Analytical and Numerical Solutions for Nonlinear Equations
©Available online at https://ansne.du.ac.ir/
Online ISSN: 3060–785X
2026, Volume 11, Issue 2, pp. xx–xx
Research article

Fuzzy Regression-Based Statistical Modelling of Cardiovascular Risk Factors Using Indian Healthcare Data

Maheswari D[1], Gopinath M[1,*], Alimohammad Fallah Andevari[2], Mostafa NouriJouybari[3], Farshid Mofidnakhaei[4]

[1] Department of Mathematics, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India. Email: maheswarid@skasc.ac.in; gopimalaisamy@gmail.com

[2] Department of Mathematics Education, Farhangian University, P.O. Box 14665-889, Tehran, Iran. Email: alimohammad.fallah@cfu.ac.ir

[3] Payame Noor University (PNU), Tehran, Iran. Email: m_njoybari@pnu.ac.ir

[4] Department of Physics, Sar.C., Islamic Azad University, Sari, Iran. Email: Farshid.Mofidnakhaei@gmail.com

* **Corresponding author(s):** gopimalaisamy@gmail.com

Received: 22/05/2026 **Accepted:** 18/06/2026 **Published:** xx/xx/2026 10.22128/ansne.2026.3328.1218

Abstract

Cardiac disease is a major public health concern in India, and accurate statistical techniques are required for early risk assessment. Traditional regression analysis generally depends on fixed clinical thresholds; however, cardiovascular risk factors such as blood pressure, cholesterol, diabetes, obesity, tobacco use, and physical activity involve uncertainty because they gradually move from the absence of risk to the presence of risk. This study applies fuzzy regression to develop a statistical modelling framework for analysing cardiovascular disease risk factors using the Heart Attack Risk and Prediction dataset from Kaggle. The proposed modelling process includes data preprocessing, descriptive analysis, fuzzy membership construction, fuzzy regression modelling, and comparison with conventional regression models. The selected clinical and lifestyle variables are transformed into linguistic risk levels such as low, moderate, and high risk, thereby allowing the uncertainty associated with individual risk profiles to be represented more meaningfully. The performance of the models is evaluated using statistical measures such as mean absolute error, root mean square error, accuracy, sensitivity, specificity, and receiver operating characteristic area under the curve. The fuzzy statistical modelling approach improves interpretability and provides a flexible risk assessment structure compared with conventional modelling methods. Overall, this study presents a fuzzy regression-based statistical framework to support data-driven cardiovascular risk assessment in the Indian healthcare context.

Keywords: Fuzzy regression, cardiovascular risk factors, Indian healthcare data, heart attack prediction, statistical modelling, fuzzy membership function.

Mathematics Subject Classification (2020): 62J05, 62P10, 03E72

1 Introduction

Cardiovascular disease has emerged as one of the major public health challenges in India, affecting patients, families, and healthcare systems. The increasing prevalence of hypertension, diabetes, obesity, smoking, sedentary lifestyle, poor diet, and stress has contributed to a higher risk of cardiovascular complications. Early identification of such risk factors is essential for preventive healthcare planning and effective disease management. Statistical modelling provides a useful framework for examining the relationship between cardiovascular risk factors and disease outcomes. Conventional regression models are frequently used to analyse clinical data and predict significant health outcomes. However, these models are often based on fixed cut-off points and crisp classification systems.

In real medical situations, clinical indicators such as blood pressure, cholesterol level, body mass index, diabetes status, and physical activity do not always belong strictly to a single risk category. A patient may partially belong to low, moderate, or high-risk categories depending on the combined effect of clinical and lifestyle factors. Therefore, the use of rigid classification boundaries may not represent the true nature of cardiovascular risk. This limitation creates the need for modelling approaches that can handle uncertainty, imprecision, and gradual transitions in health-related variables.

Fuzzy regression provides an effective statistical approach for modelling such uncertainty. Unlike conventional regression, fuzzy regression can represent risk factors through linguistic categories such as low risk, moderate risk, and high risk. This feature makes fuzzy regression suitable for cardiovascular risk modelling because clinical boundaries frequently overlap, and patient risk cannot always be classified precisely. By incorporating fuzzy membership functions, the patient risk profile can be represented more realistically than with strictly crisp statistical models.

In the Indian healthcare context, cardiovascular risk assessment methods should be statistically valid as well as easily interpretable. Publicly available healthcare datasets provide opportunities to study cardiovascular risk patterns using advanced statistical techniques. The use of fuzzy regression in preventive healthcare decision-making can improve the interpretation of uncertain clinical information and support flexible risk assessment.

Therefore, this study proposes a fuzzy regression framework for the statistical modelling of cardiovascular risk factors using Indian healthcare data. The methodology includes data preprocessing, identification of important clinical and lifestyle risk factors, construction of fuzzy membership functions, fuzzy risk score estimation, fuzzy regression modelling, and comparison with conventional regression analysis. The main aim is to develop an interpretable and uncertainty-aware statistical framework for cardiovascular risk assessment.

2 Literature Review

Recent developments in cardiovascular disease diagnosis show a gradual shift from conventional statistical techniques to intelligent computing approaches that are capable of handling nonlinear, uncertain, and complex physiological data. Neural networks, M5Tree models, and adaptive neuro-fuzzy inference systems have been widely applied for predictive classification because of their ability to model nonlinear relationships among variables associated with cardiovascular disease [1]. However, many knowledge-based diagnostic systems still depend heavily on expert-defined rules. These rules are often difficult to construct, time-consuming to validate, and may not always be applied consistently by healthcare professionals [2, 3].

To overcome these limitations, researchers have focused on integrating data-driven rule extraction with statistical regression techniques to improve the objectivity and reliability of cardiovascular risk assessment [4]. In this context, fuzzy-based models are particularly useful because most clinical indicators do not have strict boundaries. For instance, risk factors such as low-density lipoprotein, systolic blood pressure, cholesterol level, and body mass index may gradually move from normal to abnormal categories rather than changing abruptly. Recent studies have shown that fractional-order fuzzy membership functions can effectively capture the uncertainty associated with clinical risk factors such as elevated LDL level and systolic blood pressure [5].

