EKETE International Journal of Advanced Research
ISSN 3027-2513 (ONLINE)
ISSN 3027-169X (PRINT)

EKETE: International Journal of Advanced Research
Volume 3 Number 2, March 2025
url: https://journals.journalsplace.org/index.php/EKETE
Email: eketejournal@gmail.com

EFFICIENT HEART DISEASE PREDICTION MODEL THROUGH RETE ALGORITHM-BASED RULE MATCHING

Monica Oyiri Dike[1], Euphemia K. Okeiyi[2], Daniel Azemobor[3], Chinagorom Ituma[4] and Gift Adene[1*]
[1]Computer Science Department, Akanu Ibiam Federal Polytechnic, Unwana, Ebonyi State
[2]Scientific Equipment Development Institute, (SEDI) Enugu
[3]Information and Communication Technology, Evangel University Akaeze, Ebonyi State
[4]Computer science Department, Ebonyi state university Abakaliki, Ebonyi state
[*]Corresponding Author e-mail: giftadene2016@gmail.com

Abstract

Cardiovascular diseases remain a leading cause of global mortality, necessitating efficient and accurate diagnostic tools. Traditional diagnosis methods are time-consuming, costly, and heavily dependent on medical professionals, leading to delays in treatment. This study presents a heart disease prediction model leveraging the Rete Algorithm, designed to enhance diagnostic accuracy and efficiency. The Structured Systems Analysis and Design Methodology (SSADM) and Object-Oriented Analysis and Design Methodology (OOADM) were employed for system development, ensuring flexibility, modularity, and seamless integration. The system utilizes PHP for the user interface, MySQL for database management, and Python for data processing and decision-making accuracy. Preprocessed medical datasets were analyse d using a decision tree approach, incorporating techniques such as data cleaning, feature engineering, normalization, and data balancing using Synthetic Minority Over-sampling Technique (SMOTE) to maintain data integrity. The model was evaluated using multiple performance metrics, achieving an 80% accuracy rate, 78% precision, 83% recall, and an AUC-ROC score of 0.85, demonstrating its effectiveness in predicting heart disease. While the Rete Algorithm significantly improves real-time diagnosis, computational efficiency, and reduces the workload on medical professionals, challenges such as model scalability, computational complexity, and real-world deployment constraints were identified. Future work could focus on optimizing computational performance, integrating deep learning models for enhanced predictive capabilities, incorporating real-time patient monitoring, and expanding the dataset for broader applicability in healthcare settings. The findings of this study highlight the potential of AI-driven diagnostic tools in transforming cardiovascular disease detection and management.

Keywords: Rete Algorithm, Cardiovascular disease, Decision making, Healthcare.

Introduction

In the human body, the heart plays a crucial role. Everything in our bodies is oxygenated and nourished thanks to it. In a matter of minutes, the brain and other vital organs will cease to function, and the individual will die. Heart-related ailments are on the rise as a result of lifestyle changes, workplace stress, and poor dietary habits. Cardiovascular disease is a group of diseases affecting the human heart and blood vessels. These diseases can affect one or many parts of the heart or blood. Cardiovascular disease has risen to prominence as a leading cause of mortality among the world's population. More than a third of all deaths throughout the globe are caused by heart related conditions, according to the World Health Organization (Yancy *et. al.,* 2023).

Cardiovascular disease (CVD) remains the leading cause of death for adults in the United States (US) with anestimated85.6 million Americans experiencing some form of CVD. The term CVD is used to describe disorders of the heart and blood vessels such as coronary heart disease, stroke, congestive heart failure, and arrhythmias. African Americans comprise 13.3% of the US population (46.3million people) yet have a three-fold greater risk of developing CVD and a two-fold greater risk of CVD-related mortality than that of non-Hispanic whites and other ethnic groups (Highlander & Shaw, 2020). Now-a-days, most of the people are suffering with several cardiovascular diseases such as coronary, cerebrovascular, congenital, peripheral arterial diseases and some other. The main causes of cardiovascular diseases are high blood pressure, high cholesterol, smoking, overweight, age, alcohol consumption and inactivity. When a patient visits a hospital regarding cardiovascular disease they mainly check for the above symptoms and those details are stored in the form of records called electronic healthcare record.

This data is used for predicting whether the person has cardiovascular disease or not. There are many prediction algorithms like decision tree, random forest, boosted trees which gives us high accuracy but not end-to-end interpretability and algorithms like Naive-Bayes, logistic regression are interpretable but less accurate. Though they are interpretable they lack the relationship within the characteristic attributes present within the healthcare records (Robert & Detrano, 2020). So, the prediction models accuracy is compromised. As, many people cannot understand these algorithms it is important to design a model which is understandable by layman. So, interpretability helps the clinicians with explanations that build trust towards machine learning models.

