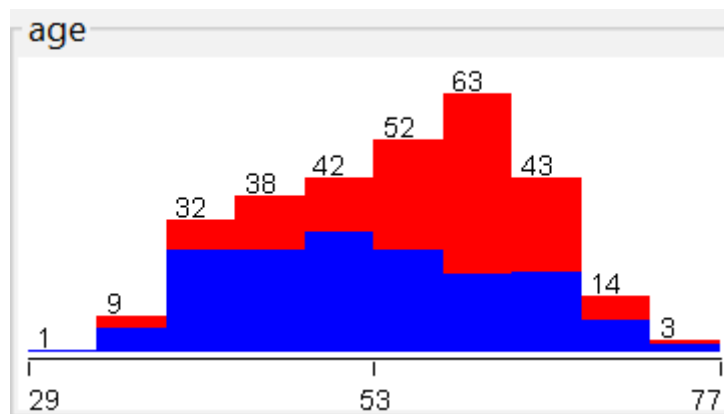


Figure 2: Age Distribution of Patients in the Heart Disease Dataset

This figure shows the distributions of the patients age from datasets that used for heart disease predictions. The histogram presents the frequency of the patients from various age ranges, the colors represent different classification group from datasets. This visualize indicates that most of the patients fall in the middle age to older adult range, that age is important factor associated with heart disease. This graph helps understand how age values are distributed before applying data mining classification models.

Data of Condition

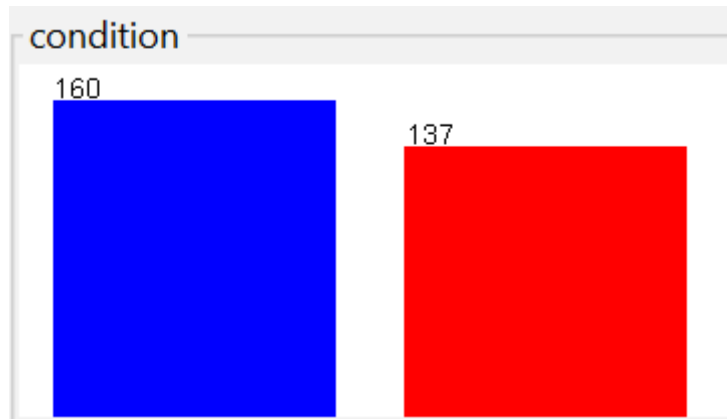


Figure 3: Distribution of Heart Disease Condition in the Dataset

The figure shows the condition of heart disease in the data set used for the analysis. The bar chart displays two groups of patients: those with heart disease (in red) and those without heart disease (in blue). The numbers obtained in the result section suggest a fairly even distribution with 160 patients in one category of condition and 137 patients in the other category. This distribution is crucial for classification modeling because models that are trained using imbalanced data are less reliable and perform poorly when used to predict the risk factors for a heart disease.

Dataset of Sex Distribution

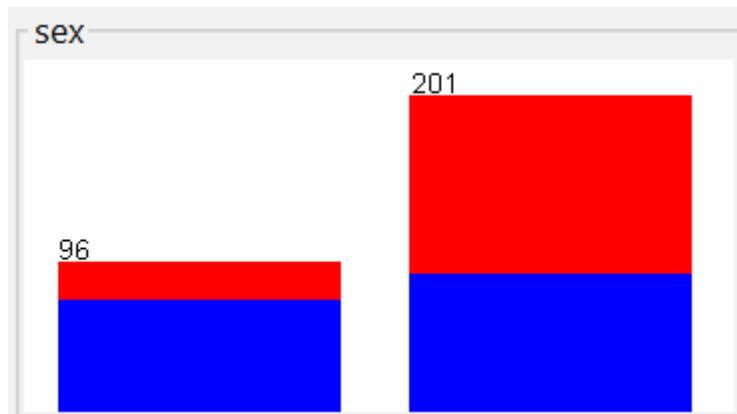


Figure 4: Sex Distribution of Patients in the Heart Disease Dataset

The figure illustrates the distribution of patients based on sex in the dataset used for heart disease analysis. The bar chart shows the number of male and female patients, represented by different colors. The visualization indicates that the dataset contains more male patients than female patients, which may influence the patterns observed in heart disease prediction. Understanding the distribution of sex in the dataset is important because gender is considered a significant factor in cardiovascular disease risk analysis.