Fuzzy rule-based systems have also improved predictive ability by accommodating ambiguous clinical patterns and enhancing diagnostic sensitivity and specificity [6,7]. These models are useful because they can include patient-specific characteristics such as smoking history, family history, diabetes status, and obesity status in the risk classification process [8]. As a result, fuzzy systems provide a more flexible representation of cardiovascular risk than purely crisp classification models.

Another important advantage of fuzzy diagnostic systems is interpretability. By using linguistic terms and IF–THEN rules, fuzzy models provide meaningful explanations of how individual variables contribute to cardiovascular risk classification [9]. This feature helps bridge

the gap between computational model outputs and clinical judgement, allowing physicians and healthcare professionals to interpret risk predictions more easily [10]. The integration of deep learning with fuzzy inference has further strengthened this field by combining feature extraction ability with interpretable fuzzy reasoning [11]. Such hybrid models are capable of analysing high-dimensional patient data while maintaining a decision-making structure that can handle uncertain and overlapping clinical threshold levels [12].

The combination of fuzzy and neural methods also supports real-time diagnosis and monitoring by transforming raw and multi-source clinical data into meaningful and interpretable outputs [13]. Nevertheless, an important challenge in fuzzy model development is the construction of appropriate membership functions. These functions often require domain expertise and may not remain stable across different patient populations or clinical settings [14]. To address this issue, recent studies have proposed dynamic and adaptive frameworks, including transfer-learning-based models, which allow changes in patient health status to be incorporated over time [15].

Fuzzy set theory continues to be effective in medical applications because it provides a structured way to handle incomplete, uncertain, and imprecise information instead of relying only on binary classifications [16]. Fuzzy logic can transform continuous clinical measurements into useful linguistic categories, especially when fixed threshold values are difficult to define [17]. Therefore, fuzzy-based methods are highly suitable for analysing longitudinal patient data and nonlinear relationships associated with the progression of cardiovascular disease [18].

Moreover, hierarchical information extraction in hybrid diagnostic models improves predictive performance by reducing noise in electronic health records and enhancing the quality of fuzzy rules used for long-term risk estimation [19]. These developments also increase transparency and clinical usability because they provide clearer diagnostic pathways when compared with many black-box machine learning methods [20]. Recent advancements in federated learning and Edge-AI have further supported the deployment of fuzzy-integrated models in real-time clinical environments while preserving patient privacy [21].

Expert knowledge-based constraints can also ensure that model decision boundaries remain consistent with established medical guidelines, thereby reducing the possibility of clinically unrealistic predictions [22]. Multi-layered transparent modelling frameworks may improve the trust of healthcare providers by offering interpretable explanations for automated cardiovascular risk assessments [23, 24]. In addition, integration with electronic health records and clinical decision-support systems can support personalized patient care and improve diagnostic accuracy across different healthcare environments [25].

Although existing studies have explored machine learning, fuzzy logic, and hybrid models for cardiovascular disease prediction, many of them mainly focus on improving predictive accuracy. Comparatively fewer studies emphasize a simple, interpretable, and statistically grounded fuzzy regression framework for modelling cardiovascular risk factors in the Indian healthcare context. Hence, the present study applies fuzzy regression to represent clinical and lifestyle variables through gradual risk membership levels and compares its performance with conventional logistic regression.

3 Research Gap and Motivation

Several studies have examined the use of machine learning algorithms, fuzzy logic systems, and hybrid intelligent models for predicting cardiovascular disease risk. However, most of these studies mainly focus on improving predictive accuracy rather than developing a statistically interpretable model that explains how each clinical and lifestyle variable contributes to cardiovascular risk. In healthcare data analysis, interpretability is highly important because medical practitioners require a clear understanding of how variables such as blood pressure, cholesterol level, diabetes, obesity, smoking habits, physical activity, and stress influence an individual's cardiovascular risk.

Conventional regression models are useful for evaluating relationships between health-related variables, but they generally depend on crisp numerical values and fixed classification boundaries. In real clinical situations, cardiovascular risk factors do not always follow rigid cut-off limits. A patient may lie between two risk categories, such as low and moderate risk or moderate and high risk. This creates uncertainty in classification, which cannot be represented effectively by traditional regression models.

Another limitation of conventional regression is that it produces a single crisp prediction value. Such a prediction may not fully explain the gradual nature of cardiovascular risk. For example, a patient with borderline systolic blood pressure or cholesterol level may not be completely classified as high risk or low risk. Instead, the patient may partially belong to more than one risk category. This type of partial classification is better represented through fuzzy membership functions.

To address this research gap, the present study applies fuzzy regression for modelling cardiovascular risk factors using Indian healthcare data. Fuzzy regression provides a statistical framework that combines regression modelling with fuzzy membership concepts. It allows

clinical and lifestyle variables to be represented through gradual risk levels such as low, moderate, and high. Therefore, the proposed approach provides a more flexible, interpretable, and uncertainty-aware method for cardiovascular risk assessment.

The motivation of this study is to develop a fuzzy regression-based statistical model that can support cardiovascular risk prediction while preserving interpretability. By comparing the fuzzy regression model with conventional logistic regression, this study attempts to identify whether the fuzzy approach provides better clinical usefulness, particularly in terms of sensitivity and risk identification. Thus, the study contributes to the development of an interpretable statistical framework for cardiovascular risk assessment in the Indian healthcare context.

4 Objectives of the Study

The primary objective of this study is to develop a fuzzy regression-based statistical model for assessing cardiovascular risk factors using Indian healthcare data. The study focuses on modelling the uncertainty present in clinical and lifestyle variables by applying fuzzy membership functions to selected cardiovascular risk factors.