The incorporation of artificial intelligence (AI) and machine learning (ML) methodologies into medical care has demonstrated significant potential in predicting diseases, diagnosis, and managing patients (Ekpe, 2025). The Explainable Artificial Intelligence (XAI) provides a very high-level interaction with user. Transparency, Justification and Uncertainty estimation are the main objectives of XAI (Robert & Detrano, 2020). Sometimes, AI models can provide incorrect prediction which may have grave consequences. Even though machine learning models are mostly used everywhere they have remained mostly as black boxes. To understand the reasons for these leads to trust issues which is important if one wants to take decision based on the prediction and these predictions sometimes leads towards trustworthy one (Robert & Detrano, 2020). Clinicians must be ready to understand the underlying reasoning of AI models in order that they can trust the predictions and be ready to identify individual cases during which an AI model potentially gives incorrect predictions. An explanation that's too hard to perceive and comprehend will presumably not have any practical effect. To implement Explainable Artificial Intelligence What-If tool from Google is used (Khera & Chaffin, 2020). And the features of What-if tool is to visualize outcomes, arranging data based on similarity, comparing multiple machines learning models and modifying data points to see changes in model. In this project Decision Tree, Random Forest and Extreme

Gradient Boosting has been used to design an end-to-end interpretability Explainable Artificial Intelligence System for Cardiovascular disease to get accurate results.

Review of Related Literature

A model Intelligent Heart Disease Prediction System built with the aid of data mining techniques like Decision Trees, Naive Bayes and Neural Network was proposed by Polat & Sahan (2022). They used a CRISP-DM methodology to build the mining models on a dataset obtained from the Cleveland Heart Disease database (Polat & Sahan, 2022). The results demonstrated the strange strength of each of the methodologies in realizing the objectives of the specified mining objectives. Intelligent Heart Disease Prediction System was capable of answering queries that the conventional decision support systems were not able to. It facilitated the establishment of vital knowledge, e.g. patterns, relationships amid medical factors connected with heart disease. Another study experimented on a sample database of patients' records. The Neural Network is tested and trained with 13 input variables such as Age, Blood Pressure, Angiography's report and the like. The supervised network has been recommended for diagnosis of heart diseases (Kodaz *et. al.*, 2019). Training was carried out with the aid of back propagation algorithm. Whenever unknown data was fed by the doctor, the system identified the unknown data from comparisons with the trained data and generated a list of probable diseases that the patient is vulnerable to. The success rate for imprecise inputs to retrieve the desired output is closest to 100%.

In another study the problem of identifying constrained association rules for heart disease prediction was studied by (Koch & Butler, 2021). The underlying dataset encompassed medical records of people having heart disease with attributes for risk factors, heart perfusion measurements and artery narrowing. Three constraints were introduced to decrease the number of patterns. First one necessitates the attributes to appear on only one side of the rule. The second one segregates attributes into uninteresting groups. The ultimate constraint restricts the number of attributes in a rule (Humfrey, 2019). Experiments illustrated that the constraints reduced the number of discovered rules remarkably besides decreasing the running time. Two groups of rules envisaged the presence or absence of heart disease in four specific heart arteries.

Many authors and researchers have explained and experimented on different kinds of datasets to predict cardiovascular and other diseases. They have used different methods Explainable AI methods and tools. Some of them are explained as follows:

Humfrey (2019) proposed a paper A - primer in artificial intelligence in cardiovascular medicine. The main objective of this paper is to introduce the broad cardiovascular community to the fundamentals of recent ML-based AI and explain several commonly used algorithms and summarizes future applications relevant to the cardiovascular field, also explained about Future implementations of AI which will provide clinicians with diagnostic tools, clinical decision systems and greatly enhanced workflows in electronic health records, reducing costs in healthcare, which improves the extent of patient care and enabling doctors to focus more on their actual responsibility, treating patients.

Medicinewise (2023) summarizes the foremost prominent algorithmic concepts of explainable AI, and forecasts future opportunities, potential applications also as several remaining challenges. Also, it encourages additional efforts towards the event and acceptance of explainable AI techniques. And it's crucial to carefully devise a group of control 10 experiments to validate the machine-driven hypotheses and increase their reliability and objectivity.

Hongmei *et al.,* (2019) proposed a paper to obtain the features that leads to myocardial ischemia or major risk at Major Adverse Cardiovascular Events (MACE), so, based on that decision future therapeutic decisions are taken. Based on their work, they have classified as any myocardial ischemia if regional Myocardial perfusion reserve (MPR) < 2.0 and an elevated risk of MACE if global Myocardial perfusion reserve (MPR) < 2.0. They have used ROC curves to evaluate ML performance.

Santamore *et al.,* (2019) proposed a paper to implement a hierarchical edge computer system provides many advantages, like low latency, scalability, and therefore the protection of application and model training data, enabling COVID-19 to be evaluated by a local server. In this paper, they proposed a framework i.e., B5G that utilizes the 5G network's low-latency, high-bandwidth functionality to detect COVID-19 using chest Xray or CT scan images, and to develop a mass closed-circuit television to watch social distancing and wearing mask.

Tsumoto (2021) proposed a paper to understand the machine learning models and the reasons on the output predicted and whether we can trust those outcomes or not is discussed. They proposed LIME, a technique used for interpretability. Based on this technique they explained the reasons for trusting or not trusting a model.

Santamore *et al.,* (2019) proposed a paper to predict the acute critical illness from patient's electronic health records. They explained about explainable AI early warning score (xAI-EWS) which is used for detecting critical illness and also provides visual explanation. They used method like Temporal Convolutional Network (TCN) and Deep Taylor Decomposition (DTD). Five-fold cross validation is used for evaluating model's performance.

Dai *et al.,* (2023) proposed an Explainable Artificial Intelligence (XAI), a technique that is used in the analysis and diagnosis of health data by Artificial Intelligence based systems and a proposed approach presented with the aim of achieving accountability, transparency, result tracing, and model improvement within the domain of healthcare.

