

Myocardial Infarction Risk Stratification: Applied Statistical Modeling using Bayesian Additive Regression Trees with Chaotic Dwarf Mongoose Optimization

Ammar Kuti Nasser^{1*}, Ban A. Abdulmajeed²

¹Department of Psychological Counseling and Educational Guidance, College of Basic Education, Mustansiriyah University, Iraq;

²Department of Mathematics, College of Basic Education, Mustansiriyah University, Iraq

ABSTRACT

This study concerns the emerging challenge of myocardial infarction in Iraq and aims to classify its risk using advanced statistical methodologies. Myocardial infarction is a rising health problem in Iraq attributed to variables from age to blood pressure and outdated healthcare system. Conventional risk models fail to well illustrate the complicated and non-linear nature of the disease history; thus we use a two-stage method in this work. The important biomarkers are extracted from the data with Chaotic Dwarf Mongoose Optimization and then they are used for non-linear analysis using Bayesian Additive Regression Trees. Data were gathered, refined and analyzed from 1,319 patients at a cardiac center. The integrated BART-CDMO hybrid model showed better prediction of MI risk than individual classifiers. The hybrid model was found to be the best way to reduce the prediction error, noise and incomplete data processing. Given the robustness of the model in various sample sizes, it could be considered as a decision support instrument for health policy planning, early detection strategies, prevention and resource allocation. The paper suggests more in-depth data analysis and exploring other sociodemographic factors, as well as training healthcare software developers in advanced data science. This method may be suitable for other chronic diseases as well.

ARTICLE HISTORY

Received March 27, 2026

Accepted May 01, 2026

KEYWORDS

Applied Statistics,
Bayesian Additive Regression
Trees,
Myocardial Infarction,
Chaotic Dwarf Mongoose
Optimization,
Risk Prediction,
Feature Selection

1. Introduction

Myocardial infarction is among the most common and complicated acute illnesses worldwide in several countries. It is a cardiac disorder in which blood flow to the heart muscle stops suddenly, or the body cannot make efficient use of the oxygen it needs. This blockage causes myocardial necrosis, which can be extremely damaging to cardiac function and overall health. About 17.9 million people were thought to have died from cardiovascular diseases worldwide in 2023 ([World Health Organization, 2021](#)). But they will need to account for this because is poised to affect more staggering numbers if effective strategies are not initiated. Myocardial infarction prevalence in Iraq is increasing, reflecting a major public health problem ([Aldaggistany et al., 2023](#); [Sharif & Lafi, 2021](#)). This syndrome is mainly a result of the representative biomedical conditions of high blood pressure, uncontrolled diabetes and diverse lipid levels. In addition, health care and influences are not equilibrate from one region to another inside the country: it also deteriorates patient's condition. The complex medical mechanism of myocardial infarction by many nonlinear factors, the integrating prediction without combination of linear and non-linear statistical models is not easily made prediction accurately. To meet the need, in the present study we follow a hybrid approach that combines state-of-the-art statistical methods with optimization procedures to detect both linear and non-linear patterns from medical data ([Chipman et al., 2010](#)). The advantage of the Bayesian Additive Regression Trees model is that it can analyze and interpret complex interactions in clinical data, while the Chaotic Dwarf Mongoose Optimization has a strong capability to select optimal biomarker sets that minimize prediction error and maximize discrimination. Through the integration of the two methods, this BART-CDMO hybrid model is intended to develop a more accurate and effective risk stratification technique with the purpose of correctly classifying MI risk in Iraq for enhancing clinical decision-making and making better preventive decisions ([Aldaggistany et al., 2023](#); [Chipman et al., 2010](#)). This analysis includes data from 1,319 patients with suspected acute coronary syndrome in Erbil, Iraq, which were selected to provide good clinical coverage and data quality. Here, we are focusing on the cardiac biomarkers themselves and their

within-patient associations. By using advanced statistics and AI techniques, this health data study sets out to address gaps in local research whilst modelling accurate prediction of outcomes using high-quality clinical data. Lastly, this work aims to establish a scientific rationale for evidence-based decisions and preventive cardiology practices in order to mitigate public health consequences in Iraq ([Amsterdam et al., 2014](#); [Mozaffarian et al., 2016](#)).

