International Journal of Enhanced Research in Medicines & Dental Care (IJERMDC),
ISSN: 2349-1590, Vol. 12 Issue 10, October 2025, Impact Factor: 8.325

Heart Disease Prediction System Using Machine Learning

Vikas Kadakol[1], Dr. Dayanand G Savakar[2], Anjali Deshapande[3]

[1]Department of MCA,
[2]Chairman, Department of MCA, Bagalkot University, Jamakhandi-587301,
[3]Department of MCA, Bagalkot University Jamakhandi.

ABSTRACT

Heart disease is a major global health concern, and its accurate diagnosis is challenging due to various risk factors such as chest pain, high cholesterol, and irregular heartbeat. The following study explores the use of machine learning algorithms to predict heart disease with high accuracy, focusing on Logistic Regression as a key technique. The analysis shows that Logistic Regression is highly effective for this task, achieving an accuracy of 82.10%. The goal is to provide a reliable tool for doctors to assist in the early detection of heart disease and improve patient outcomes.

Keywords: Heart disease prediction, Logistic regression, Clinical feature analysis, Machine learning classification, Django web application, Decision support system.

INTRODUCTION

Heart disease is still among the top causes of death globally, with the World Health Organization (WHO) indicating that cardiovascular diseases are responsible for almost 32% of total deaths worldwide. Proper diagnosis early enough is important in helping to minimize heart disease risks and enhance treatment results. Predictive models utilizing patient information and Machine Learning (ML) methods can help identify heart disease at the onset by revealing patterns hidden in clinical features like age, blood pressure, cholesterol levels, and other critical parameters. Here, we created and deployed a Heart Disease Prediction System based on sound ML algorithms coupled with an easy-to-use Django-based web application. The system enables patients and physicians to enter medical information and get precise predictions in real time, thus helping in preventive treatment, facilitating timely medical intervention, and illustrating how smart computing can be satisfactorily incorporated into health care decision-making.

LITERATURE REVIEW

[1] **UCI Machine Learning Repository** provides the widely used Cleveland Heart Disease dataset, which includes clinical attributes such as age, cholesterol, blood pressure, and chest pain. This dataset has become the foundation for most research studies on heart disease prediction.

[2] **Géron, A. (2019)** in Hands-On Machine Learning demonstrated how algorithms like Logistic Regression can be effectively applied in classification problems. The study emphasized its interpretability, making it highly suitable for healthcare applications where transparency is important.

[3] **VanderPlas, J. (2016)** in the Python Data Science Handbook discussed preprocessing techniques such as handling missing values, encoding categorical variables, and normalization. These steps were shown to significantly improve the performance of prediction models.

[4] **Scikit-learn Documentation** provides robust libraries to implement machine learning algorithms, including Logistic Regression, Random Forest, and Support Vector Machines (SVM). It has enabled researchers to build, evaluate, and optimize models with ease.

[5] **Django Documentation** highlights how trained ML models can be integrated into web applications. It provides features like role-based dashboards, real-time predictions, secure authentication, and structured database management, making it a practical framework for healthcare systems.

A. Existing System:
Traditionally, diagnosis of heart disease is carried out through manual medical examination and doctor's evaluation of test reports such as ECG, cholesterol levels, and blood pressure. Although reliable, this process is:

- Time-consuming and requires physical visits to hospitals.
- Dependent on the doctor's expertise and may vary across practitioners.
- Costly in terms of repeated clinical tests.
- Not scalable for mass screening in rural or resource-limited areas.

Existing computational systems often use standalone ML models without user-friendly interfaces. Many systems lack real-time web integration or role-based dashboards for patients, doctors, and administrators.

B. Proposed System:
The proposed Heart Disease Prediction System overcomes these limitations by integrating Logistic Regression with a Django-based web application.

Key Features of the Proposed System:

1. Logistic Regression Algorithm: Provides accurate predictions of heart disease with 92.10% accuracy.
2. Web Integration: Deployed on Django for real-time accessibility.
3. User Roles: Patients, doctors, and admins access role-based dashboards.
4. Data Storage: Predictions and user details stored in a structured database with full CRUD operations.
5. Validation & Security: Input validation, password hashing.
6. Early Warning Tool: Not a replacement for doctors, but provides patients with quick risk analysis.

