

Republic of Iraq
Ministry of Higher Education and
Scientific Research
AL-Mustaqbal University / College of Science
Department of Intelligent Medical Systems



**Design and Implementation of a Predictive System for Heart Disease
Using Machine Learning Techniques**

A Project

**Submitted to the Department of Intelligent Medical Systems, AL-Mustaqbal
University/ College of Science, as part of the requirements for obtaining a
Bachelor's degree in Intelligent Medical Systems.**

By

**Azhar Jabbar Obaid
Ihsan Haitham Ibrahim
Ahmed Khudhair Abdulameer**

Supervised

Asst. Lect. Ali Saleem Haleem

م 2026

١٤٤٧ هـ

Abstract

Cardiovascular diseases remain the leading cause of mortality worldwide, creating a critical need for early detection and effective risk prediction strategies. Recent advances in artificial intelligence and data-driven healthcare have enabled the development of predictive systems capable of supporting clinical decision-making through automated analysis of patient data. Machine learning, in particular, offers the ability to model complex nonlinear relationships among multiple cardiovascular risk factors, thereby improving prediction accuracy beyond traditional statistical approaches.

Despite these advancements, several challenges persist in heart disease prediction, including limitations of conventional risk assessment models, underutilization of structured clinical data, and the need for scalable and interpretable predictive frameworks. Many existing systems focus primarily on algorithmic performance without providing integrated, user-friendly environments suitable for real-world clinical use.

The objective of this research is to design and implement a machine learning-based Heart Disease Prediction System (HDPS) that enables accurate classification of cardiovascular risk using structured clinical attributes while providing an accessible web-based decision-support interface. The study aims to compare multiple supervised learning algorithms and identify the most suitable model for deployment in predictive healthcare applications.

To achieve this objective, a dataset of 5000 anonymized patient records containing demographic and physiological cardiovascular indicators was prepared and processed through a comprehensive preprocessing pipeline including encoding, normalization, and feature selection. Four supervised machine learning algorithms—Logistic Regression, Random Forest, Gradient Boosting, and XGBoost—were trained and evaluated. The predictive engine was integrated into a Streamlit-based web application supporting real-time individual prediction and batch analysis. Feature importance analysis was employed to provide global interpretability of model behavior.

Model performance was evaluated using standard classification metrics including Accuracy, Precision, Recall, F1-score, and Area Under the ROC Curve (AUC). Experimental results demonstrated high predictive performance across all models, with all four classifiers achieving accuracy exceeding 96%. Among these models, the Random Forest classifier achieved the best overall performance, reaching an accuracy of **97.4%**, precision of **97.31%**, recall of **96.88%**, and an F1-score of **97.09%**. The developed system also achieved low prediction latency with an average response time of **0.42 seconds**, while maintaining strong usability characteristics through its web-based interface. These results indicate that the proposed system can serve as an effective machine learning-based decision-support tool for early cardiovascular risk screening and assist healthcare professionals in improving early diagnosis and preventive care.