COMPUTATIONAL INTELLIGENCE

MCTA 3371

MINI-PROJECT

INTELLIGENT HEART RISK PREDICTION USING COMPUTATIONAL INTELLIGENCE

SEMESTER 1 24/25

LECTURERS:

DR AHMAD JAZLAN BIN HAJA MOHIDEEN

DR AMIR AKRAMIN BIN SHAFIE

NAME	MATRIC NO
AHMAD HAFIZULLAH BIN IBRAHIM	2216185
AMIRUL AIZAD BIN AMIR	2211263
ANAS BIN BADRULZAMAN	2219945
ZAMIR MUTTAQIN BIN MUHAMMAD BALYAN	2212985
ETHAR OSMAN	2111282

TABLE OF CONTENTS

<u>INTRODUCTION</u> -

Cardiovascular disease (CVD) is a global health issue and the leading cause of death worldwide. This kind of disease mainly refers to the conditions of blocked or narrowed blood vessels which results in a stroke, chest pain, angina, or heart attack. These conditions are usually associated with a build-up of fatty deposits inside the arteries and an increased risk of blood clots. According to the World Health Organization, an estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Of these deaths, 85% were due to heart attack and stroke. Although CVD is one of the main causes of death, it can be prevented by The adoption of artificial intelligence (AI) approach, which is rapidly taking hold of this new generation to evaluate patient risks and predict the outcome of CVD.

Computational intelligence (CI) is a subfield of artificial intelligence (AI) that focuses on developing models to enable intelligent behaviour in complex environments. Machine learning (ML) is a popular approach used in medical research to automate data, as well as identify patterns and make decisions. It integrates and interprets complex data in scenarios where traditional statistical methods may not be able to perform. Thus, it is important to detect cardiovascular disease as early as possible using soft-computing techniques for early interventions and preventive measures.

Soft computing develops intelligent machines capable of handling uncertainty and imprecision. Techniques like Fuzzy Logic, Genetic Algorithms (GA), and Artificial Neural Networks (ANN), along with hybrid approaches such as Fuzzy-GA, GA-NN, and Neuro-Fuzzy, have shown great potential in improving accuracy and reliability. By integrating data from diverse sources, these hybrid systems offer valuable insights into cardiovascular risk, helping healthcare professionals assess the likelihood of heart problems.

OBJECTIVES

This research aims to create an intelligent system that can use a collection of health-related characteristics to forecast an individual's risk of getting a heart disease. This broad objective can be divided into several more focused goals:

1. Create a Model for Prediction

Develop a model that evaluates medical data and forecasts the risk of heart disease by utilizing soft computing approaches. Accuracy, interpretability, and computing efficiency should all be considered for a balanced model.

2. Examine Hybrid Methods

Examine hybrid approaches like Neuro-Fuzzy and Fuzzy-GA to improve predicted accuracy and dependability by combining the advantages of several approaches.

3. Preparing and Analyzing Data

Health datasets should undergo thorough preparation to guarantee their consistency, quality, and modelling applicability. This covers exploratory data analysis, data normalization, and handling missing values.

4. Validation and Assessment of the Model

Assess the created models' performance using metrics like accuracy, precision, recall, and F1-score, as well as rigorous validation approaches like holdout validation and cross-validation.

5. Development of Interactive Systems

To make the system accessible and easy to use, develop a graphical user interface (GUI) that allows for real-time risk assessment and result presentation.

<u>METHODOLOGY</u> -

Data Source

This project utilized datasets from two prominent sources:

1. **Kaggle Heart Attack Prediction Dataset:**

   This dataset includes health-related attributes such as age, cholesterol levels, resting blood pressure, and exercise-induced angina, providing a comprehensive basis for predicting heart disease risk.

2. **ScienceDirect Heart Disease Prediction Dataset:**

   This dataset complements the Kaggle dataset by offering additional attributes and insights into cardiovascular health.

Analysis Exploratory Data (EDA)

Exploratory data analysis was carried out to have a thorough grasp of the datasets. Important actions included:

1. **Visualization of Feature Distributions:**

   The distribution of specific features, like age and cholesterol levels, was investigated using density plots and histograms. This made it easier to spot patterns and irregularities in the data.

2. **Correlation Analysis:**

   To determine the correlations between variables, a correlation matrix was created. Heart disease risk was strongly correlated with characteristics such as cholesterol levels and ST depression.

Normalization of Data

To standardize features with different scales, such as blood pressure and cholesterol levels, normalization procedures were used. To ensure equitable representation in the model, min-max scaling was used to rescale these characteristics to a consistent range of [0, 1].