Datasets of Cholesterol

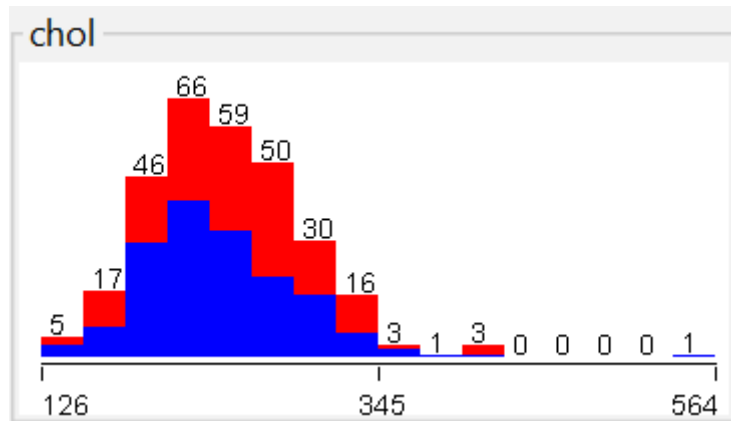


Figure 5: Chest Pain Type Distribution in the Heart Disease Dataset

The figure shows the distribution of chest pain types among patients in the heart disease dataset. The bar chart displays the number of cases for each chest pain category, represented by different colored bars. The visualization helps identify which chest pain types are more common among patients and highlights their potential relationship with heart disease risk. Understanding this distribution is important because chest pain type is a significant clinical indicator used in predicting cardiovascular conditions.

Different Algorithm Classification

The data mining analysis identified significant patterns and clinical markers connected with heart disease risk. The study evaluated four classification algorithms using thirteen clinical attributes across a dataset of 297 instances. The table 1 shown, Random Forest algorithms appeared as the most effective model, it reaches the highest accuracy of 88.96% and ROC area of 0.96%. In addition, the attribute analysis revealed nine primary predictors of heart disease, with heart stress test results and (Thal) and maximum heart rate (Thalach) this two identify as most critical clinical markers for early detection.

Table 3. Performance Evaluation of Classification Models

Algorithms	Accuracy	Precision	Recall	F-Measure	ROC Area
Random Forest	88.96%	0.890	0.890	0.890	0.960
PART Classifier	85.23%	0.856	0.852	0.854	0.923
J48 Decision Tree	85.23%	0.845	0.842	0.843	0.931
Naïve Bayes	78.54%	0.791	0.791	0.788	0.874

In addition, Figures 6–9 present the WEKA visualization of the selected algorithms. This figure based on the results further prove the numerical comparison in table 3



Figure 6. J48 Decision Tree Results

This result of the classification process in the WEKA data mining tool with J48 Decision Tree algorithm is displayed in figure 6. The scatter graph shows how the actual values in the classes compare with the predicted values in the classes for heart disease conditions. These points are instances, with colors varying based on classification categories. The dots along the diagonal are correctly classified and the other points are misclassified. This visualization can be used to evaluate and compare the performance and accuracy of the J48 classification model in predicting heart disease risk factors.