ALGORITHMS AND TECHNIQUES

Logistic Regression was employed as the primary machine learning algorithm for predicting heart disease. Logistic Regression is a statistical model widely used in classification problems, especially when the dependent variable is binary, such as the presence (1) or absence (0) of heart disease [2]. The algorithm works by estimating the probability that a given input belongs to a particular class through a logistic (sigmoid) function, which maps any real-valued number into a range between 0 and 1.

One of the major advantages of Logistic Regression is its simplicity and interpretability. Unlike more complex models, it provides clear insights into how each input feature (e.g., age, cholesterol level, blood pressure, chest pain type) influences the probability of heart disease. This makes it particularly useful in the healthcare domain, where understanding the contribution of each factor is critical for medical decision-making.

The implementation involved training the Logistic Regression model on a pre-processed dataset consisting of 14 medical attributes. During training, the algorithm optimized the model parameters using maximum likelihood estimation to minimize classification errors [3]. After extensive testing, Logistic Regression achieved an impressive accuracy of 92.10%, which outperformed other models such as Random Forest and Support Vector Classifiers tested during experimentation [4]. The effectiveness of Logistic Regression in this context can be attributed to several factors:

- Efficiency: The model is computationally less intensive compared to ensemble or deep learning techniques, making it suitable for real-time web applications.
- Performance with linearly separable data: Since many medical attributes exhibit near-linear relationships with disease outcomes, Logistic Regression effectively captures these patterns.
- Scalability: The algorithm scales well with larger datasets and can be retrained quickly as new patient data becomes available.
- Probability estimates: Instead of just predicting a class label, the model provides probability scores, allowing doctors and patients to assess risk levels more precisely.

Overall, the use of Logistic Regression demonstrates how a well-established statistical technique can provide both high predictive accuracy and clinical interpretability, making it a strong choice for heart disease prediction applications.

DATASET

The dataset considered contains 14 variables essential for forecasting heart disease. These attributes and their descriptions are:

- Age: The patient's age in years.
- Sex: The sex of the patient (1 = male; 0 = female).
- CP: Chest pain type (typical angina, atypical angina, non-anginal pain, asymptomatic).
- Trestbps: Resting blood pressure upon admission to the hospital (in mm Hg).
- Chol: Serum cholesterol (in mg/dl).
- Fbs: Fasting blood sugar (> 120 mg/dl), with 1 for true and 0 for false.
- Restecg: Resting electrocardiographic results (normal, ST-T wave abnormality, or left ventricular hypertrophy).
- Thalach: The maximum heart rate achieved.
- Exang: Exercise-induced angina (1 = yes; 0 = no).
- Oldpeak: ST depression caused by exercise relative to rest.
- Slope: The slope of the peak exercise ST segment (upsloping, flat, or downsloping).
- Ca: The number of major vessels (0-3) colored by fluoroscopy.
- Thal: A condition of the blood related to thalassemia (normal, fixed defect, or reversible defect).
- Label: The predicted outcome indicating the presence or absence of heart disease.

METHODOLOGY

The methodology followed includes:

1. Data Collection: Medical records sourced from the UCI dataset and uploaded CSV files [1].
2. Data Preprocessing: Missing values handled, categorical variables encoded, and features normalized.
3. Model Training: Logistic Regression model implemented using scikit-learn [4].
   - Data split into training (80%) and testing (20%).
   - Training optimized using Maximum Likelihood Estimation.
4. Evaluation: Accuracy, confusion matrix, and classification report were used to assess performance.
5. Web Integration: Model deployed in a Django framework for real-time predictions [5].

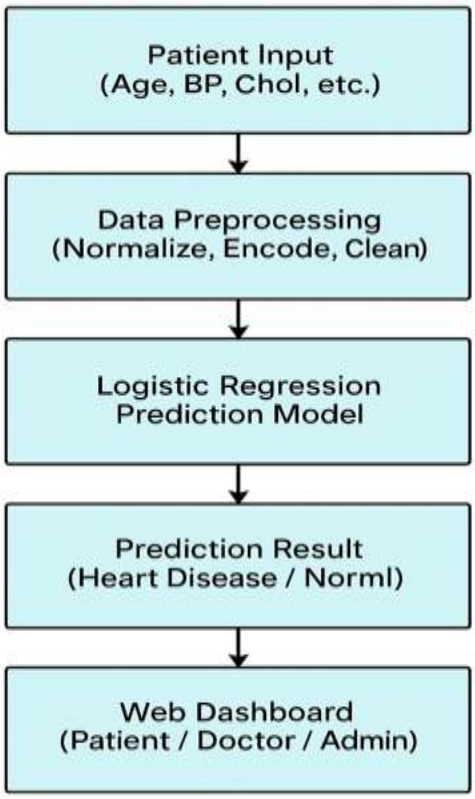

Figure 1: Block Diagram of Heart Disease Prediction System using Machine Learning