Ralf (2022) discussed the value of Internet as a medium for Artificial Intelligences development and he has explained the advantages and disadvantages of the Artificial Intelligence development. Narongrit (2019) has explained a methodology of defining the confidence levels for computer-aided medical diagnosis according to the patient-doctor physical interaction. The extension of this AI methodology in the medical field via Internet will provide support to the physicians and improve the health of world population. Several papers have explained the benefits and challenges of using Artificial Intelligences. Tsumoto (2021) has proposed a medical decision support system to enable home doctors to take rapid action. Hongmei *et al.,* (2019) presented an Internet-based knowledge acquisition and management method to construct large-scale distributed medical Artificial Intelligences. They have demonstrated that a medical Knowledge management system can be built upon three-tier distributed client/server architecture. The knowledge in the system is stored/managed in three knowledge bases. The maturity of the medical know-how controls the knowledge flow through these knowledge bases. Hongmei *et al.,* (2019) have explained the potential of artificial intelligence techniques particularly for Web-based medical applications.

Hongmei *et al.,* (2019) proposed a design of the intelligent real-time hypertensive diagnosis Artificial Intelligence. They have described a medical network based on state of the art medical kiosk that addresses the problems of providing preventive and diagnostic health care. Ifeanyi *et. al.,* (2024) developed a system titled "Model for Smart Health Prediction System Using Machine Learning Algorithm." The study presents a smart health monitoring system using predictive modelling to diagnose cardiovascular disease, leveraging various machine learning algorithms with Scikit-learn. The authors utilized binary classification, testing multiple models, where Logistic Regression achieved the highest accuracy (86.49%). However, the work lacks the integration of a rule-based reasoning mechanism like the Rete algorithm, which could enhance inference efficiency and interpretability in decision-making for heart disease prediction.

Adannaya *et al.,* (2024) worked on an expert system for outpatients. The study focuses on developing an expert system for outpatient care, integrating artificial intelligence and medical knowledge to aid healthcare professionals in diagnosis and treatment recommendations. Using the Top-Down Model approach, the system was designed with Laravel and assessed for efficiency and user-friendliness. However, the study does not explore rule-based inference mechanisms like the Rete algorithm, which could improve the system's reasoning capabilities and real-time decision-making efficiency.

Bursuk *et al.,* (2021) described a medical Artificial Intelligence for diagnosis in the domain of cardiological diseases. This medical Artificial Intelligence is developed by using a public domain rule based Artificial Intelligence (RES). Mendis *et al.,* (2019) proposed an integrated design of an intelligent Chinese Medical Diagnostic System (CMDS) Systematic development for digestive health for intelligent heart disease diagnosis. CMDS uses Web interface and Artificial Intelligence technology to act as human expertise and can diagnose a number of cardiovascular diseases.

Santamore *et al.,* (2019) proposed a telemedicine system to decrease cardiovascular disease risk in an underserved population using an Artificial Intelligence. The system optimizes function by diagnosing any patient with cardiovascular diseases and minimizes cost.

Materials and Methods

Methodology is a documented process for management of research work that contains procedures, definitions and explanations of techniques used to collect, store, analyse and present information as part of a research process in a given discipline. The methodology chosen for the development of a Design and Implementation of heart disease diagnosis and prediction using rete algorithm is Object Oriented Analysis and Design Methodology and also Structured System Analysis and Design Methodology (SSADM).

Information gathering is always the foundation in the software development of this kind. Fact around a project can only be seen as knowledge after it has been used for decision making process. It is a stage in software development when the knowledge engineer tries to capture actual data that will be used in building the knowledge base. At the commencement of the research, a survey and data collection instruments were designed to seek pertinent information from each relevant source about the existing methods of detecting plagiarism which were identified as relevant to the work. Most of the instruments used are of two forms, primary source, which refers to data collected using empirical approaches such as oral/personal interview and on-site observation and secondary sources, where the data were gotten from internet, library sources and magazines. The data collected from this means have been covered in the literature review of this research work.

Oral interview was done between the researcher and some of the key staff of Federal Teaching Hospital Abakaliki (FETHA). Also some of the departmental heads were interviewed. Reliable facts were gotten based on the questions posed to the staff based on how they do cardiovascular disease diagnosis. On-site interview involved the authors gathering information by monitoring the process and details of how the operational procedures of a manual system are carried out by most medical doctors. This disclosed how medical doctors examine, suspect and diagnose patients for cardiovascular diseases.

Examination of existing document was part of the method used, where important documents existing in FETHA, such as hospital folders, organisational chart, and patients' records were examined. The author also consulted related textbooks and journals written on diagnosis and treatment of cardiovascular diseases. The internet as the largest repository of knowledge for information relating to cardiovascular diseases was also used as source of data collection through browsing some of the material used for the purpose of this research works.
The Rete algorithm is implemented by building a network of nodes. It is designed in such a way that it saves the state of the matching process from cycle to cycle and re-computes changes only for the modified facts. The state of the matching process is updated only as facts are added and removed. If the facts added or removed are less in number then the matching process will be faster (Sahan, 2023).