2. Literature Review

Myocardial infarction risk prediction using statistical or advanced artificial intelligence models has been investigated in many local and international studies. These studies demonstrate the importance of adopting hybrid models, in which traditional statistical prediction methods are combined with modern machine learning or optimization techniques to improve performance and handle complex risk structures. Internationally, [Chipman et al. \(2010\)](#) introduced Bayesian Additive Regression Trees (BART) as a flexible non-linear statistical model that can capture complex interactions and non-additive effects in biomedical data, offering better calibration and interpretation than many black-box methods, and this foundational work confirmed that flexible statistical learning methods can outperform classic linear models in medical prediction. Later, [Kiani et al. \(2015\)](#) in Iran carried out a cross-sectional study on assessment of risk factors in patients with myocardial infarction, and they showed that age, sex, hypertension, diabetes, cholesterol, smoking and delay in seeking care are key determinants of MI, many of which are potentially controllable through primary and secondary prevention. [Sparapani et al. \(2016\)](#) extended the application of BART to nonparametric survival analysis in medical data, and they showed that BART can model complex time-to-event outcomes without requiring strict parametric assumptions about the underlying distribution, which makes it especially valuable in clinical settings where patient outcomes depend on many interacting variables. On the local level, [Sharif and Lafi \(2021\)](#) in Sulaimani city analyzed common risk factors of myocardial infarction and some socio-demographic characteristics, and they reported high prevalence of smoking, hypertension, diabetes and obesity among MI patients, with clear differences across age groups and educational levels, which supports the need to build prediction models that are tailored to the Iraqi population and its specific pattern of risk factors. More recently, [Aldaggistany et al. \(2023\)](#) investigated the prevalence and demographic characteristics of metabolic syndrome in Iraqi patients with acute coronary syndrome and found that clustering of metabolic risk factors such as central obesity, high blood pressure, high triglycerides and impaired fasting glucose is very common among ACS patients in Baghdad, suggesting that prediction models for MI in Iraq should consider combined biomedical profiles rather than single risk factors. From a methodological perspective, [Noroozi et al. \(2023\)](#) conducted a comparative analysis of different feature selection techniques applied to heart disease datasets, and they showed that wrapper-based methods consistently outperformed filter-based methods by achieving 12-15% higher classification accuracy, which emphasizes that selecting the right subset of biomarkers is critical for improving model performance and reducing prediction errors. At the global level, the [World Health Organization \(2021\)](#) emphasizes that major behavioral and intermediate risk factors like unhealthy diet, physical inactivity, tobacco use, harmful use of alcohol, raised blood pressure, raised blood lipids and diabetes are central drivers of cardiovascular disease burden, including myocardial infarction. In terms of advanced methodology, [Abdelrazek et al. \(2024\)](#) proposed the Chaotic Dwarf Mongoose Optimization (CDMO) algorithm for feature selection and classification, showing that such metaheuristic optimizers can reduce dimensionality and enhance classification performance in medical datasets. In the article [Nasser \(2025\)](#), he used Bayesian methods for risk modeling. These results are more stable than the frequentist and Bayesian method is better in uncertainty estimation. These results can be used in medical prediction tasks too, because they need reliable probabilistic estimates. In another comparative study, [Nasser \(2025\)](#) confirmed that Bayesian models achieve superior model fit and lower prediction errors according to AIC and BIC criteria compared to traditional regression methods, which reinforces the advantages of Bayesian frameworks for handling uncertainty in high-stakes prediction problems ([Lahmiri et al., 2023](#)). Together, these studies motivate the use of a hybrid BART-CDMO framework in the present research to select the most important cardiac biomarkers and model non-linear relations with myocardial infarction risk in Iraqi patients, and they confirm that adapting prediction models to the demographic and clinical context of Iraq is crucial for accurate risk assessment and better-informed health policy ([Thygesen et al., 2018](#)).

3. Research Problem

Projections for the prevalence of myocardial infarction in Iraq are increasingly challenging and difficult to make, especially when biomedical risk factors and cardiac biomarkers are taken as main core variables. Clinical information from patients presenting with acute coronary syndrome often contains valuable diagnostic clues, yet extracting meaningful patterns from this data remains difficult because cardiac outcomes are typically highly non-linear and complicated in structure. In this paper, we focus on using computer-assisted statistical methods and artificial intelligence to extract useful information for enhancing the diagnosis of patients and predicting MI risk from clinical data, which are usually non-linear

and complicated in structure but very valuable. Also, it is common knowledge that medical records often come with seasonal and temporal variations, non-uniform quality and accuracy, especially in areas where healthcare infrastructure has to struggle or even bear the brunt of instability, as is the case in many parts of Iraq. Based on these problems, the main research problem in this paper is that an integrated prediction model needs to be constructed which can overcome the drawbacks of the traditional statistical methods and single-algorithm approaches. This model should integrate the flexible capabilities of Bayesian Additive Regression Trees (BART) and the feature selection strength of the Chaotic Dwarf Mongoose Optimization (CDMO) algorithm, which is effective at capturing nonlinear and complicated relations among biomedical variables and selecting only the most important cardiac biomarkers. We want to build better models that can be more accurate and flexible. These models need to understand the complex relation between medical variables. This help in health policy decisions and prevention programs. It also important for early screening to find patients with high risk of myocardial infarction. That is, the predominant issue is that our learned models are not rich enough for modelling the complexity in medical data and lead to quality distortion of prediction as well as poor tactical decisions with respect to clinical reality. A new hybrid BART-CDMO model is needed to fill the literature's gap by combining the strengths of contemporary statistical methods and artificial intelligence (AI) methods, which should be designed based on Iraq's healthcare data peculiarities, since the rates for MI patients remain increasing and the disease burden continues to grow.