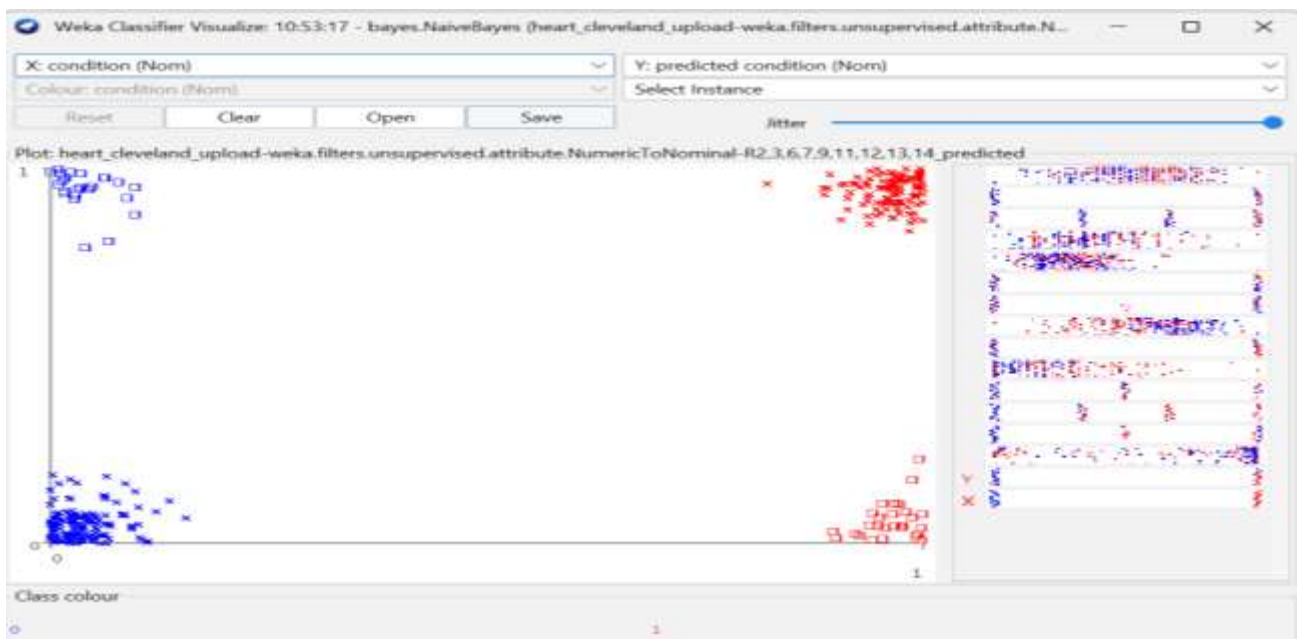


Figure 7. Naïve Bayes Results

The Naïve Bayes algorithm was implemented in the data mining tool WEKA to visualize the classification in this figure. The scatter diagram shows the relationship between the actual class value (X-axis) and the predicted class value (Y-axis) for the class for the heart disease condition. Every point is a patient instance in the data set with the different colors representing classes in the data set. Points coming from a similar area mean a correct classification, and points that are spread out mean that they are misclassified. This visualization gives an idea of how well the Naïve Bayes model performs in predicting the risk factors for heart disease from the healthcare dataset.