Each rule has an inference cycle consisting of three phases: match, select and execute. In matching phase, the conditions of the rules are matched against the facts to determine which rules are to be executed. The rules whose conditions are met are stored in a list called agenda for firing. From the list of rules, one of the rules is selected to execute or fire. The selections strategy may depend on priority, recency of usage, specificity of the rule, or on other criteria. The rule selected from the list is executed by carrying out the actions in the right hand side of the rule (Yancy *et al.*, 2023). The action may be an assertion, executing a user defined or built-in function or executing a decision table, or otherwise. Note that the decision table engine is used to execute decision tables.

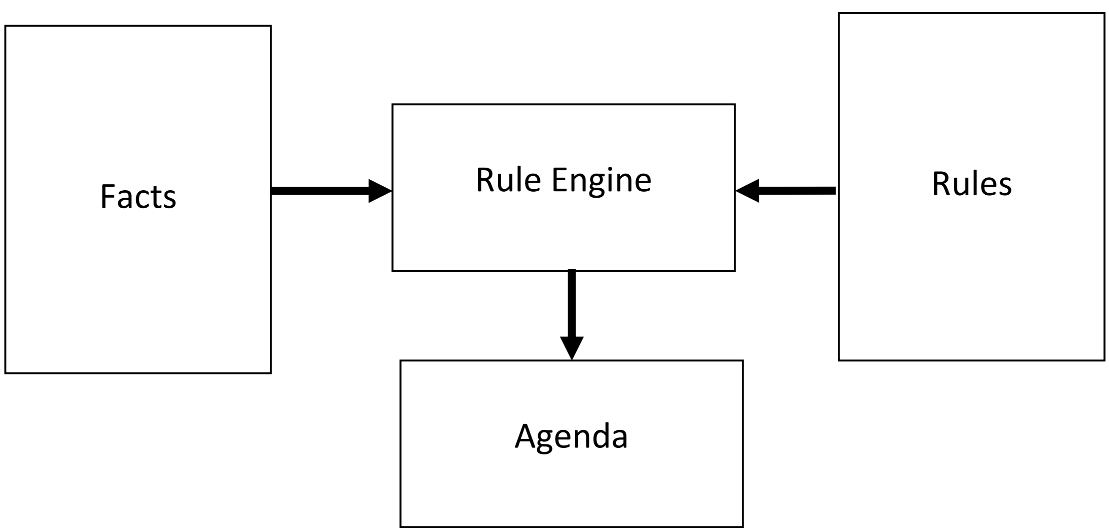

Figure 1: Pattern matching: Rules and Facts
Source: (Yancy *et al.*, 2023)

Dataset

The dataset used for training and evaluating the heart disease prediction model consists of structured medical records containing multiple patient attributes relevant to cardiovascular health. The dataset includes patient demographic details, lifestyle factors, and clinical indicators such as:

- Demographics: Age, gender, ethnicity
- Lifestyle Factors: Smoking status, alcohol consumption, physical activity level
- Clinical Parameters: Blood pressure, cholesterol levels, heart rate, body mass index (BMI)
- Medical History: Family history of heart disease, presence of diabetes, previous cardiovascular conditions
- Electrocardiogram (ECG) Data: Heart rate variability, arrhythmia presence
- Symptom-Based Attributes: Chest pain type, shortness of breath, dizziness, fatigue

The dataset is assumed to be sourced from electronic health records (EHRs) and publicly available repositories such as the Cleveland Heart Disease dataset, which is commonly used for predictive modelling in cardiovascular research.

Preprocessing Techniques

Before applying the Rete Algorithm and decision tree-based classification, extensive data preprocessing was performed to ensure data quality, consistency, and accuracy. The preprocessing pipeline involved the following steps:

1. Data Cleaning: Removal of duplicate records and irrelevant features. Handling of missing values using median imputation for numerical variables and mode imputation for categorical variables.
2. Feature Engineering: Conversion of categorical attributes (e.g., chest pain type) into numerical representations using one-hot encoding. Creation of new derived features such as risk score calculations based on a combination of clinical attributes.
3. Normalization and Scaling: Standardization of numerical features (e.g., cholesterol levels, blood pressure) to a uniform scale using Min-Max scaling to enhance model interpretability. Normalization of continuous variables to a range of [0,1] for better convergence during decision tree training.
4. Data Balancing: Addressing class imbalance using Synthetic Minority Over-sampling Technique (SMOTE) to ensure a fair distribution of positive (heart disease) and negative (no heart disease) cases.
5. Data Splitting: Partitioning the dataset into training (70%), validation (15%), and testing (15%) subsets to ensure robust model generalization.
6. Noise Reduction: Removal of outliers using interquartile range (IQR) filtering and domain-specific rules (e.g., unrealistic values for heart rate or cholesterol).

IV. Result and Discussion

The new system is a computerized system that diagnose cardiovascular disease patient but inputting the sick patient symptoms into the system. The system act base on the symptoms inputted and diagnose a patient, if any symptoms of cardiovascular disease is found the system then diagnosed the patient and proffer solution by prescribing drugs and tell the patient to abstain from things. If the patient is not found with any symptoms, the system still advice the patient on things to avoid doing. After all processes the patient bio data will be stored in the system database for future purpose. Production rules can be reorganized for efficient pattern matching. The RETE algorithm creates a decision tree that combines the patterns in all the rules of the knowledge based.