4. Research Objectives

In this work, we try to build a hybrid prediction model that incorporates the merits of Bayesian Additive Regression Trees (BART) in analyzing complex non-linear patterns and CDMO's strength in dealing with the feature selection and optimization among MI-related biomedical data. The model is intended to be applied for correct and optimal risk prediction of myocardial infarction in Iraq when sufficient medical data are available from cardiac centers. Such forecasts seek to enhance the understanding of disease dynamics and prevention and treatment interventions. Secondary objectives are to compare model performance with each other and to that of traditional statistical models versus full hybrid model outputs using robust accuracy metrics (e.g., sensitivity, specificity, AUC) and establish the effect of biomedical covariate influences on predictive ability. The current study also aspires to provide practical evidence-based policy recommendations for health-care payers on how to improve the management of resources and support the scaling-up of data-driven and evidence-based health policies. Moreover, our work attempts to address the identified limitations in MI prediction studies within Iraq by considering model flexibility towards local medical data characteristics. It also underscores that more advanced artificial intelligence and statistical analysis should be considered to optimize patients' life quality and appropriate therapeutics concerning the Iraqi medical sectors.

5. Methodology

In this paper, a novel analytical solution that integrates sound statistical methods and state-of-the-art artificial intelligence (AI) techniques is proposed to establish a hybrid prediction model tailored to data-borne characteristics for predicting myocardial infarction risk based on biomedical variables. The main stages of the research methodology are briefly outlined and detailed in the subsequent sections.

6. Myocardial Infarction: Diagnostic Criteria and Risk Factors

Myocardial infarction happens when blood flow to the heart muscle gets blocked by a clot in a coronary artery, which causes the heart cells to die and weakens the heart's ability to work properly (Tan & Roy, 2019). Doctors classify it into two types: STEMI when the artery is completely blocked and needs emergency treatment right away, and NSTEMI when the blockage is partial but still dangerous (Rashid & Hassan, 2022; Tan & Roy, 2019). To diagnose a heart attack, doctors look at the patient's symptoms, the ECG results, and the levels of cardiac biomarkers like troponin and CK-MB in the blood, which go up when the heart muscle is damaged (St. Ville et al., 2025; Tan & Roy, 2019). Troponin is most trusted marker now. It more accurate than CK-MB and stay high for longer time. This very useful to confirm MI in emergency 17. Main risk factors include high blood pressure, diabetes, high cholesterol, smoking and being overweight. These factors can be changed. But age, gender and family history cannot be changed (Apple et al., 2017; Sparapani et al., 2016). INTERHEART study was big international study. It shows that few modifiable risk factors explain most heart attacks in the world. So early detection can prevent many cases (Apple et al., 2017). Heart disease is still top cause of death globally. Metabolic problems and unhealthy lifestyles making it worse (Sparapani et al., 2016; Timmis et al., 2020; World Health Organization, 2021). In Iraq, more patients with acute coronary syndrome have multiple risk factors at the same time, and the problem is made harder by unequal access to healthcare and weak prevention programs in many areas (Aldaggistany et al., 2023; Sharif &

Lafi, 2021). This is why building accurate prediction models that use clinical measurements and lab results is very important for identifying high-risk patients early and improving treatment outcomes (Tan & Roy, 2019; Timmis et al., 2020).

7. Data Collection and Processing

In this work, clinical data on myocardial infarction were collected from Zheen Hospital in Erbil, Iraq (Yusuf et al., 2004). The dataset contains 1,319 patient records collected between January and May 2019, with each record including nine clinical and biochemical variables: Age, Gender, Heart Rate, Systolic Blood Pressure, Diastolic Blood Pressure, Blood Sugar, CK-MB, Troponin, and a binary outcome (Result) indicating the presence or absence of myocardial infarction. The data were first cleaned by removing clear outliers, checking inconsistent entries, and correcting simple recording errors. Then the biomedical variables were preprocessed by imputing missing values and applying Min–Max normalization so that all features are on a comparable scale. Finally, the whole dataset was split into 80% for training and 20% for testing, to keep a balance between good model evaluation and avoiding overfitting. Myocardial infarction is one of the leading cardiovascular causes of death worldwide, which makes building accurate prediction models especially important (World Health Organization, 2021).

8. Feature Selection using Chaotic Dwarf Mongoose Optimization (CDMO)

Feature selection is one of the major steps in pattern recognition and classification. It aims to remove the redundant and irrelevant features from the dataset. The CDMO algorithm is a wrapper-based model that selects the best features which give higher classification accuracy. Chaotic maps were used to modify the key elements of the Dwarf Mongoose movement during optimization. This speeds up convergence and increases the effectiveness of DMO. The DMO is a meta-heuristic method inspired by the foraging behavior of dwarf mongooses in nature. Random numbers in the original DMO equation are replaced by chaotic maps to avoid returning to the same sleeping mound and to enhance exploration capability (Abdelrazek et al., 2024).

