

Design and Implementation of Retinal Image Based Heart Disease Prediction and Risk Management Using Machine Learning

Mrs. Smita S Wagh¹, Mr. Tejas Khatkale², Ms. Devshri Khairnar³, Mr. Yash Mali⁴, Ms. Sanika Najan⁵
^{1,2,3,4,5}*Department of Computer Engineering JSPM's Jayawantrao Sawant College of Engineering, Pune, India*

Abstract: Cardiovascular diseases like hypertension and heart attacks impact microvascular structures. Fundus imaging helps detect retinal blood vessel abnormalities linked to these conditions. Machine learning and AI identify preclinical symptoms beyond human observation. This study uses vessel segmentation to analyse retinal vasculature, aiding early heart disease detection, especially in young individuals. Retinal imaging supports diagnosis, treatment planning, and clinical assessment across ophthalmology and cardiology. The findings emphasize vessel segmentation's role in medical diagnostics.

Index Terms: AI diagnosis, Cardiology, Cardiovascular disease, Fundus imaging, Heart attack, Hypertension, Machine learning, medical image processing, Ophthalmology, Retinal blood vessels, Retinal vessel segmentation.

1. INTRODUCTION

The heart is a type of muscular organ which circulates blood into the body and forms the core component of the body's cardiovascular system which also includes lungs. The cardiovascular system also includes a system of blood vessels, for instance, veins, arteries, and capillaries. The blood vessels supply blood throughout the body. Irregularities in normal circulation of blood from the heart result in a number of different kinds of heart diseases which are popularly referred to as cardiovascular diseases (CVD). Heart diseases are the principal causes of mortality across the globe. In the World Health Organization (WHO) survey, it has been reported that 17.5 million total deaths all over the world are due to heart attacks and strokes. Over 75% of cardiovascular disease deaths happen primarily in middle-income and low-income nations. Furthermore, 80% of the fatalities caused by CVDs are due to stroke and heart attack. Hence, early prediction of cardiac abnormalities and prediction tools for heart diseases can save a lot of life and assist

physicians to develop an effective treatment protocol which reduces the death rate by cardiovascular diseases. Thanks to the evolution of advanced healthcare systems, there is ample patient data available today (i.e. Big Data in Electronic Health Record System) that can be utilized for modeling predictive models for cardiovascular diseases. Machine learning or data mining is one discovery process for analyzing big data from a diverse point of view and condensing it into valuable information. "Data Mining is a non-trivial extraction of implicit, previously unknown and potentially useful information about data". Today, a tremendous amount of data relating to disease diagnosis, patients etc.

1.1 MOTIVATION

One of the most significant challenges confronting healthcare organizations (hospitals, medical centers) is providing quality services at reasonable prices. Quality service means correctly diagnosing patients and treating them with effective treatments. Wrong clinical decisions can result in catastrophic outcomes which are thus unacceptable. Hospitals should also reduce the expense of clinical tests. They can do so by using suitable computer-based information and/or decision support systems.

Majority of the hospitals nowadays utilize some form of hospital information systems for handling their health care or patient information [12]. Such systems usually produce copious amounts of data that occur in the forms of numbers, text, chart and images. Sadly, most of these data are hardly exploited to inform decision making at a clinical level. There is substantial hidden information buried in these data that remains considerably untapped. This poses a significant question: "How do we transform data into meaningful information that can empower healthcare

professionals to make smart clinical decisions? "This is the primary incentive for this research."

1.2 PARAMETER BASED Vs RETINAL IMAGE BASED PREDICTION

Parameter-Based Prediction: The specific measurable feature extracted from the retinal images should rather be critical in determining what level of a patient's engagement with AVR, vessel width, tortuosity, and the presence of microaneurysms-characteristic signs of cardiovascular risk. Since these are considered as biomarkers in the context of a machine learning model to classify the patient at a higher or lesser risk of heart disease, then what is of importance in the extent of the patient's engagement to real-time image analysis becomes relevant. The retinal features can be understood along with the benefits associated with parameter-based prediction; the clinicians can relate these directly to known risk factors associated with cardiovascular disease. This is too reliant on proper segmentation and feature extraction that relies more on the quality of images that could introduce variability in case of manual doing or through not so optimal algorithms.