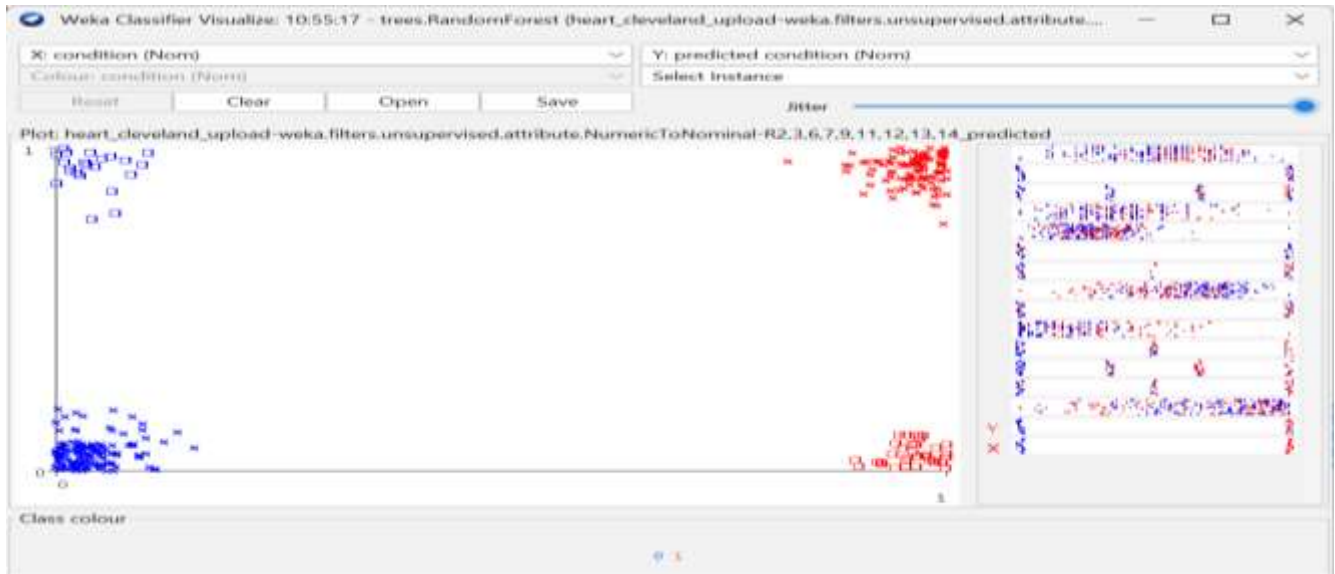


Figure 8. Random Forest Results

The figure shows the classification visualization of the WEKA data mining tool and Random Forest algorithm. The scatter plot shows the actual heart disease state (X-axis) and the predicted heart disease state (Y-axis) from the model. The data points correspond to patient records from the data set, and the colors indicate different class categories. The points are grouped in their respective areas, demonstrating the ability of the Random Forest model to classify numerous instances accurately. A few isolated instances are misclassified cases. This visualization shows how well the Random Forest algorithm works on the healthcare dataset to predict the risk factors associated with heart disease.



Figure 9. PART Classifier Results

This figure shows the classification visualization created by the PART (Partial Decision Tree) rule-based algorithm of WEKA. The graph plots the predicted health outcome of heart disease (Y-axis) against the actual health outcome of heart disease (X-axis). Each point on the plot is a single individual in the data set, the color of the point indicating the class (presence or absence of heart disease). The results demonstrate a good agreement between the blue and red points in their areas, indicating that the PART algorithm performed well in correctly classifying the instances. There are some scattered points outside the main clusters, though, which suggest low levels of misclassification. This indicates that the PART classifier is reasonably accurate at finding patterns in the data set, but it is not quite as accurate as some other ensemble learners, like Random Forest. In general, the visualization shows that the rule base classification would be an effective method for predicting heart disease states from the attributes available.

Interpretations of the Results in the Context of Objectives

This study's main goal was to create and assess a data mining-based classification model that might use WEKA's machine learning techniques to predict cardiac disease. Four classification algorithms—J48 Decision Tree, Naïve Bayes, Random Forest, and PART—were used to accomplish this goal, and their predicted performance and visualization outcomes were evaluated. According to the testing findings, Random Forest outperformed the other algorithms in terms of classification. The majority of cases were accurately identified, with little overlap between anticipated and actual classes, according to the display of classifier errors. This indicates that the model is the most dependable algorithm for predicting heart disease situations because of its high capacity to find patterns and relationships within the dataset. Additionally, the J48 Decision Tree method performed fairly well, as evidenced by the visualization's obvious clustering of correctly identified examples. This implies that the decision tree strategy can successfully extract decision rules from the dataset, but because of the simplicity of its tree structure in comparison to ensemble methods, it may still result in some misclassifications. The Naïve Bayes method, on the other hand, performed quite poorly, as evidenced by its visualization, which shows more cases outside of the primary classification clusters. Based on decision tree principles, the PART rule-

based classifier performed moderately and produced classification patterns that were comparable to J48. The results support the study's goal of identifying the best machine learning algorithm for heart disease prediction. The visualization demonstrates that many instances were correctly predicted, but a few misclassified points suggest that rule-based approaches may still struggle with certain complex patterns in the dataset. The results verify that ensemble learning techniques, such as Random Forest, outperform single-model classifiers in terms of predicted accuracy and robustness. These findings demonstrate how data mining methods can be used in healthcare systems to help with early detection and decision assistance for the diagnosis of heart disease.