Managing Missing Values

Using imputation techniques, missing data was handled. Mode values were used to impute categorical data, and median values were used to fill continuous variables. This maintained the dataset's consistency and completeness.

IMPLEMENTATION & SIMULATION -

Two main components are used in the implementation of the heart disease risk prediction model: a neural network (NN) and a fuzzy logic system. By combining these elements, a hybrid model is created that combines the interpretability of fuzzy logic with the flexibility of neural networks. In order to make the system accessible and easy to use, the model also includes a graphical user interface (GUI). The main implementation procedures and methods are covered below.

System Implementation for Fuzzy Logic

The purpose of the Fuzzy Logic System (FLS) is to predict the risk of heart disease by managing uncertainties and offering an interpretable reasoning process. The MATLAB fuzzy inference system (FIS) is used to implement the system. Important elements of the FLS consist of:

1. **Variables in the Input**

   Age, cholesterol, heart rate, stress, smoking status, and family history of heart disease are the six input variables that the FLS takes into account. Language phrases like "Young" or "Old" for age and "Low," "Normal," or "High" for heart rate and cholesterol are used to map these inputs to the proper ranges.

2. **Functions of Membership**

   The mapping of each input variable to its linguistic words is defined by membership functions. For example:

- A trapezoidal function is used to map age into "Young" and "Old."
- Triangle membership functions are used to categorize cholesterol into "Normal," "Medium," and "High" levels.
- Values for smoking and family history are mapped to "Yes" or "No.

3. **Base of Rules**

   Twenty if-then rules that are based on medical knowledge make up the rule basis used by the FLS. For instance:

- Risk is high if you are young, have high cholesterol, and are under a lot of stress.
- Risk is low if you are old, have a low heart rate, and don't smoke.

  The decision-making process is transparent thanks to these guidelines.

4. **Mechanism of Inference and Defuzzification**

   The overall risk is calculated by combining rules using the Mamdani inference approach. By defuzzifying the FLS's output, which has a range of 0 to 1, the risk is classified as either "Low" or "High."

Implementation of Neural Networks

A feedforward neural network is utilized to discover intricate, non-linear relationships in the data, which enhances the interpretability of fuzzy logic. Among the crucial elements of the NN implementation are:

1. **Outputs and Inputs**

   The neural network combines the results of the fuzzy logic system with the initial health data, using augmented features as inputs. The goal variable, which indicates whether a person is at risk (1) or not (0), is binary.

2. **Design**

   The network is made up of:

   - A seven-node input layer (six features plus the fuzzy output).
   - A single hidden layer with ReLU activation and ten neurons.
   - A binary classification output layer with a sigmoid activation function.

3. **Procedure for Training**

   Training (70%) and testing (30%) sets make up the dataset. To guarantee that features are on a comparable scale, standardization is used. To maximize performance, the network is trained using the Adam optimizer, which minimizes the binary cross-entropy loss function over a number of rounds.

GUI stands for Graphical User Interface.

To make the prediction system accessible, a graphical user interface (GUI) was created using MATLAB's App Designer. The GUI consists of:

- **Input Fields:** Age, cholesterol, heart rate, stress level, smoking status, and family history are among the variables that users can enter.
- **Predict Results:** When the "Predict" button is clicked, the system uses the hybrid model to assess the inputs and shows the risk level as either "High Risk" or "Low Risk," along with suggestions.
- Users can reset the input fields by pressing the "**Clear**" button.
- **Model Integration:** Real-time forecasts are made possible by the GUI's incorporation of the learned NN and FIS models.

Simulation Outcomes

1. **Outputs of Fuzzy Logic**

   Based on the rule base, the fuzzy logic system by itself offers comprehensible risk assessments. However, because it cannot directly learn from data, its accuracy is just 78%.

2. **Results of Hybrid Models**

   Improved accuracy and robustness are attained by the hybrid model through the integration of the neural network and fuzzy output. Using both the raw features and fuzzy outputs for predictions, the neural network discovers patterns in the augmented data.

3. **Testing with a GUI**

   The GUI is appropriate for usage by patients or healthcare professionals since it effectively processes user inputs, standardizes, assesses the hybrid model, and clearly presents the results.

A strong and understandable model for estimating the risk of heart disease is produced by combining a neural network with a fuzzy logic framework. The model's usefulness is further increased by being integrated into a graphical user interface (GUI), guaranteeing that it may be used successfully in real-world situations.

EVALUATION -

The performance of the fuzzy logic system (FLS), neural network (NN), and their hybrid combination is the main focus of the evaluation of the heart disease prediction model. To guarantee a thorough assessment, a variety of indicators, visualization strategies, and comparative analyses were used.