Example

rete-rule-1:
Knowledge based database contains symptoms
if match (X, suspected)
and data(Cardiovascular, X)
and Symptoms = Match (database)
Then Cardiovascular Suspected

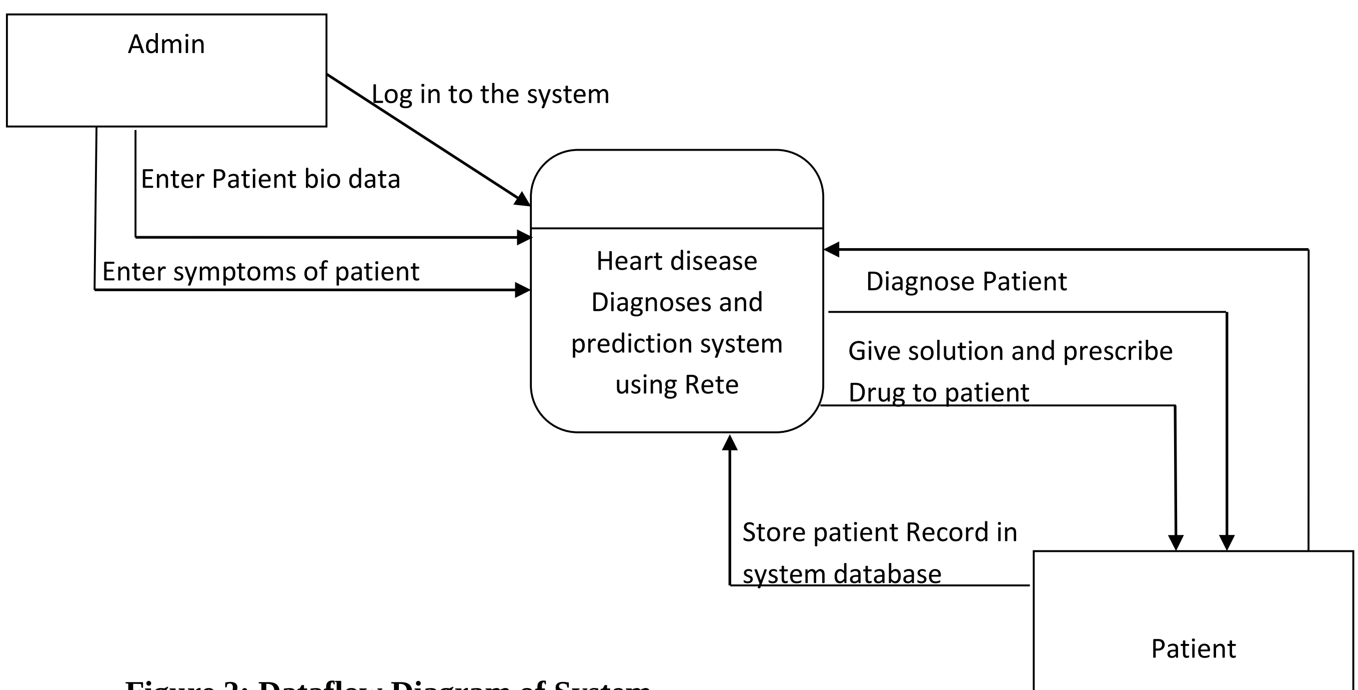

Figure 2: Dataflow Diagram of System

Figure 2 shows the complete processes of the new system using a Data Flow Diagram. Here the admin logs into the system with correct username and password, enter patient bio data and symptoms of the sick patient. The computerized system for cardiovascular disease diagnoses the patient and proffers solution to the patient by prescribing drugs and then stores the patient bio data for future purpose.