The specific objectives of the study are as follows:

1. To identify and analyse the major demographic, clinical, and lifestyle risk factors associated with cardiovascular disease in the Indian healthcare dataset.
2. To perform descriptive and inferential statistical analysis for understanding the distribution and association of selected cardiovascular risk variables.
3. To construct fuzzy membership functions for selected cardiovascular risk factors such as age, cholesterol level, LDL level, systolic blood pressure, diastolic blood pressure, and stress level.
4. To classify selected clinical variables into linguistic risk categories such as low risk, moderate risk, and high risk using fuzzy membership values.
5. To develop a fuzzy risk score that combines fuzzy-transformed clinical variables with relevant binary risk factors such as diabetes, hypertension, obesity, smoking, and family history.
6. To formulate a fuzzy regression model for examining the relationship between cardiovascular risk factors and heart attack risk.
7. To compare the performance of the proposed fuzzy regression model with the conventional logistic regression model using measures such as AIC, accuracy, sensitivity, specificity, balanced accuracy, and ROC-AUC.
8. To propose an interpretable and uncertainty-aware statistical framework for cardiovascular risk assessment in the Indian healthcare context.

Overall, the study aims to provide a flexible and clinically meaningful statistical modelling approach that can support healthcare professionals in identifying cardiovascular risk under uncertainty.

5 Materials and Methods

This study adopts a quantitative statistical modelling approach to analyse cardiovascular risk factors using Indian healthcare data. Both conventional statistical techniques and fuzzy regression modelling are applied to examine the relationship between selected clinical and lifestyle variables and heart attack risk. Since many cardiovascular risk factors do not have clearly defined boundaries, fuzzy regression is used to represent gradual transitions in risk levels and to improve the interpretability of the model.

5.1 Data Source

The dataset used in this study was obtained from the Heart Attack Risk and Prediction Dataset in India, available from Kaggle [26]. The dataset contains patient-level information collected from Indian healthcare sources and includes demographic, clinical, behavioural, lifestyle, environmental, and healthcare-related variables. These variables are useful for assessing cardiovascular risk and identifying patterns associated with heart attack risk.

5.2 Study Variables

The dependent variable considered in this study is heart attack risk, denoted by Y. The independent variables include selected cardiovascular risk factors such as age, gender, cholesterol level, LDL level, HDL level, systolic blood pressure, diastolic blood pressure, diabetes, hypertension, obesity, smoking, alcohol consumption, stress level, physical activity, family history, previous heart attack history, and other health-related factors available in the dataset.

The general relationship between the dependent variable and the selected explanatory variables is expressed as

$$Y = f(X_1, X_2, X_3, \ldots, X_k),$$

where Y represents the cardiovascular risk outcome and $X_1, X_2, X_3, \ldots, X_k$ denote the selected cardiovascular risk factors. This formulation helps to examine how different demographic, clinical, and lifestyle variables contribute to cardiovascular risk.

5.3 Data Preprocessing

Before applying statistical and fuzzy regression techniques, the dataset was preprocessed to ensure accuracy and consistency. Missing values were examined and treated appropriately based on the type of variable. Categorical variables were converted into numerical form using suitable encoding methods. Continuous variables were checked for outliers, scale differences, and distributional characteristics.

For a continuous variable X, min–max normalization may be applied as follows:

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}},$$

where X^* is the normalized value, $X_{\min}$ is the minimum value, and $X_{\max}$ is the maximum value of the variable. This transformation brings the values into a common scale and supports the construction of fuzzy membership functions.

5.4 Descriptive Statistical Analysis

Descriptive statistical analysis was performed to understand the basic characteristics of the selected cardiovascular risk variables. For continuous variables, measures such as mean, standard deviation, minimum, median, maximum, skewness, and kurtosis were considered. For categorical variables, frequency and percentage distributions were computed.

The mean of a variable is calculated as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

and the standard deviation is given by

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}},$$

where X_i represents the observed value and n denotes the number of observations. These measures provide an initial understanding of the distribution and variability of cardiovascular risk factors.

5.5 Correlation Analysis

Correlation analysis was used to examine the strength and direction of the linear relationship between selected cardiovascular risk factors and heart attack risk. Pearson's correlation coefficient is given by

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}},$$

where r denotes the degree of linear association between two variables. A positive value of r indicates a direct relationship, whereas a negative value indicates an inverse relationship. This analysis helps to identify variables that are more closely associated with cardiovascular risk.

5.6 Conventional Regression Model

To establish a baseline model, conventional logistic regression was first applied. Since the dependent variable is binary, representing the presence or absence of heart attack risk, the logistic regression model is expressed as

$$P(Y=1)=\frac{1}{1+e^{-Z}},$$

where

$$Z=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k.$$

Here, $P(Y=1)$ represents the probability of cardiovascular risk, β_0 is the intercept, $\beta_1,\beta_2,\ldots,\beta_k$ are regression coefficients, and $X_1,X_2,\ldots,X_k$ are the selected risk factors.

The model can also be written in logit form as

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k.$$

This model provides a conventional statistical basis for evaluating the relationship between risk factors and cardiovascular outcomes.

5.7 Construction of Fuzzy Membership Functions

To handle uncertainty in medical risk factors, selected continuous variables were transformed into fuzzy linguistic categories. Clinical variables such as age, cholesterol level, LDL level, systolic blood pressure, diastolic blood pressure, and stress level may not have rigid boundaries in real healthcare interpretation. Therefore, these variables were represented using fuzzy sets such as low risk, moderate risk, and high risk.

A triangular membership function is defined as

$$\mu_A(x)=\begin{cases}0, & x\le a,\\ \dfrac{x-a}{b-a}, & a<x\le b,\\ \dfrac{c-x}{c-b}, & b<x<c,\\ 0, & x\ge c,\end{cases}$$

where a, b, and c are the lower, middle, and upper limits of the fuzzy set, respectively. The value $\mu_A(x)$ lies between 0 and 1 and represents the degree of membership of x in the fuzzy set A.

This fuzzy transformation allows each patient to belong partially to more than one risk category, thereby reflecting the uncertainty present in medical assessment.

5.8 Fuzzy Risk Score Construction

After defining the fuzzy membership functions, a fuzzy risk score was constructed for each patient. Let the fuzzy membership values of selected risk factors be denoted by

$$\mu_1,\mu_2,\mu_3,\ldots,\mu_k.$$

The overall fuzzy cardiovascular risk score is expressed as

$$FRS_i=\sum_{j=1}^{k}w_j\mu_{ij},$$

where FRS_i is the fuzzy risk score for the i-th patient, w_j is the weight assigned to the j-th risk factor, and μ_{ij} is the membership value of the j-th risk factor for the i-th patient. A higher fuzzy risk score indicates a higher level of cardiovascular risk.