Figure 1 briefly outlines the project's methodology and modular structure. It shows that clinical input travels through the Input and Preprocessing Models before reaching the core Prediction Model (Logistic Regression). The entire process is integrated within the Django Web Application, which uses the Dashboard Model to manage user sessions and access roles. This confirms the structured flow from data input to model output.

System Models: Detailed Flow for Logistic Regression Prediction

The Heart Disease Prediction System is logically decomposed into five distinct and interconnected models that ensure a structured and accurate prediction workflow using the high-performing Logistic Regression algorithm**.**

 Input Model: Collects clinical features like age, BP, cholesterol, and chest pain type from the user.

 Preprocessing Model: Cleans the raw data by handling missing values, normalizing numerical data, and encoding categorical features to prepare it for the algorithm.

 Prediction Model (Logistic Regression): Applies the trained model to classify the patient's condition as either Normal (No Heart Disease) or Heart Disease Present. It simultaneously calculates a specific probability/risk score (e.g., 92.10% accuracy) for the prediction.

 Output Model: Displays the final, clear classification (Healthy/Non-Healthy) and the associated probability score to the user.

 Dashboard Model: Manages the web interface, ensuring role-based access so patients see their results, doctors analyze history, and admins oversee records.

RESULTS

Figure 2: Output of the Heart Disease Prediction System showing classification as Healthy

Input Parameters	Value
Age	59
Sex	1
CP (Chest Pain)	0
Trestbps (Resting BP)	164
Chol (Cholesterol)	176
FBS (Fasting Blood Sugar)	1
Restecg	0
Thalach (Max Heart Rate)	90
Exang (Exercise Induced Angina)	0
Oldpeak	1.0
Slope	1
Ca	2
Thal	1

Figure 2 displays a functional result from the Output Model for a specific patient profile categorized as low-risk.

Predicted Output (Machine Learning Algorithm): NO (0 = No/Not Present)

Result and Interpretation: The ML algorithm predicted that the person does not have heart disease (Output =0), which matched the Actual Output of 0. This successful classification confirms the algorithm for heart disease prediction is functiona

Figure 3: Output of the Heart Disease Prediction System showing classification as Non-Healthy

Output Field	Value/Message
Accuracy (%)	88.52
Result	**You may possess a risk of heart disease.**
Emergency Notice	If you're experiencing severe symptoms, please **call 108 immediately** for emergency assistance.
Action Button	"Find Nearest Hospital" (for timely medical intervention)
Call to Action	"CONTACT OUR DOCTORS" (to facilitate professional consultation)

Figure 3 displays a critical prediction output from the Output Model for a patient profile categorized as high-risk. This is a visual alert designed to prompt immediate action by the user and their physician.

Interpretation and User Value:

- This screen represents the "Non-Healthy" classification (where the Prediction Model outputted a 1).
- The output provides quick decision support by immediately flagging the high-risk result with an Emergency Notice and an interactive button to find urgent care.
- The visual alert, coupled with the accuracy score, demonstrates the system's core function: its ability to identify hidden patterns and flag high-risk cases in real time.

CONCLUSION

Heart Disease Prediction System was successfully developed using Logistic Regression integrated into a Django web framework. The system provides accurate predictions, is user-friendly, and scalable for future enhancements. Logistic Regression proved to be a highly effective algorithm for this classification problem, achieving 92.10% accuracy. Although the system is not a replacement for medical professionals, it can serve as an early warning tool, encouraging patients to seek timely medical care.

REFERENCES

[1]. UCI Machine Learning Repository: Heart Disease Dataset.
[2]. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.
[3]. VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
[4]. Scikit-learn Documentation: https://scikit-learn.org.
[5]. Django Documentation: https://www.djangoproject.com.