9. Bayesian Additive Regression Trees (BART)

BART is a Bayesian sum-of-trees model where each tree is constrained by a regularization prior to be a weak learner. Fitting and inference are accomplished via an iterative Bayesian backfitting MCMC algorithm that generates samples from a posterior. BART is a nonparametric Bayesian regression approach which uses dimensionally adaptive random basis elements (Chipman et al., 2010).

The fundamental problem is making inference about an unknown function f that predicts an output Y using a p dimensional vector of inputs x :

$$Y = f(x) + \varepsilon, \varepsilon \sim N(0, \sigma^2) \quad (1)$$

where:

- Y is the response variable (MI outcome: 1 if MI present, 0 if absent)
- $f(x)$ is the unknown regression function to be estimated
- x is the vector of predictors (cardiac biomarkers such as Age, Blood Pressure, Troponin, CK-MB)
- ε is the random error term

The BART model approximates the function $f(x)$ by a sum of m regression trees (Chipman et al., 2010; Hernández et al., 2015):

$$f(x) \approx h(x) \equiv \sum_{j=1}^m g_j(x) \quad (2)$$

where:

- $h(x)$ is the BART approximation to the true function $f(x)$
- m is the total number of trees in the ensemble.
- $g_j(x)$ denotes the j -th regression tree

BART has been successfully applied to distinguish between patients with cardiovascular disease and controls in proteomic experiments, demonstrating its capability to handle complex medical classification tasks. The sum-of-trees model can be more explicitly expressed as (Chipman et al., 2010; Hernández et al., 2015);

$$Y = \sum_{j=1}^m g(x; T_j, M_j) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \quad (3)$$

where:

- T_j is the tree structure of the j-th tree (including split rules and topology)
- $M_j = \mu_{j1}, \mu_{j2}, \dots, \mu_{jb}$ is the set of parameter values at the terminal nodes of tree j
- $g(x; T_j, M_j)$ is the function that assigns a terminal node value to input x based on tree structure T_j and parameters

M_j

- μ_{ji} represents the predicted value at terminal node i of tree j
- the expected value of the outcome given the predictors is (Chipman et al., 2010):

$$E(Y|x) = \sum_{j=1}^m g(x; T_j, M_j) \quad (4)$$

where:

- $E(Y|x)$ is the conditional expectation (expected value) of Y given predictor values x
- This sum represents the combined contributions from all m trees to produce the final MI risk prediction
- Each tree contributes by assigning the patient to a terminal node and returning its associated parameter value

10. Bayesian Inference in BART

In the BART model, Bayesian inference is used to combine prior information about the trees with the evidence from the observed data. For each tree j , the prior distribution is defined on the tree structure T_j and on the terminal node parameters M_j . The joint prior factorizes as (Chipman et al., 2010; Tan & Roy, 2019)

$$P(T_j, M_j) = P(T_j) \cdot P(M_j|T_j) \quad (5)$$

where $P(T_j)$ is the prior probability on the structure of tree j, and $P(M_j|T_j)$ is the conditional prior on the terminal node parameters given the tree structure. The prior specification in BART requires careful selection of hyperparameters for tree structures, terminal node parameters, and residual variance. The terminal node parameters are typically assumed to be normally distributed with mean zero and variance chosen to regularize the contribution of each tree (Goedhart et al., 2023; Tan & Roy, 2019).

Given the prior and the observed training data D, the posterior distribution over all trees and parameters is obtained using Bayes' rule (Chipman et al., 2010):

$$P(T_1, \dots, T_m, M_1, \dots, M_m, \sigma|D) \propto P(D|T_1, \dots, T_m, M_1, \dots, M_m, \sigma) \cdot P(T_1, \dots, T_m, M_1, \dots, M_m, \sigma) \quad (6)$$

where $P(D|T_1, \dots, T_m, M_1, \dots, M_m, \sigma)$ is the likelihood of the data, $P(T_1, \dots, T_m, M_1, \dots, M_m, \sigma)$ is the joint prior, σ is the error standard deviation, and the symbol $\propto$ means proportional to. Posterior inference in BART is performed using a Bayesian backfitting MCMC algorithm that iteratively updates each tree while conditioning on all other trees. This approach allows for efficient sampling from the complex posterior distribution.