9. Key Findings

According to the study's findings, Random Forest outperformed the other studied classification algorithms in terms of predictive ability, exhibiting greater accuracy in identifying cases of heart disease in the Philippine healthcare dataset. Additionally, the J48 Decision Tree and PART algorithms demonstrated dependable performance, accurately categorizing numerous cases and identifying distinct patterns between individuals with and without heart disease. The Naïve Bayes classifier, on the other hand, showed less predictive accuracy. This could be because it assumed that qualities were independent, which might not adequately reflect the intricate interactions found in medical data. Additionally, the investigation showed that a number of clinical characteristics in the dataset made a substantial contribution to the prediction of heart disease risk factors, enabling machine learning models to differentiate between various patient states. Overall, the findings show that data mining methods are useful instruments for evaluating medical data and can aid in the early identification and decision-making of heart disease risk prediction in the Philippine healthcare setting.

10. Implications

The study's findings demonstrate how data mining methods might enhance healthcare analytics and disease prediction in the setting of Philippine healthcare. The effective use of classification algorithms like J48, Naïve Bayes, Random Forest, and PART shows that machine learning models are capable of analyzing patient health data and finding patterns linked to risk factors for heart disease. These results suggest that healthcare organizations can promote early identification and preventative healthcare programs by using predictive models. Medical practitioners can offer prompt treatments, lifestyle advice, and suitable treatment options by identifying individuals who are more likely to develop heart disease. Additionally, the study indicates that the efficiency of medical diagnostic and decision-making processes can be improved by incorporating data mining technology into healthcare information systems. This strategy can help data-driven healthcare solutions that improve patient outcomes and fortify the Philippines' entire health healthcare management system.

11. Limitations

There are several limitations of this study that should be considered when interpreting the findings. First, the analysis was conducted using a limited healthcare dataset which may not fully represent the Philippine population. The dataset may also be lacking some medical or lifestyle variables such as diet, physical activity and family medical history which may affect heart disease risk factors. Second, the study only tested four classification algorithms including J48 Decision Tree, Naïve Bayes, Random Forest and PART using the WEKA data mining tool. Other machine learning techniques or advanced

deep learning models were not included that may generate different or better predictive results. Finally, the quality and completeness of the dataset used for analysis affect the results. Missing, inconsistent or limited data may affect the accuracy and reliability of the predictive models. Therefore, future research can leverage larger and more comprehensive healthcare datasets to improve the robustness of prediction models for heart disease.

12. Conclusions

Conclude your research paper here. To conclude, the accomplished demonstrated the effectiveness of the data mining in identify the specific attributes associated with heart disease risk. By assessing the thirteen patient attributes from various classification algorithms, the research identified that nine specific factors significantly influence heart disease diagnosis. Particularly, **Thalassemia (Thal)**, **Maximum Heart Rate (Thalach)**, **Chest Pain Type (CP)**, **Number of Major Vessels (CA)**, **ST Depression (Oldpeak)**, **Age**, **Serum Cholesterol**, **Resting Blood Pressure**, and **Exercise-induced Angina** that identify as the primary risk predictors. Moreover, the comparative analysis of algorithms revealed that Random Forest provided a imperative insight for the feature importance, Naïve Bayes achieved the strongest predictive accuracy of 83.11%. The integration of rigorous data pre-processing accompanied by classification models such as **J48**, **Naïve Bayes**, **Random Forest**, and **PART** secures a higher standard of healthcare data analysis. The findings provide a data-driven foundation for Filipino healthcare practitioners to improve diagnostic accuracy by prioritizing the nine critical risk factors during the patients screening and for early intervention programs