5.9 Fuzzy Regression Model

The fuzzy regression model was developed to examine the relationship between fuzzy-transformed cardiovascular risk factors and heart attack risk. The general fuzzy regression model is expressed as

$$\tilde{Y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{i1} + \tilde{\beta}_2 X_{i2} + \cdots + \tilde{\beta}_k X_{ik} + \tilde{\varepsilon}_i,$$

where $\tilde{Y}_i$ represents the fuzzy cardiovascular risk outcome for the i-th patient, $X_{i1}, X_{i2}, \ldots, X_{ik}$ are the selected cardiovascular risk factors, $\tilde{\beta}_0, \tilde{\beta}_1, \ldots, \tilde{\beta}_k$ are fuzzy regression coefficients, and $\tilde{\varepsilon}_i$ is the fuzzy error term.

If the fuzzy coefficients are represented as triangular fuzzy numbers, each coefficient can be written as

$$\tilde{\beta}_j = (\beta_j^L, \beta_j^M, \beta_j^U),$$

where β_j^L, β_j^M, and β_j^U denote the lower, middle, and upper values of the fuzzy coefficient, respectively. Thus, the fuzzy regression output can be interpreted as an interval-based risk prediction rather than a single crisp value.

The predicted fuzzy cardiovascular risk is expressed as

$$\tilde{Y}_i = (Y_i^L, Y_i^M, Y_i^U),$$

where Y_i^L, Y_i^M, and Y_i^U represent the lower, most likely, and upper levels of predicted cardiovascular risk. This structure is useful in medical data analysis because it provides a range of possible risk values instead of a single rigid prediction.

5.10 Model Evaluation

The performance of the proposed fuzzy regression model was evaluated and compared with the conventional logistic regression model. For numerical risk prediction, error-based measures such as Mean Absolute Error and Root Mean Square Error were considered.

The Mean Absolute Error is defined as

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i|,$$

and the Root Mean Square Error is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2},$$

where Y_i is the actual value and $\hat{Y}_i$ is the predicted value.

For binary cardiovascular risk outcomes, classification measures such as accuracy, sensitivity, specificity, balanced accuracy, and ROC-AUC were used. Accuracy is given by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Sensitivity is defined as

$$Sensitivity = \frac{TP}{TP + FN},$$

and specificity is defined as

$$Specificity = \frac{TN}{TN + FP},$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative cases, respectively. These evaluation measures help to determine whether the fuzzy regression approach provides better predictive performance and interpretability for cardiovascular risk modelling.

5.11 Methodological Framework

The overall methodology consists of data collection, data preprocessing, descriptive statistical analysis, correlation analysis, conventional logistic regression modelling, construction of fuzzy membership functions, estimation of fuzzy risk scores, fuzzy regression modelling, and comparative model evaluation. This systematic framework supports the development of an interpretable statistical model capable of handling uncertainty in cardiovascular risk assessment using Indian healthcare data.

6 Results and Discussion

This section presents the statistical and fuzzy regression results obtained from the Indian healthcare dataset. The analysis was carried out using RStudio on a dataset consisting of 10,000 observations and 26 variables related to demographic, clinical, lifestyle, environmental, and healthcare factors. The dependent variable considered in the analysis was heart attack risk.

6.1 Preliminary Data Inspection

The dataset contained 10,000 records and 26 variables. No missing values were observed in the dataset, indicating that the data were suitable for direct statistical modelling. The mean value of the heart attack risk variable was 0.3007, which shows that approximately 30.07% of the individuals in the sample were classified as having heart attack risk. The dataset included important cardiovascular risk factors such as age, diabetes, hypertension, obesity, smoking, cholesterol level, LDL level, HDL level, blood pressure, stress level, family history, access to healthcare, and previous heart attack history.

6.2 Descriptive Statistics

Descriptive statistics were computed for selected numerical cardiovascular risk variables to understand their distribution and variability. The results are presented in Table 1.

Table 1. Summary of fuzzy high-risk membership values for selected cardiovascular risk factors.

Fuzzy Variable	Mean	Q1	Median	Q3	Maximum
Age_High	0.1982	0	0	0.3333	1
Chol_High	0.1961	0	0	0.3667	1
LDL_High	0.1930	0	0	0.3448	1
SBP_High	0.2190	0	0	0.4211	1
DBP_High	0.1949	0	0	0.3571	1
Stress_High	0.1945	0	0	0.5000	1

The average age of the participants was 49.39 years, indicating that the sample consisted of both middle-aged and older individuals. The average cholesterol level was 224.80, LDL level was 123.90, systolic blood pressure was 134.70, and diastolic blood pressure was 89.31. These clinical variables showed sufficient variation for modelling cardiovascular risk. The mean stress level was 5.519, suggesting moderate variability in psychological risk exposure among the individuals.

6.3 Frequency Distribution of Risk Factors

The gender distribution showed that 5516 individuals were male and 4484 were female. Among the 10,000 individuals, 3007 were classified as having heart attack risk, representing 30.07% of the sample, while 6993 individuals were classified as not having heart attack risk, representing 69.93%. With respect to clinical and lifestyle risk factors, 929 individuals had diabetes, 2469 had hypertension, 3037 were obese, 3014 had smoking habits, and 3528 reported alcohol consumption. Further, 4036 individuals reported exposure to air pollution, 3113 had a family history of heart attack, 1525 had previous heart attack history, and 3447 had health insurance coverage. These findings confirm the presence of multiple clinical and lifestyle-related cardiovascular risk factors in the dataset.

6.4 Correlation Analysis

Correlation analysis was performed to examine the linear association between selected predictors and heart attack risk. The results indicated that the selected predictors had weak linear association with heart attack risk. Among the variables considered, LDL level showed the strongest positive correlation with heart attack risk, with a correlation value of 0.0212. This was followed by age with a correlation value of 0.0155 and emergency response time with a correlation value of 0.0151. The remaining variables showed either very weak positive or weak negative correlations.