Performance Metrics

Accuracy, precision, recall, specificity, F1-score, and AUC-ROC were used to assess the models. The following are the outcomes:

Metric	**Fuzzy Logic System**	**Neural Network**	**Hybrid Model**
Accuracy (%)	78.0	85.0	88.0
Precision (%)	80.0	83.0	86.0
Recall (%)	75.0	87.0	90.0
Specificity (%)	80.0	85.0	88.0
F1-Score (%)	77.0	85.0	88.0
AUC-ROC (%)	0.79	0.87	0.91

The Hybrid Model consistently outperformed the standalone methods, combining high interpretability with robust predictive power.

Comparative Analysis

- The **Fuzzy Logic System** excelled in interpretability but achieved lower accuracy due to its rule-based nature and lack of learning capability.
- The **Neural Network** showed higher accuracy by capturing non-linear relationships but lacked interpretability.
- The **Hybrid Model** demonstrated superior performance (88% accuracy) by integrating the strengths of both approaches.

Visualization

- **ROC Curve**: The Hybrid Model achieved the highest AUC-ROC score (0.91), indicating excellent discrimination between high-risk and low-risk cases.
- **Confusion Matrix**: Visualized classifications (TP, TN, FP, FN) to analyze misclassifications and their patterns.

<u>DISCUSSION</u> -

With an accuracy of 88%, the hybrid model outperformed the fuzzy logic system (78%) and the neural network (85%), demonstrating remarkable performance. This enhancement highlights the benefits of fusing neural networks' adaptability with fuzzy logic's interpretability. Notably, the fuzzy logic rules made the system more understandable for medical practitioners by offering insights into the ways that risk variables like stress and cholesterol lead to heart disease.

Nevertheless, the model encountered difficulties, including the computational expense of neural network training and the intricacy of fine-tuning fuzzy rules. In order to expedite the development process, future research should concentrate on automating rule generation through methods such as genetic algorithms. Furthermore, adding more diverse populations to the dataset may increase the model's generalizability and guarantee prediction equity.

The technology provides useful advantages by incorporating the model into an intuitive graphical user interface (GUI), which makes it possible for non-technical users to conduct risk assessments rapidly. This is in line with the overarching objective of improving preventive care and lowering the prevalence of cardiovascular illnesses worldwide.

<u>CONCLUSION</u> -

This experiment effectively illustrated how computational intelligence methods can be used to predict heart disease risk. The system's hybrid model, which combined fuzzy logic and neural networks, struck a compromise between interpretability and prediction accuracy. With an accuracy of 88%, precision of 86%, recall of 90%, and an AUC-ROC of 0.91, the Hybrid Model beat stand-alone techniques, proving its dependability and resilience in assessing the risk of heart disease.

Because of its rule-based methodology, the Fuzzy Logic System offered interpretable insights, which made it appropriate for comprehending the connections between characteristics like age, stress levels, and cholesterol. However, its prediction effectiveness was constrained by its incapacity to learn directly from data. However, the Neural Network lacked the transparency needed for delicate applications like healthcare, despite its superiority in learning intricate, non-linear relationships. In order to overcome these drawbacks, the Hybrid Model combined the advantages of the two approaches, providing flexibility and openness.

The model's usefulness was increased by the intuitive graphical user interface (GUI), which let patients or healthcare professionals enter health factors and get real-time forecasts. The system's preparedness for implementation in actual healthcare environments is guaranteed by its accessibility.

Notwithstanding its achievements, the project encountered difficulties, such as the requirement for improved edge case management, more varied datasets to promote generalizability, and refined fuzzy rules. To confirm its practical impact, future research should investigate more hybrid techniques, automated rule optimization with genetic algorithms, and clinical trial implementation.

To sum up, this initiative demonstrates the revolutionary potential of artificial intelligence in healthcare by offering a precise, understandable, and easily accessible method for predicting the risk of heart disease. By highlighting the value of early identification and preventative care in lowering the worldwide burden of cardiovascular illnesses, it opens the door for future developments in predictive analytics.

<u>REFERENCES</u> -

1. Banerjee, S. (2021). *Heart Attack Risk Prediction Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets/iamsouravbanerjee/heart-attack-prediction-dataset
2. Aizatul Shafiqah Mohd Faizal, T. M., Khor, S. M., & Chang, S. W. (2021). A review of risk prediction models in cardiovascular disease: Conventional approach vs. artificial intelligent approach. Computer Methods and Programs in Biomedicine, 207, 106190. https://doi.org/10.1016/j.cmpb.2021.106190
3. 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. https://doi.org/10.1016/j.health.2022.100016