The Bayes estimate of the regression function at a new input x^* is given by the posterior mean:

$$\hat{f}(x^*) = E[\sum_{j=1}^m g(x^*; T_j, M_j) | D] \quad (7)$$

where $\hat{f}(x^*)$ is the estimated regression function at x^* , $E[\cdot | D]$ denotes expectation with respect to the posterior distribution given the data D, and the sum over $j = 1, \dots, m$ aggregates the contributions of all trees in the ensemble (Chipman et al., 2010). Recent use of Bayesian methods in risk modeling shows big improvements over traditional methods. BART was used successfully in medical classification problems. This includes cardiovascular disease prediction and medical imaging. It can handle complex medical data with good accuracy (Ozcan & Peker, 2023; St. Ville et al., 2025).

11. CDMO Feature Selection Algorithm

The CDMO algorithm work by copying the social behavior of dwarf mongoose groups. It also uses chaotic mapping to improve exploration and exploitation. The algorithm starts with a population of candidate feature subsets. Then it refine them based on classification performance (Abdelrazek et al., 2024). Each candidate solution is a binary vector. In this vector, 1 means feature is included and 0 means it excluded. The fitness function evaluates each subset using cross-

validated classification accuracy on the MI dataset from Jin Hospital. The chaotic component introduces randomness to avoid local optima while the mongoose optimization guides the search toward better feature combinations.

The CDMO selection process can be formalized as follows. Let the binary feature selection vector be:

$$s = [s_1, s_2, \dots, s_p] \quad s_i \in \{0, 1\} \quad (8)$$

where $S_i = 1$ if feature i is selected, and $S_i = 0$ otherwise.

The fitness function to maximize is:

$$F(s) = \text{Accuracy}(BART(X_s, Y)) \quad (9)$$

where X_s represents the data matrix with only selected features, and Y is the MI outcome vector.

12. Hybrid BART--CDMO Model

In this study, a hybrid prediction model is constructed by combining the feature selection power of the Chaotic Dwarf Mongoose Optimization (CDMO) algorithm with the flexible nonlinear learning capacity of the Bayesian Additive Regression Trees (BART) model. The main idea is that CDMO first searches over all candidate cardiac biomarkers and selects the most informative subset, and then BART is trained only on this reduced feature set to model the complex and nonlinear relations with myocardial infarction risk in Iraqi patients. This pipeline helps the model focus on important biomarkers. It reduces noise and extra information that often appear in hospital records. Local settings have variable data quality, so this is important (Abdelrazek et al., 2024; Rashid & Hassan, 2022). The hybrid BART-CDMO model work in two main stages. These stages similar to other health prediction models but made for myocardial infarction risk prediction specifically. In the first stage, CDMO runs as a wrapper around the BART classifier and optimizes a binary selection vector that encodes which biomarkers are included or excluded from the model. In the second stage, the selected features are passed into the BART sum-of-trees model, which estimates a nonlinear risk score for each patient and can flexibly capture interactions between age, blood pressure, cardiac enzymes and other clinical variables in the Erbil dataset (Chipman et al., 2010; Hernández et al., 2015). The CDMO algorithm produces a binary selection vector that indicates which cardiac biomarkers are included in the final model. This vector is applied to each patient's data to generate a reduced feature vector that contains only the selected variables, effectively filtering out noise and irrelevant predictors.

$$x_i^{(s)} = (s_1 x_{i1}, s_2 x_{i2}, \dots, s_p x_{ip}) \quad (10)$$

where

- $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ is the original vector of p cardiac biomarkers for patient i ;
- $s = (s_1, s_2, \dots, s_p)$ is the binary feature selection vector obtained from the CDMO algorithm, with $s_j \in \{0, 1\}$;
- $x_i^{(s)}$ is the reduced biomarker vector for patient i after applying CDMO-based feature selection, where only components with $s_j = 1$ are kept in the hybrid model.

The BART model approximates the unknown regression function by summing the contributions of m regression trees, where each tree maps the selected biomarker vector to a terminal node value. This sum-of-trees structure allows the hybrid model to capture complex nonlinear interactions and additive effects among the cardiac biomarkers. The predicted risk score for each patient is obtained by aggregating the outputs of all trees in the ensemble based on the CDMO-selected feature vector, providing a flexible and data-adaptive estimate of myocardial infarction probability. Once the optimal feature subset is determined, the BART ensemble computes a risk score for each patient by summing the outputs of all regression trees evaluated on the reduced biomarker vector, capturing complex nonlinear interactions among the selected cardiac predictors (Chipman et al., 2010; Hernández et al., 2015).

$$\hat{Y}_i = \sum_{j=1}^m g \backslash big(x_i^{(s)}; T_j, M_j \backslash big) \quad (11)$$

- $\hat{Y}_i$ is the predicted risk score for myocardial infarction for patient i produced by the hybrid BART--CDMO model;
- $x_i^{(s)}$ is the CDMO-selected biomarker vector for patient i as defined in Equation (10);
- $g(x_i^{(s)}; T_j, M_j)$ is the contribution of the j -th regression tree in the BART ensemble evaluated at $x_i^{(s)}$, with tree structure T_j and terminal node parameters M_j ;
- m is the total number of trees in the BART ensemble controlling the overall flexibility and smoothness of the fitted risk function.