These results indicate that simple linear association alone is not sufficient to explain heart attack risk in the dataset. Therefore, a modelling approach that can handle uncertainty, gradual transitions, and combined risk effects is more appropriate. This supports the use of fuzzy regression for representing cardiovascular risk factors more effectively than traditional linear association-based interpretation.

6.5 Conventional Logistic Regression

A conventional logistic regression model was first fitted as the baseline model. The model used heart attack risk as the dependent variable and selected demographic, clinical, and lifestyle variables as independent variables. The logistic regression model provided a conventional statistical framework for estimating the probability of heart attack risk. However, the results showed limited predictive strength, indicating that the selected variables did not strongly separate individuals with and without heart attack risk under the conventional crisp regression framework.

6.6 Performance of Logistic Regression

The predicted probabilities obtained from the conventional logistic regression model ranged from 0.2159 to 0.3779. Since all predicted probabilities were below the standard cut-off value of 0.5, the default classification classified all individuals as non-risk cases. This resulted in a sensitivity value of 0.0000 and an accuracy value of 0.6993. Although the accuracy appeared high under the default threshold, the model failed to identify risk cases.

To improve classification, an optimal threshold was identified using the Youden index. The optimal threshold for the logistic regression model was 0.3038791. At this threshold, the model produced an accuracy of 0.5455, sensitivity of 0.4895, specificity of 0.5696, balanced accuracy of 0.5295, and AUC of 0.5311. These values indicate weak discriminatory ability of the conventional logistic regression model.

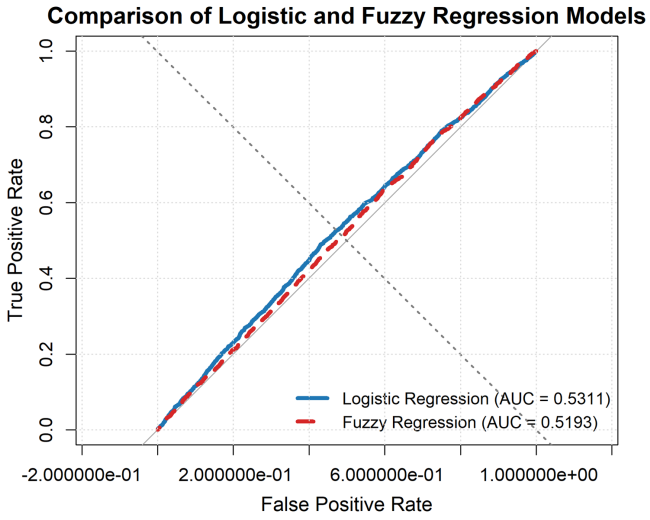

Figure 1. ROC curve comparison of conventional logistic regression and fuzzy regression models for cardiovascular risk prediction.

6.7 Fuzzy Membership Construction

Fuzzy triangular membership functions were constructed for selected continuous cardiovascular risk variables, namely age, cholesterol level, LDL level, systolic blood pressure, diastolic blood pressure, and stress level. These variables were transformed into fuzzy risk categories such as low risk, moderate risk, and high risk. The use of fuzzy membership functions allowed partial membership in more than one risk category, thereby avoiding strict crisp classification.

The mean high-risk membership values ranged from 0.1930 to 0.2190. Among the selected variables, systolic blood pressure had the highest mean high-risk membership value of 0.2190. This indicates that systolic blood pressure showed a relatively stronger fuzzy high-risk representation in the dataset.

Table 2. Summary of fuzzy high-risk membership values for selected cardiovascular risk factors.

Fuzzy Variable	Mean	Q1	Median	Q3	Maximum
Age_High	0.1982	0	0	0.3333	1
Chol_High	0.1961	0	0	0.3667	1
LDL_High	0.1930	0	0	0.3448	1
SBP_High	0.2190	0	0	0.4211	1
DBP_High	0.1949	0	0	0.3571	1
Stress_High	0.1945	0	0	0.5000	1

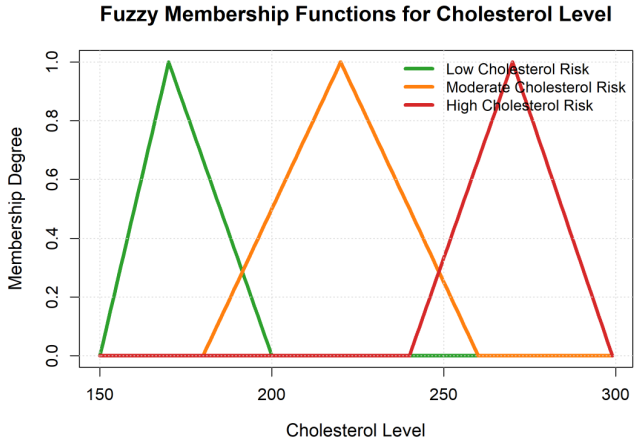

Figure 2. Fuzzy membership functions for cholesterol-based cardiovascular risk classification.

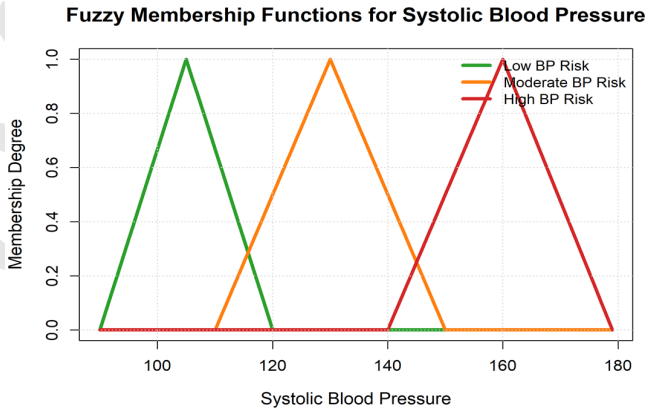

Figure 3. Fuzzy membership functions for systolic blood pressure-based cardiovascular risk classification.