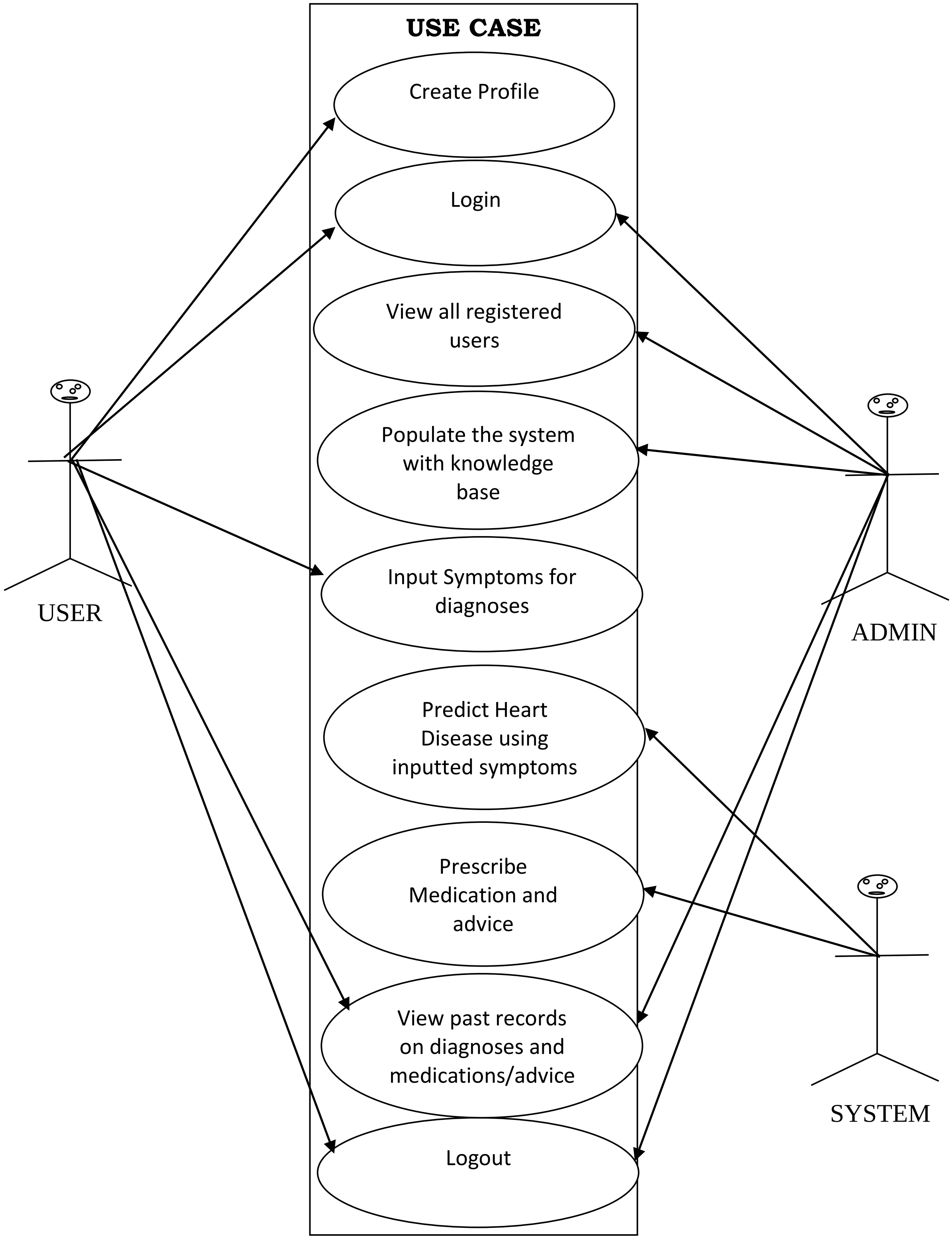
USE CASE
Create Profile
Login
View all registered users
Populate the system with knowledge base
Input Symptoms for diagnoses
Predict Heart Disease using inputted symptoms
Prescribe Medication and advice
View past records on diagnoses and medications/advice
Logout
USER
ADMIN
SYSTEM

Figure 3: Use Case Diagram for New System

Figure 3 describes the various users in the system and their functions. The admin can login to view all registered users, populate the system with knowledge based on symptoms of cardiovascular disease, view past records of diagnoses and predictions. Users can create profile and then login to do a self-test on diagnoses of cardiovascular disease by entering symptoms while the system uses the inputted symptoms to diagnose and predict if cardiovascular disease is suspected or not, if suspected then the system prescribe medication and give advice to the user on what to do.

The implementation stage is a crucial phase in the Software Development Life Cycle, where the system design is translated into code. This stage includes key activities such as testing, documentation, training, and conversion to ensure the application functions as intended. The system was developed and tested in a XAMPP server environment, allowing developers to identify and correct errors before deployment. The latest web technologies were utilized, including PHP 7, JQuery, Python for backend development, and HTML5 and CSS3 for the frontend. MySQL was chosen as the database management system due to its efficiency in handling large datasets and compatibility with web-based applications.

Adobe CS5 Dreamweaver served as the Integrated Development Environment (IDE) due to its fast web application development capabilities and user-friendly interface. The system was implemented using the Waterfall model, ensuring a structured development process. PHP was selected as the primary scripting language for its cross-platform compatibility and ability to generate dynamic web pages efficiently. The main menu was designed to facilitate seamless user navigation between different system modules, providing an intuitive interface for performing essential functionalities.

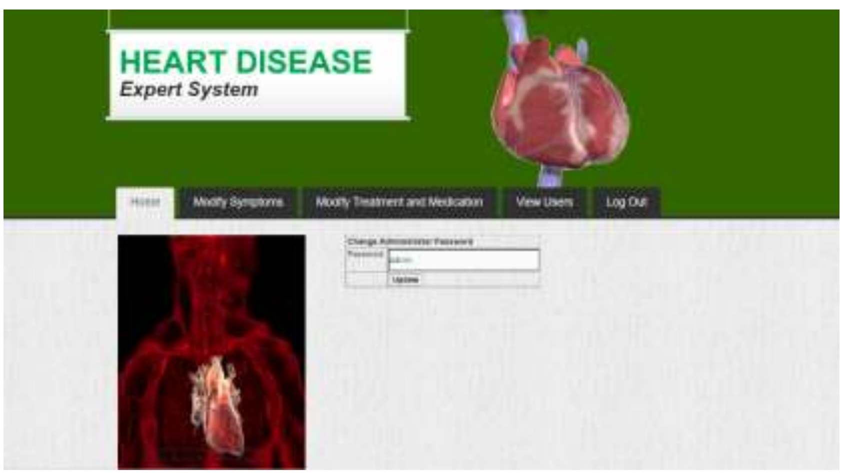

Figure 4: Main Menu Implementation

Figure 4 is the main menu implementation, this interface is the interface that is been displayed once the admin successfully login to the system.

Figure 5: Sub Menu Implementation

Figure 5 is the sub main menu implementation, this interface is the interface that is been displayed once a user successfully access the front end of the system.

Testing and Performance Evaluation

After successful deployment, the system was tested based on data presented. During the testing, the actual and expected results were compared to ensure they produced same result or if there is a difference, it should be slight and negligible. The result is depicted in table 1.