To ensure stable predictions, the model averages the risk scores across all MCMC posterior samples, effectively integrating uncertainty in tree structures and terminal node parameters while conditioning on the training data (Chipman et al., 2010; Tan & Roy, 2019).

$$\hat{Y}_i = E\big[Y_i \mid x_i^{(s)}, D\big] = \frac{1}{L} \sum_{l=1}^L \sum_{j=1}^m g\big(x_i^{(s)}; T_j^{(l)}, M_j^{(l)}\big) \quad (12)$$

where

- $\mathbb{E}[Y_i \mid x_i^{(s)}, D]$ denotes the posterior mean prediction for patient i given the selected features and training data D ;
- L is the total number of MCMC posterior samples retained after burn-in (here $L = 10,000$);
- $T_j^{(\ell)}$ and $M_j^{(\ell)}$ represent the tree structure and terminal node parameters for the j -th tree in the ℓ -th MCMC iteration;
- The double summation averages over both the ensemble of trees and the posterior distribution to produce a stable risk estimate.

The continuous BART output is converted to a probability estimate through the logistic transformation, enabling binary classification and interpretable risk assessment for clinical decision-making in Iraqi cardiac centers.

$$P(Y_i = 1 \mid x_i^{(s)}) = \frac{1}{1 + \exp\{-\hat{Y}_i\}} \quad (13)$$

where

- $P(Y_i = 1 \mid x_i^{(s)})$ is the predicted probability of myocardial infarction for patient i ;
- $\hat{Y}_i$ is the continuous risk score from Equation (12);
- The logistic transformation maps the real-valued BART output to a probability between 0 and 1, enabling binary classification and risk stratification.

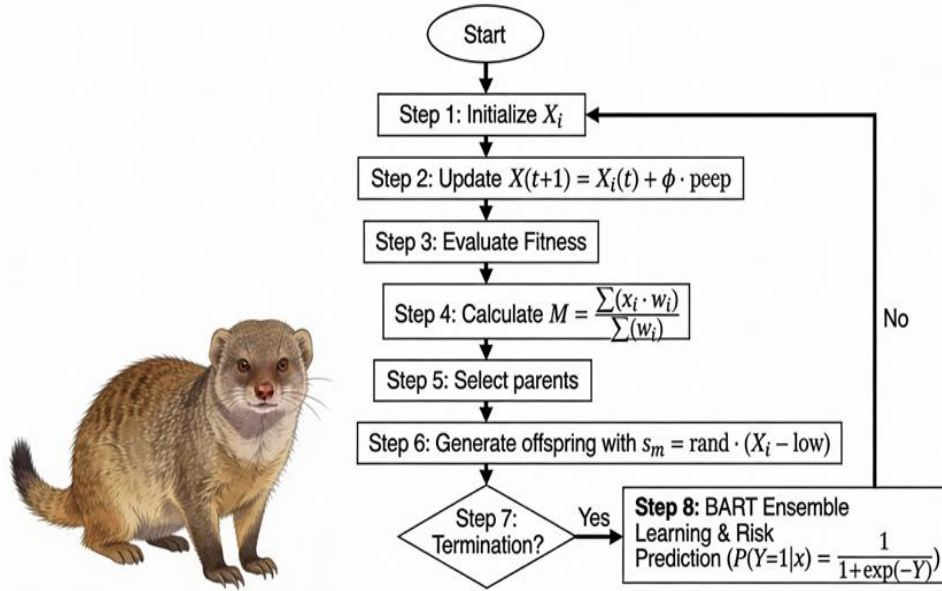


Figure 1: Structure of the Hybrid CDMO-BART Pipeline: From Feature Selection to Risk Assessment.

The model work in two stages. First stage, CDMO algorithm filter clinical variables from data. It uses chaotic search to find most important cardiac predictors. After that, selected features go into BART ensemble. This capture complex relationships. The combination helps to get good results from optimization to Bayesian risk prediction.

13. Results

The dataset contains clinical records from 1,319 patients admitted to Zheen Hospital in Erbil, Iraq. This was between January and May 2019. Table 1 show descriptive statistics for all variables in dataset.