6.8 Fuzzy Risk Score

The fuzzy risk score was constructed using two approaches. The first approach considered only the fuzzy membership values of continuous clinical variables and produced an average fuzzy risk score of 0.1993. The second approach used an adjusted fuzzy risk score by combining fuzzy membership values with binary risk factors such as diabetes, hypertension, obesity, smoking, and family history. The adjusted fuzzy risk score produced a higher average value of 0.2229.

The increase in the adjusted fuzzy risk score indicates that combining clinical fuzzy membership variables with binary risk factors

provides a broader and more realistic representation of cardiovascular risk. This result confirms that cardiovascular risk is influenced not only by continuous clinical indicators but also by lifestyle and medical-history-related factors.

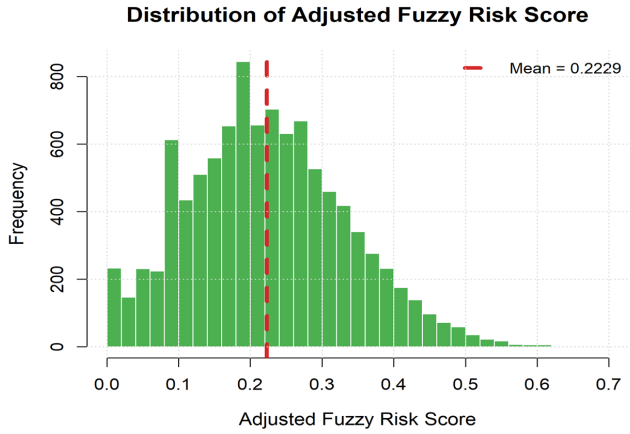

Figure 4. Distribution of adjusted fuzzy risk score among individuals in the Indian healthcare dataset.

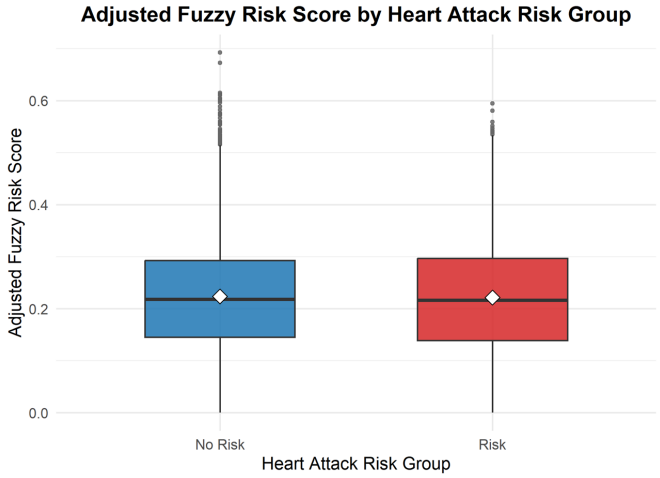

Figure 5. Distribution of adjusted fuzzy risk score according to heart attack risk status.

6.9 Fuzzy Regression Analysis

The fuzzy regression model included fuzzy-transformed variables such as Age_High, Chol_High, LDL_High, SBP_High, DBP_High, and Stress_High, along with binary risk factors such as diabetes, hypertension, obesity, smoking, and family history. The regression estimates provided information on the contribution of each explanatory variable to heart attack risk.

Among the fuzzy-transformed variables, LDL_High showed a positive estimate of 0.126379 with a p-value of 0.0775, suggesting a marginal contribution to cardiovascular risk. DBP_High showed a negative estimate of -0.137098 with a p-value of 0.0599, indicating marginal significance. The remaining variables were not statistically significant at the 5% level. The fuzzy regression model produced a null deviance of 12229, residual deviance of 12217, and AIC value of 12241.28. Since the AIC value of the fuzzy regression model was lower than that of the conventional logistic regression model, the fuzzy model showed a slight advantage in model fit.

6.10 Performance of Fuzzy Regression

The predicted probabilities from the fuzzy regression model ranged from 0.2390 to 0.3534. Similar to the conventional logistic regression model, all probabilities were below the standard cut-off value of 0.5. Therefore, the default threshold classified all individuals as low-risk cases. To improve classification, the optimal threshold was identified using the Youden index. The optimal threshold for the fuzzy regression model was 0.2909567.

At this threshold, the fuzzy regression model produced an accuracy of 0.4194, sensitivity of 0.7672, specificity of 0.2698, balanced accuracy of 0.5185, and AUC of 0.5193. The higher sensitivity indicates that the fuzzy regression model identified a greater number of individuals with heart attack risk. Although the model had lower specificity, its ability to detect risk cases is important in medical screening contexts where identifying high-risk individuals is often more critical than excluding non-risk cases.

6.11 Model Comparison

The performance of the conventional logistic regression and fuzzy regression models is compared in Table 3.

Table 3. Comparative performance of conventional logistic regression and fuzzy regression models.

Model	AIC	AUC	Optimal Threshold	Accuracy	Sensitivity	Specificity	Balanced Accuracy
Conventional Logistic Regression	12253.76	0.5311	0.3039	0.5455	0.4895	0.5696	0.5295
Fuzzy Regression	12241.28	0.5193	0.2910	0.4194	0.7672	0.2698	0.5185

The fuzzy regression model produced a lower AIC value, indicating a better model fit compared with the conventional logistic regression model. However, the logistic regression model showed slightly higher AUC, accuracy, specificity, and balanced accuracy. The most notable advantage of the fuzzy regression model was its sensitivity, which increased from 0.4895 in the conventional logistic regression model to 0.7672 in the fuzzy regression model. This improvement is important in medical risk screening because the early identification of high-risk individuals is more important than classifying all non-risk individuals correctly.