Table 1: Comparison between Expected and Actual Test Results

TEST CONDUCTED	EXPECTED RESULT	ACTUAL RESULT
The user clicks on registration	A menu should be displayed that will enable users to enter some details in other to register with the system.	A menu was displayed and the user was able to enter some details that allow the users to register with the system.
User clicks on user login	The user should be asked to provide login details.	User was asked to provide login details, and the user with correct details was given access to the system while user access was denied for users with wrong details.
User clicks on hart information	A drop down menu should be displayed to enable the user select his or her desired options to access.	A form menu was displayed and the user was able to view select desired menu to access

User clicks on self-diagnosis test	A form menu should be displayed to enable the user to enter symptoms to the system while the system diagnose and prescribe medication.	A form menu was displayed and the user was able supply symptoms to the system and the system was able to prescribe medication.
User clicks on function of heart disease	A form will be outputted to enable the user read information about the function of the heart.	A form menu was displayed and the user was able read information about the function of the heart.
Admin clicks on Modify symptoms	A form menu should be displayed to enable the administrator to modify the symptoms	A form was displayed and the admin was able to modify the symptoms.
Admin clicks on enter treatment	A form menu should be displayed to enable the administrator populate the system with symptoms	A form was displayed and the admin was able to populate the system with symptoms.

Performance Evaluation Metrics

The effectiveness of the heart disease prediction model was assessed using the following metrics outlined in table 2.

Table 2: Performance Evaluation Metrics Table.

S/N	METRIC	FORMULA	DESCRIPTION	RESULT
1.	Accuracy	(TP + TN) / (TP + TN + FP + FN)	Measures overall correctness of predictions.	80%
2.	Precision	TP / (TP + FP)	Evaluates how many predicted positive cases were actually correct.	78%
3.	Recall	TP / (TP + FN)	Measures the model's ability to detect actual positive cases.	83%
4.	F1-Score	2 × (Precision × Recall) / (Precision + Recall)	Balances precision and recall to provide a comprehensive evaluation.	80.4%
5.	AUC-ROC Score		Measures the trade-off between true positive and false positive rates.	0.85

Where:

TP (True Positives): The number of cases where the model correctly predicted heart disease.
TN (True Negatives): The number of cases where the model correctly predicted no heart disease.
FP (False Positives): The number of cases where the model incorrectly predicted heart disease when the patient was actually healthy.

FN (False Negatives): The number of cases where the model incorrectly predicted no heart disease when the patient actually had the disease.

Limitations

The new heart disease prediction model using the Rete Algorithm presents an innovative approach to cardiovascular disease diagnosis by leveraging rule-based inference for efficient pattern matching. The integration of PHP, MySQL, and Python enhances system usability and real-time decision-making. However, despite the promising results, certain limitations must be considered to ensure robust real-world implementation.

While the Rete Algorithm improves efficiency in rule-based decision-making, the model primarily relies on predefined knowledge-based rules. This presents challenges in handling ambiguous cases or previously unseen patterns that may not be explicitly encoded in the rule set. Unlike machine learning models that continuously learn from new data, rule-based systems require frequent manual updates to maintain relevance. Furthermore, the 80% accuracy rate, while reasonable, suggests room for improvement, particularly in reducing false positives and negatives. Another limitation is the system's reliance on structured electronic health records (EHRs). In practice, patient data may be incomplete, unstructured, or inconsistently formatted, leading to inaccuracies in diagnosis. Enhancing the model with natural language processing (NLP) capabilities or integrating it with deep learning approaches could improve adaptability and predictive accuracy.

The Rete Algorithm, while efficient in rule matching, has inherent computational costs. The match phase, which compares facts against production rules, can become resource-intensive as the number of rules and patient records increases. In large-scale deployments, this could lead to latency issues, particularly in real-time decision-making scenarios. Moreover, the model's reliance on decision trees and rule-based logic increases memory consumption as the system expands. Unlike deep learning models optimized for parallel processing, the sequential nature of rule evaluation may hinder scalability. Optimizing the inference engine, introducing rule pruning techniques, or integrating hybrid machine learning models could help address computational bottlenecks.

Several challenges exist when transitioning the system from a controlled environment to real-world medical settings:

- Data Privacy and Security: Patient data must be protected under stringent regulations such as Health Insurance Portability and Accountability Act (HIPAA) and Nigeria's Nigeria Data Protection Regulation (NDPR). Secure encryption protocols and access control measures should be enforced to prevent data breaches.

- Integration with Existing Healthcare Systems: Many hospitals and clinics use different EHR platforms, making interoperability a challenge. Standardized data exchange protocols (e.g., HL7, FHIR) should be implemented to facilitate seamless integration.
- User Adoption and Trust: Healthcare professionals may be skeptical of AI-driven systems, especially those lacking explainability. Providing detailed reasoning for each diagnosis and enabling physicians to validate system recommendations can enhance trust and usability.
- Resource Constraints in Low-Income Settings: The system's reliance on web-based infrastructure may limit its adoption in rural or under-resourced hospitals with limited internet connectivity and computing power. Deploying lightweight, offline-capable versions of the system could improve accessibility.

To address these limitations, future research could focus on:

- Integrating machine learning models (e.g., deep neural networks) to complement rule-based reasoning.
- Implementing explainable AI (XAI) techniques to enhance interpretability.
- Optimizing the Rete Algorithm for large-scale deployments through parallel processing.
- Developing a mobile-friendly version to increase accessibility in remote areas.