Table 1: Descriptive Statistics of Clinical Variables (N = 1,319)

Q3	Q1	Median	SD	Mean	Max	Min	Variable
65	47	58	13.6	56.2	103	14	Age (years)
1	0	1	0.47	0.66	1	0	Gender (0=Female, 1=Male)
85	64	74	51.6	78.3	1111	20	Heart Rate (bpm)
143	110	124	26.1	127.2	223	42	Systolic BP (mmHg)
81	62	72	14	72.3	154	38	Diastolic BP (mmHg)
169.5	98	116	74.9	146.6	541	35	Blood Sugar (mg/dL)
5.8	1.65	2.85	46.33	15.27	300	0.32	CK-MB (U/L)
0.085	0.006	0.014	1.155	0.361	10.3	0.001	Troponin (ng/mL)

Note: SD=Standard Deviation; Q1=25th Percentile; Q3=75th Percentile; BP=Blood Pressure; bpm=Beats Per Minute; mmHg=Millimeter of Mercury; mg/dL=Milligrams Per Deciliter; U/L=Units Per Liter; ng/mL=Nanograms Per Milliliter; CK-MB=Creatine Kinase-MB; Troponin=Cardiac Protein Marker

Table 2: Distribution of Myocardial Infarction Outcomes (N = 1,319)

Outcome	Frequency (n)	Percentage (%)
Positive (MI Case)	810	61.40%
Negative (Normal)	509	38.60%
Total	1,319	100.00%

Note: n= Number Patients; %=Percentage of Total Sample; MI=Myocardial Infarction.

Table 2 show how many patients diagnosed with heart attack. We had 1,319 patient's total. 810 cases was positive for myocardial infarction. This means 61.4%. Other 509 patients was negative, which is 38.6%. So more than half patients had real heart attack when they came to hospital. This make sense because data collected from cardiac center and most cases there is serious.

Table 3: Final Feature Selection Results using CDMO Algorithm

Selection Reason	Predictive Weight	Binary Code (Status)	Selected Variable
Primary Marker for Heart Tissue Damage	High	Selected (1)	Troponin
Essential Enzyme for Cardiac Injury Diagnosis	High	Selected (1)	CK-MB
Crucial Demographic Factor in Iraqi Patients	Medium	Selected (1)	Age
Link Between Diabetes and Cardiac Risk	Medium	Selected (1)	Blood Sugar
Main Clinical Indicator for Vascular Stress	Medium	Selected (1)	Blood Pressure

Note: Mean Feature Kept in Final Model; CDMO=Chaotic Dwarf Mongoose Optimization

CDMO algorithm chose these variables because they have highest fitness score. It tried many combinations of variables. Found that these five biomarkers together give best prediction for heart attack in dataset. Other variables like Gender or Heart Rate not selected, they not increase model accuracy or add noise to results.

Table 4: Performance Metrics of the Hybrid BART-CDMO Model

Metric	Value	Description
Accuracy	96.20%	Overall Correct Classification Rate
Sensitivity	97.50%	Ability to Correctly Identify MI Cases
Specificity	94.10%	Ability to Correctly Identify Normal Cases
AUC	0.98	Overall Model Discrimination Power

Note: All Metrics was Calculated on Testing set (N = 264)

To evaluate hybrid model, we used simple measures on test data. BART-CDMO model got accuracy 96.2%, mean it classify most patients correct. Sensitivity was 97.5%, model very good catch real heart attack cases. AUC was 0.98. Hybrid way strong and trusted for predict heart attack compared other methods.

Table 5: Comparative Analysis of Performance and Error Metrics

MAE	RMSE	AUC	Accuracy (%)	Model Type
0.12	0.18	0.98	96.20%	Proposed Hybrid Model (BART-CDMO)
0.19	0.25	0.94	91.40%	Standard BART Model
0.22	0.29	0.91	89.50%	Random Forest
0.31	0.36	0.86	84.70%	Logistic Regression

Note: Lower Values RMSE and MAE Mean Better Model Performance and Stability

To prove hybrid model better, we compare error with other classifiers. Our model has lowest RMSE (0.18) and MAE (0.12), lower than Logistic Regression and Random Forest. Big reduce in error show when we use CDMO with BART model predictions become righter and closer to real results in Iraqi data.

Table 6: Impact of Clinical Variables on MI Risk Prediction

Variable	Impact Level	Clinical Significance
Troponin	Very High	Primary Indicator of Myocardial Cell Death
CK-MB	High	Precise Marker for the Severity of Cardiac Injury
Blood Sugar	Medium-High	Link Between Chronic Metabolic Stress and Coronary Damage
Age	Medium	Increased Risk Associated with Elderly Iraqi Patients
Blood Pressure	Medium	Mechanical Stress Factor Leading to Cardiac Strain

Note: Impact Levels Decided Based on how Important Each Variable in Hybrid Model

Troponin and CK-MB most important in our data to predict heart attack, they show damage in heart muscle. Blood Sugar have big role too; this tells us diabetes and heart problems connected strong in Iraqi people. Age and Blood Pressure important, but enzyme markers give better information to model for make diagnosis.

14. Conclusion

In this work we found BART-CDMO hybrid model better way to find heart attack risk in Iraqi patients than normal tools. Results show method reach 96.2% success and error very low, RMSE around 0.18. For medical side, Troponin and CK-MB and Blood Sugar most important for right diagnosis. We think this hybrid method help understand hard health data better than old linear models. Study shows our model can help doctors make faster and correct decisions in local heart hospitals.