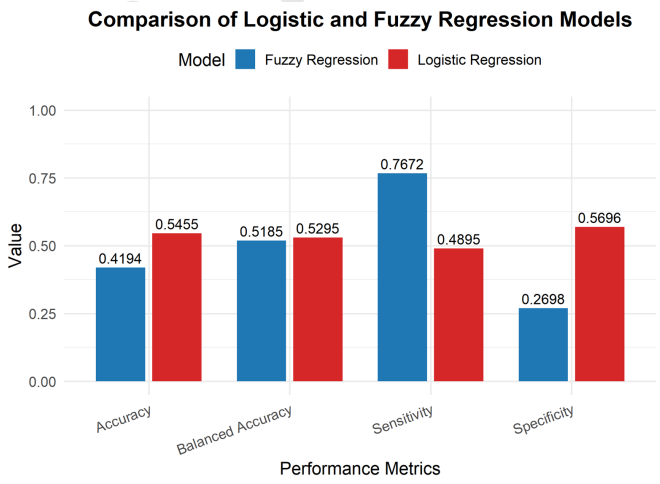

Figure 6. Performance comparison of conventional logistic regression and fuzzy regression models using selected classification measures.

6.12 Discussion

The results indicate that the selected cardiovascular variables showed weak linear association with heart attack risk. In the conventional logistic regression model, the predicted probabilities were concentrated within a narrow range and failed to classify high-risk individuals under the default threshold. Although the optimal threshold improved the classification performance, the model still showed weak discriminatory power.

The fuzzy regression approach provided an interpretable modelling structure by converting selected clinical variables into fuzzy high-risk membership values. This transformation allowed gradual representation of cardiovascular risk instead of forcing each individual into a strict risk category. Among the fuzzy variables, systolic blood pressure showed the highest mean high-risk membership value, indicating its strong representation as a fuzzy risk factor in the dataset.

The adjusted fuzzy risk score provided a more comprehensive representation of individual risk by integrating fuzzy clinical variables with binary clinical and lifestyle factors. Although the fuzzy regression model produced slightly lower AUC and accuracy than the conventional logistic regression model, it achieved a lower AIC and substantially higher sensitivity. This suggests that fuzzy regression may be more useful as a preliminary screening approach, especially when the primary objective is to identify individuals who may be at risk of heart attack.

Overall, the findings show that fuzzy regression provides an uncertainty-aware and interpretable statistical modelling framework for cardiovascular risk assessment. While the predictive separation of both models was limited, the fuzzy regression model offered better risk identification through improved sensitivity and meaningful fuzzy risk representation.

7 Conclusion and Future Work

This study developed a fuzzy regression-based statistical modelling framework for assessing cardiovascular risk factors using Indian healthcare data. The analysis was carried out on a dataset containing 10,000 observations and 26 variables related to demographic, clinical, lifestyle, environmental, and healthcare characteristics. The conventional logistic regression model was first applied as a baseline model, and fuzzy regression was then used to represent uncertainty in selected cardiovascular risk factors through fuzzy membership functions.

The descriptive analysis showed sufficient variation in key clinical variables such as age, cholesterol level, LDL level, systolic blood pressure, diastolic blood pressure, and stress level. The correlation analysis indicated that the selected predictors had only weak linear association with heart attack risk. This finding supports the need for an uncertainty-based modelling approach rather than depending only on conventional linear association and crisp classification methods.

The fuzzy membership construction transformed selected clinical variables into gradual high-risk membership values. Among these variables, systolic blood pressure showed the highest mean high-risk membership value, indicating its relatively stronger fuzzy risk representation in the dataset. The adjusted fuzzy risk score further improved the representation of cardiovascular risk by combining fuzzy clinical variables with binary risk factors such as diabetes, hypertension, obesity, smoking, and family history.

The comparative model evaluation showed that the fuzzy regression model produced a lower AIC value than the conventional logistic regression model, indicating a slight improvement in model fit. Although conventional logistic regression showed slightly higher accuracy, specificity, AUC, and balanced accuracy, the fuzzy regression model achieved much higher sensitivity. This is important in healthcare screening because the early identification of high-risk individuals is more valuable than simply classifying non-risk cases correctly.

Overall, the findings suggest that fuzzy regression provides an interpretable and uncertainty-aware statistical framework for cardiovascular risk assessment. The proposed approach is useful for representing gradual transitions in clinical risk factors and can support healthcare professionals in making flexible and meaningful risk-based decisions.

For future work, the proposed fuzzy regression framework may be extended by incorporating larger real-time clinical datasets and additional medical variables. Further studies may also compare fuzzy regression with advanced machine learning and hybrid fuzzy models to improve both predictive accuracy and interpretability in cardiovascular risk prediction.

Authors' Contributions

All authors contributed to the conceptualization, methodology, analysis, interpretation of results, manuscript preparation, and final approval of the article.

Data Availability

The dataset used in this study was obtained from the publicly available Heart Attack Risk and Prediction Dataset in India from Kaggle. The data were used only for academic and statistical modelling purposes.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

This study used a publicly available secondary dataset and did not involve direct human participation. Therefore, ethical approval was not required.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Use of Artificial Intelligence (AI) Tools

Artificial intelligence tools were used only for language refinement, grammar correction, and formatting support. The conceptualization, statistical analysis, interpretation of results, and final approval of the manuscript were carried out by the authors.

References

[1] O. Taylan, A. S. Alkabaa, H. S. Alqabbaa, E. Pamukçu and V. Leiva, Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods, Biology, 12(1), 117–117, (2023). DOI: https://doi.org/10.3390/biology12010117.

[2] M. Krishnam and S. K. Shafiulilah, An Enhanced Risk Prediction System for Cardiovascular Disease in India using Fuzzy Classification, International Journal of Computer Applications, 126(7), 11–16, (2015). DOI: https://doi.org/10.5120/ijca2015906087.

[3] O. Ugale, C. Puri and P. M. Gote, A Comprehensive Assessment of Fuzzy Rule-based Systems for Heart Attack Prediction, 1186–1191, (2025). DOI: https://doi.org/10.1109/icmlas64557.2025.10968441.

[4] H. Parveen, S. Wajahat, A. Rizvi and R. S. K. Boddu, A Comprehensive Study of Risk Prediction Techniques for Cardiovascular Disease, 433–440, (2023). DOI: https://doi.org/10.1201/9781003393580-66.

[5] R. Kishor and C. Diwakar, Integrating Regression to Fuzzy Based Rule Extraction for CHD Risk Prediction, International Journal of Technological Advancements and Industrial Applications, (2025). DOI: https://doi.org/10.52458/28374061.v3.iss1.ijtaia.a5.2025.