Image-Based Prediction: This is an image-based prediction that involves a model learning patterns directly from raw or pre-processed retinal images which bypasses the actual manual feature extraction step. Those deep architectures, especially CNNs, are often used for these because it has the capability of detecting complex and latent patterns in images that may not be explicitly apparent in some images. The approach allows for more accuracy and robustness because the model automatically captures intensive and subtle features in retinal images. An important drawback is to reduce the interpretability of image-based prediction: deep learning models are highly esteemed as "black boxes," and which features may have led to a certain prediction is somewhat hard to understand clinically. That demand for large amounts of data and computationally expensive resources in training may correspond to costs not all healthcare facilities can afford.

Table no. 01 Parameter Based Vs Retinal Image Based Prediction

ASPECTS	PARAMETER BASED PREDICTION	IMAGE BASED PREDICTION
---------	----------------------------	------------------------

Feature Extraction	It is a manual or algorithm-based extraction of some parameters.	Model learns features directly from images.
Interpretability	High; individual parameters are understandable for clinicians.	Lower; Model decisions are less interpretable.
Accuracy	Moderate; this depends on how the features are extracted with accuracies at every point.	High; The model was able to identify autonomous complex patterns.
Model Complexity	Usually much lower, especially for simpler ML models.	High, requiring Deep learning models like CNNs.
Dependency on image quality	Sensitive to the quality of the image at feature extracting stage.	Needs high quality images but may generalize better.
Computational requirement	It demands fewer computing resources; hence, lower.	It has higher requirements; it requires big data sizes and more powerful GPUs
Data requirements	Moderate; can work even with small amounts, if the parameters are consistent.	High, deep learning models require a large amount of datasets.
Scalability	Highly- Scalable in low resource settings.	This application requires low scalability; it's hardware is special.

2. PRELIMINARIES

This section outlines the challenges and solutions in developing an ML-based system using retinal imaging for heart disease prediction. The goal is to create a non-invasive, accessible, and accurate diagnostic tool.

Challenges and Solutions:

1. High CVD Mortality & Economic Burden: Develop a low-cost, non-invasive diagnostic tool using retinal image analysis for early detection, especially in low-income regions.
2. Limited Access to Conventional Diagnostics: Utilize fundus imaging and ML to enable early, remote diagnosis beyond specialized cardiac centres.

- 3. Complex & Time-Consuming Medical Data Analysis: Integrate optimized ML classifiers (DNN, logistic regression) for faster and more accurate risk assessment.
- 4. Reliable Risk Estimation: Implement vessel segmentation and feature extraction to quantify biomarkers (diameter, tortuosity, branching) for precise risk scoring.
- 5. Variability & Quality Issues in Retinal Images: Apply preprocessing techniques (noise reduction, normalization) to enhance image consistency and model accuracy.
- 6. Integration with Clinical Workflows: Develop a user-friendly system (mobile/web) for seamless

adoption in healthcare settings, aiding personalized treatment plans.

7. Ethical & Privacy Concerns: Ensure strict data privacy with anonymization and secure storage to comply with medical regulations.

8. Lack of Interpretability in ML Models.

3. LITERATURE SURVEY

This review identifies AI and ML research in heart disease prediction that emphasizes deep learning, ensemble techniques, and optimization. Research responds to challenges such as class imbalance and feature selection, with the aim of early detection and better clinical usage.

LITERATURE SURVEY TABLE:

Sr.No.	Title	Proposed System	Author(s)	Parameters and method	Advantages/Disadvantages
1	A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning	Deep learning-based clinical decision support system for heart disease prediction.	A. A. Almazroi et al.	Parameters: Clinical data; Method: Deep learning neural network	Advantage: High prediction accuracy. Disadvantage: Limited dataset diversity.
2	A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method	Ensemble learning method for heart disease prediction.	Kumar, K. U. Singh, M. Kumar	Parameters: Clinical records; Method: Random Forest, Gradient Boosting	Advantage: Increased accuracy through ensemble models. Disadvantage: Computationally expensive.
3	A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks	Hybrid deep neural network system for robust heart disease detection.	M. S. A. Reshan et al.	Parameters: Patient health metrics; Method: Hybrid DNN combining CNN and LSTM	Advantage: Robust performance with complex patterns. Disadvantage: Requires significant training time.
4	An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches	Dual-stage machine learning approach for heart disease risk prediction	S. Mondal et al.	Parameters: Risk factors; Method: Stacked ML combining SVM and Logistic Regression	Advantage: Improved computational efficiency. Disadvantage: It may end up overfitting on small datasets.
5	Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV	Machine learning-based diagnosis with parameter optimization using GridSearchCV	G. N. Ahmad et al.	Parameters: Heart disease dataset; Method: SVM, Decision Trees, and GridSearchCV	Advantage: Fine-tuned models with high precision. Disadvantage: Increased complexity during parameter selection.
6	Heart Disease Prediction Using	Stacking model with balancing	A. Noor et al.	Parameters: Balancing	Advantage: Effective handling of imbalanced datasets.