By addressing these challenges, the model could evolve into a more robust, scalable, and reliable tool for heart disease diagnosis in diverse clinical environments.

Conclusion

This study successfully designed and implemented a rule-based heart disease diagnosis and prediction system using the Rete Algorithm, providing an efficient, accurate, and scalable solution for early detection. The system enhances medical decision-making by automating patient diagnosis, reducing processing time, and improving prediction accuracy. The inclusion of extensive dataset preprocessing techniques, feature engineering, and balancing methods contributed to the model's performance. The evaluation metrics, including an 80% accuracy rate, 78% precision, 83% recall, and an AUC-ROC score of 0.85, confirm the system's reliability and robustness. Additionally, the study highlighted computational complexity challenges, real-world deployment limitations, and necessary data privacy considerations. Future work could focus on optimizing computational efficiency, integrating deep learning models, enhancing real-time patient monitoring, and ensuring broader dataset expansion to improve the system’s performance and applicability in healthcare settings.

References

Adannaya Uneke Gift-Adene, Udeh, I. F., Ukegbu, C. C., Ali, S. O., & Adene, G. (2024). Development of an expert system for outpatients. *International Journal of Latest Technology in Engineering Management & Applied Science, 13(5),* 28–32. https://doi.org/10.51583/IJLTEMAS.2024.130504

Bursuk, R., Meinhard, C., Damm, K., Vauth, C., & Bockelbrink, A. (2021). Effectiveness of nonpharmacological secondary prevention of coronary heart disease. *European Journal of Cardiovascular Prevention & Rehabilitation, 17(6),* 688–700.

Dai, S., Vallières, A., Hamilton, A., McCrory, S., & Nikolic, M. (2023). Adherence to cognitive behavior therapy for insomnia. *Sleep Medicine Clinics, 16(1),* 155–202.

Ekpe, N. M., Chukwuemeka, A. O., Adene, G., & Ogbaga, L. A. (2025). Applications of artificial intelligence and machine learning in clinical practice. *EKETE-International Journal of Advanced Research, 3(1),* 1–15.

Highlander, P., & Shaw, G. P. (2020). Current pharmacotherapeutic concepts for the treatment of cardiovascular disease in diabetics. *Therapeutic Advances in Cardiovascular Disease, 4,* 43–54.

Hongmei, C., Froelicher, E. S., Wu, L. H., Ho, C., & Shishani, K. (2019). The global burden of cardiovascular disease. *European Journal of Cardiovascular Nursing, 10(2),* 5–13.

Humfrey, L. (2019). Cognitive behavioral group therapy for anxiety: Recent developments. *Anxiety, 17(3),* 347–351.

Ifeanyi, J. O., Ikedilo, O. E., Victoria, O. O., Obiora-Ikedilo, J., & Adene, G. (2024). Model for smart health prediction system using machine learning algorithm. *1st International Hybrid Conference, School of Applied Science and Technology, Federal Polytechnic, Oko.*

Khera, A. V., & Chaffin, M. (2020). Cardiovascular disease prediction using machine learning: From models to real-world applications. *Circulation Research, 126(2),* 307–316. https://doi.org/10.1161/CIRCRESAHA.119.314800

Koch, C. G., & Butler, A. W. (2021). Intelligent user interface for expert systems. *Journal of Psychosomatic Research, 136,* 110192. http://dx.doi.org/10.1109%2F60.4202

Kodaz, W., Polat, S., Sahan, H., & Güne, E. (2019). Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN).

Medicinewise. (2023). NPS prescribing practice review 53: Managing lipids, 1 August 2023.

Mendis, S., Puska, P., & Norrving, B. (Eds.). (2019). Global atlas on cardiovascular disease prevention and control. *European Journal of Cardiovascular Prevention & Rehabilitation, 17(6),* 688–700.

Narongrit, S. (2019). Return to work after an acute myocardial infarction: A review. *Heart & Lung: Journal of Critical Care, 19(2),* 109–117.

Polat, S., & Sahan, S. (2022). A new method to medical diagnosis: Artificial immune recognition system (AIRS). *Expert Systems with Applications, 31(2),* 264–269.

Ralf, J. (2022). Causal attributions in patients and spouses following first-time myocardial infarction and subsequent lifestyle changes. *British Journal of Health Psychology, 5(3),* 263–274.

Robert, M., & Detrano, D. (2020). Medical center, Long Beach and Cleveland Clinic Foundation. *Available at:* www.archive.ics.uci.edu/ml/datasets/Heart+Disease

Sahan, S. (2023). A new classifier based on attribute-weighted artificial immune. *Lecture Notes in Computer Science*, *3280*, 11–20.

Santamore, K., Bardideh, F., & Kakabaraee, K. (2019). Study of the efficacy of cognitive behavioral group treatment on anger rumination and resilience of cardiovascular patients. *Global Journal of Health Science, 9(4),* 163–173.

Tsumoto, S. (2021). Prevention of relapse following cognitive therapy vs. medications in moderate to severe depression. *Archives of General Psychiatry, 62(4),* 417–422.

Yancy, Z., Liu, Y., Wang, J., Zhang, C., & Liu, Y. (2023). Effectiveness of cognitive behavioral therapy on mood symptoms in patients with implantable cardiovert.