13. Recommendations

Based on the results of this study, some suggestions can be given for future research and practical application of heart disease prediction by using data mining techniques. First, future studies should look into the use of larger and more diverse datasets of healthcare data, especially from different hospitals or healthcare institutions in the Philippines, to improve the generalizability and reliability of predictive models. Secondly, researchers are encouraged to investigate other machine learning algorithms and advanced techniques such as Support Vector Machines, Neural Networks, or deep learning models to improve the accuracy and performance of heart disease prediction systems. Third, healthcare institutions may consider the integration of data mining and predictive analytics into their healthcare information systems to support early detection of heart disease risk factors and improve clinical decision making. Predictive models can assist medical professionals in identifying high-risk patients earlier and implementing preventive healthcare strategies. Finally, future work should also focus on incorporating other health-related factors such as lifestyle, diet, physical activity, and family history of health to gain a more holistic analysis of risk factors for heart disease and improve the prediction results.

14 References

1. Detrano R., Janosi A., Steinbrunn W., Pfisterer M., Schmid J.J., Sandhu S., Froelicher V., “International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease”, *The American Journal of Cardiology*, 1989, 64 (5), 304–310. <https://www.sciencedirect.com/science/article/pii/0002914989900249>
2. Al-Taie R.R.K., Saleh B.J., Saedi A.Y.F., Salman L.A., “Analysis of WEKA Data Mining Algorithms for Heart Disease Prediction System”, *International Journal of Electrical and Computer Engi*

- neering, 2021, 11 (2), 1735-1743. <https://ijece.iaescore.com/index.php/IJECE/article/view/22232>
3. Ramalingam V.V., Dandapath A., Raja M.K., “Heart Disease Prediction Using Machine Learning Techniques: A Survey”, International Journal of Engineering & Technology, 2018, 7 (2.8), 684-687. https://www.researchgate.net/publication/325116774_Heart_disease_prediction_using_machine_learning_techniques_A_survey
 4. Bhatia N., Singh A., “A Review of Data Mining Techniques for Heart Disease Prediction”, International Journal of Advanced Research in Computer Science and Software Engineering, 2010, 2 (2). https://www.researchgate.net/publication/275973494_Heart_Disease_Prediction_System_using_Data_Mining_Techniques
 5. Haider M., Hussain M., Insany G.P., “Heart Attack Prediction Using Machine Learning Models: A Comparative Study of Naïve Bayes, Decision Tree, Random Forest, and K-Nearest Neighbors”, Engineering Proceedings, 2025, 107 (1), 121. <https://www.mdpi.com/2673-4591/107/1/121>
 6. Saleh B.J., Saedi A.Y.F., Al-Aqbi A.T.Q., Salman L.A., “Analysis of Data Mining Techniques for Heart Disease Prediction Using WEKA”, Library Philosophy and Practice, 2020, 4032. <https://scholarworks.sjsu.edu/libphilprac/4032/>
 7. Sy M.P., Sese A.K.G., “Predictive Analytics for Cardiovascular Diseases and Diabetes Risk Scoring in the Philippines: a Management Perspective”, Current Cardiovascular Risk Reports, 2020, 14 (5). <https://link.springer.com/article/10.1007/s40471-020-00232-7>
 8. Galarosa M.B., Rosel A.M., Sabili M.A.C., “CAREdio: Health screening and heart disease prediction machine learning model”, Proceedings of the 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), 2020, 1–6. <https://doi.org/10.1109/HNICEM51456.2020.9400034>
 9. Burley C.V., Sohrabi H.R., Alshahrani M., Dunne J., Naismith S.L., Anstey K.J., “Country-specific modifiable dementia risk factors across the Western Pacific Region determined by population attributable fraction”, The Lancet Regional Health – Western Pacific, 2025, 55, 101314. <https://doi.org/10.1016/j.lanwpc.2024.101314>