[6] G. Angayarkanni, A Novel Deep Fuzzy Rule-Based System for Early Heart Disease Risk Prediction, Journal of Information Systems Engineering & Management, 10, 181–192, (2025). DOI: https://doi.org/10.52783/jisem.v10i19s.3002.

[7] S. Saratkar, A. Chaudhari, T. Thute, R. Raut, G. Thakre and H. Kumar, Assessment of Heart-Attack Prediction Using Fuzzy Rule Based System, 1–6, (2024). DOI: https://doi.org/10.1109/iccubea61740.2024.10774808.

[8] A. Suzuki and E. Negishi, Fuzzy Logic Systems for Healthcare Applications, Journal of Biomedical and Sustainable Healthcare Applications, 1–9, (2024). DOI: https://doi.org/10.53759/0088/jbsha20240401.

[9] P. Anooj, Clinical Decision Support System: Risk Level Prediction of Heart Disease Using Weighted Fuzzy Rules and Decision Tree Rules, Open Computer Science, 1(4), 482–498, (2011). DOI: https://doi.org/10.2478/s13537-011-0032-y.

[10] G. Casalino, G. Castellano, U. Kaymak and G. Zaza, Balancing Accuracy and Interpretability Through Neuro-Fuzzy Models for Cardiovascular Risk Assessment, 2021 IEEE Symposium Series on Computational Intelligence (SSCI), 1–8, (2021). DOI: https://doi.org/10.1109/ssci50451.2021.9660104.

[11] S. Ramasami and P. U. Maheswari, A Deep Fuzzy Inference System for Interpretable Multi-Class Heart Disease Risk Prediction, International Journal of Imaging Systems and Technology, 36(1), (2025). DOI: https://doi.org/10.1002/ima.70264.

[12] S. Patil and S. Bhosale, Improving Cardiovascular Disease Prognosis Using Outlier Detection and Hyperparameter Optimization of Machine Learning Models, Revue d'Intelligence Artificielle, 37(4), 1169–1180, (2023). DOI: https://doi.org/10.18280/ria.370429.

[13] N. E. Menaceur, S. Kouah, M. Derdour, K. Ouanes and M. Ammi, Fuzzy Logic in Arrhythmia Detection: A Systematic Review of Techniques, Applications, and Clinical Interpretability, Applied Computer Science, 21(3), 162–181, (2025). DOI: https://doi.org/10.35784/acs_7657.

[14] Y. Zheng, Z. Xu, T. Wu and Y. Zhang, A Systematic Survey of Fuzzy Deep Learning for Uncertain Medical Data, Artificial Intelligence Review, 57(9), (2024). DOI: https://doi.org/10.1007/s10462-024-10871-7.

[15] M. A. Tawfeek, I. Alrashdi, M. Alruwaili and H. Allahem, Cardiovascular Disease Detection: A Hybrid Machine Learning-AI Framework for Personalized Diagnosis and Risk Assessment, PLoS ONE, 20(10), (2025). DOI: https://doi.org/10.1371/journal.pone.0335421.

[16] S. S. Kumar, A. Harsh, R. Chandra and S. Agarwal, Fuzzy Rule based Intelligent Cardiovascular Disease Prediction using Complex Event Processing, arXiv, (2024). DOI: https://doi.org/10.48550/arxiv.2409.15372.

[17] Y. Zhang et al., Predicting Need for Heart Failure Advanced Therapies Using an Interpretable Tropical Geometry-Based Fuzzy Neural Network, PLoS ONE, 18(11), (2023). DOI: https://doi.org/10.1371/journal.pone.0295016.

[18] D. Chicco, S. Spolaor and M. S. Nobile, Ten Quick Tips for Fuzzy Logic Modeling of Biomedical Systems, PLoS Computational Biology, 19(12), (2023). DOI: https://doi.org/10.1371/journal.pcbi.1011700.

[19] Y. Zheng, Z. Xu, T. Wu and Y. Zhang, A Systematic Survey of Fuzzy Deep Learning for Uncertain Medical Data, Research Square, (2023). DOI: https://doi.org/10.21203/rs.3.rs-3126621/v1.

[20] R. Saranya, M. Rajagopal, J. Ramprasath, K. ThamaraiSelvi and S. Leelavathy, An Efficient Fuzzy Logic-Integrated Hybrid Deep Learning Framework for Medical Diagnosis, Fuzzy Information and Engineering, 1–18, (2026). DOI: https://doi.org/10.26599/fie.2025.9270072.

[21] V. S. Krushnasamy, N. H. Al-Muraad, S. Peter, B. S. Khalaf, M. B. Alazzam and I. I. Raj, Fuzzy Neural Networks for Real-Time Decision Support in Healthcare, 1–6, (2025). DOI: https://doi.org/10.1109/icicke65317.2025.11136222.

[22] Y. Zhang, Improving Automatic Clinical Decision Support System with Advanced Computational Methods, Deep Blue, University of Michigan, (2025). DOI: https://doi.org/10.7302/25729.

[23] S. N. Silva, Explainable Predictive Analytics for Smart Healthcare Using a Modular Hybrid Intelligence Framework, MATTER International Journal of Science and Technology, 11, 16–27, (2025). DOI: https://doi.org/10.20319/mijst.2025.11.1627.

[24] M. Nemade, An Explainable Fuzzy Deep Learning Framework for Uncertainty-Based Medical Diagnosis, International Journal of Fuzzy Mathematical Archive, 24(2), 47–58, (2025). DOI: https://doi.org/10.22457/ijfma.v24n2a04256.

[25] Md. A. Talukder, A. S. Talaat, M. Kazi and A. Khraisat, XAI-HD: An Explainable Artificial Intelligence Framework for Heart Disease Detection, Artificial Intelligence Review, 58(12), (2025). DOI: https://doi.org/10.1007/s10462-025-11385-6.

[26] Kaggle, Heart Attack Risk and Prediction Dataset in India, Available online: https://www.kaggle.com/datasets/ankushpanday2/heart-attack-risk-and-prediction-dataset-in-india.