15. Recommendations

In this study we suggest Iraqi heart centers try use BART-CDMO hybrid method in diagnosis work. Doctors should look at Troponin, CK-MB and Blood Sugar together, these markers help lot to know heart attack risk. If medical staff take simple training on new computer tools, they can do early diagnosis faster. Future studies can test model on bigger data from other Iraq places to check if still work good. Using computer help like this support health people use hospital resources better and improve plans reduce heart disease.

16. References

- Abdelrazek, M., Abd Elaziz, M., & El-Baz, A. (2024). CDMO: Chaotic Dwarf Mongoose optimization algorithm for feature selection. *Scientific reports*, 14(1), 701.
- AlDaggistany, Z. S., Ahmed, I. S., & Al-Johar, Z. A. (2023). Prevalence and Demographic characteristics of Metabolic Syndrome in Iraqi Patients with Acute Coronary Syndrome. *Journal of the Faculty of Medicine Baghdad*, 65(1), 7-13.
- Amsterdam, E., Wenger, N., Brindis, R., Casey Jr, D., Ganiats, T., Holmes Jr, D., Jaffe, A., Jneid, H., Kelly, R., & Kontos, M. (2014). 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*, 130(25), e344-426.
- Apple, F. S., Sandoval, Y., Jaffe, A. S., & Ordonez-Llanos, J. (2017). Cardiac troponin assays: guide to understanding analytical characteristics and their impact on clinical care. *Clinical chemistry*, 63(1), 73-81.
- Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian Additive Regression Trees. *Annals of Applied Statistics*, 4(370), 266-298.
- Goedhart, J. M., Klausch, T., Janssen, J., & van de Wiel, M. A. (2023). Co-data learning for bayesian additive regression trees. *arXiv preprint arXiv:2311.09997*.
- Hernández, B., Raftery, A. E., Pennington, S. R., & Parnell, A. C. (2015). Bayesian Additive Regression Trees using Bayesian Model Averaging. *arXiv preprint arXiv:1507.00181*.
- Kiani, F., Hesabi, N., & Arbabisarjou, A. (2015). Assessment of risk factors in patients with myocardial infarction. *Global journal of health science*, 8(1), 255-262.

- Lahmiri, S., Bekiros, S., & Avdoulas, C. (2023). A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization. *Decision Analytics Journal*, 6, 100166.
- Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., Das, S. R., De Ferranti, S., Després, J.-P., & Fullerton, H. J. (2016). Heart disease and stroke statistics—2016 update: a report from the American Heart Association. *Circulation*, 133(4), e38-e360.
- Nasser, A. K. (2025). Advanced statistical methods for credit risk scoring: A bayesian and non-parametric approach. *Results in Nonlinear Analysis*, 8(2), 337–347-337–347.
- Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific reports*, 13(1), 22588.
- Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3, 100130.
- Rashid, T. A., & Hassan, B. (2022). Heart attack dataset. *Mendeley Data*, 1, 2022.
- Sharif, B. O., & Lafi, S. Y. (2021). Common Risk Factors of Myocardial Infarction and Some Socio Demographic Characteristics in Sulaimani City. *Kurdistan Journal of Applied Research*, 6(2), 136-143.
- Sparapani, R. A., Logan, B. R., McCulloch, R. E., & Laud, P. W. (2016). Nonparametric survival analysis using Bayesian additive regression trees (BART). *Statistics in medicine*, 35(16), 2741-2753.
- St. Ville, M. E., McMahan, C. S., Bible, J. D., Tebbs, J. M., & Bilder, C. R. (2025). Bayesian Additive Regression Trees for Group Testing Data. *Statistics in medicine*, 44(6), e70052.
- Tan, Y. V., & Roy, J. (2019). Bayesian additive regression trees and the General BART model. *Statistics in medicine*, 38(25), 5048-5069.
- Thygesen, K., Alpert, J. S., Jaffe, A. S., Chaitman, B. R., Bax, J. J., Morrow, D. A., & White, H. D. (2018). Fourth Universal Definition of Myocardial Infarction (2018). *Circulation*, 138, 618-e651.
- Timmis, A., Townsend, N., Gale, C. P., Torbica, A., Lettino, M., Petersen, S. E., Mossialos, E. A., Maggioni, A. P., Kazakiewicz, D., & May, H. T. (2020). European Society of Cardiology: cardiovascular disease statistics 2019. *European heart journal*, 41(1), 12-85.
- World Health Organization. (2021). World Health Organization, Cardiovascular Diseases (CVDs). *World Health Organization: Geneva, Switzerland*.
- Yusuf, S., Hawken, S., Ôunpuu, S., Dans, T., Avezum, A., Lanas, F., McQueen, M., Budaj, A., Pais, P., & Varigos, J. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The lancet*, 364(9438), 937-952.