	Stacking Model With Balancing Techniques and Dimensionality Reduction	and feature reduction for improved heart disease detection.		techniques; Method: Stacking ensemble model	Disadvantage: Potential loss of key features during reduction.
7	Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases	Regression analysis to estimate overlapping symptoms of cardiovascular diseases.	S. Ghorashi et al.	Parameters: Clinical symptom data; Method: Regression analysis	Advantage: It is easy to understand and interpret Disadvantage: Limited for non-linear relationships.
8	Cardiovascular Risk Factors in Patients with Congenital Hemophilia: A Focus on Hypertension	Analysis of cardiovascular risk factors, focusing on hypertension in hemophilia patients.	M. C. Badescu et al.	Parameters: Hypertension metrics; Method: Statistical analysis	Advantage: Valuable insights into a niche population. Disadvantage: Only applies to hemophilia
9	Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm	Deep learning algorithm for coronary artery segmentation and risk prediction.	C. Xiao et al	Parameters: Coronary artery images; Method: Convolutional Neural Networks	Advantage: Accurate artery segmentation. Disadvantage: Requires extensive labeled image datasets.
10	Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare	Machine learning classification for heart disease detection in e-healthcare	J. P. Li et al.	Parameters: E-health data; Method: ML classifiers (SVM, Decision Trees)	Advantage: It is easier to integrate the application of e-health. Disadvantage: Challenges with high-scale deployment.

Table no. 02 Literature survey table

4. SYSTEM DESIGN

4.1 EXISTING SYSTEM

Medical diagnoses often rely on doctors' intuition rather than data-driven insights, leading to errors, biases, and high costs. Misdiagnoses pose severe risks, affecting patient safety and trust in healthcare.

Disadvantages:

- Late-stage disease prediction.
- Time-consuming data analysis.
- High cost and dependency on human expertise.
- Risk of fatal misdiagnosis.

4.2 PROPOSED SYSTEM

A machine learning-based heart disease prediction system using retinal images enhances accuracy and accessibility. Deep learning models analyze fundus images to detect early signs of heart disease, potentially integrating into mobile and web applications for wider use.

4.3 MODULES

The entire work of this project is divided into 4 modules. They are:

- a. Data Preprocessing: Cleans and prepares input data (tokenization, stemming, handling missing values).
- b. Feature Extraction: Uses TF-IDF, word embeddings, and image processing to extract key features.
- c. Classification: Compares multiple classifiers (Naïve Bayes, SVM, Logistic Regression, DNN) to identify the best-performing model.
- d. Prediction: Deploys the optimized model to classify heart disease risk from new data.

4.4 ALGORITHM OF PROPOSED SYSTEM

1. Data Collection: Retinal images and patient health records.
2. Preprocessing: Image resizing, normalization, and augmentation.

3. Feature Extraction: Vessel width, optic disc, microaneurysms, and tabular health data.
4. Model Selection: Chooses ML models like DNN and Logistic Regression.
5. Training: Trains models on a dataset split into training and testing sets.
6. Evaluation: Measures accuracy, precision, recall, F1-score, and AUC-ROC.
7. Prediction: Assesses heart disease risk from new data.
8. Results Interpretation: Identifies key risk factors and visualizes outcomes.
9. Deployment: Implements a user-friendly interface for healthcare professionals via web or mobile applications.

5. ML ALGORITHMS

1. Logistic Regression: Makes predictions of probabilities for binary outcomes, which makes it appropriate for medical diagnosis. Maps predictions between 0 and 1 using the logistic function and estimates the likelihood of heart disease based on characteristics such as age, blood pressure, and cholesterol.
2. Deep Neural Networks (DNNs): Advanced algorithms that identify sophisticated patterns in data. Multiple decision trees are trained, and the predictions are combined to improve accuracy and minimize overfitting. DNNs are best at identifying nuanced interactions in medical data.
3. Support Vector Machine (SVM): A classification algorithm that identifies the best hyperplane to classify heart disease and non-disease cases. Handles high-dimensional data well and identifies complex patterns well.
4. Naïve Bayes: A Bayes theorem-based probabilistic classifier under feature independence. Efficient in computation and can be employed for initial medical screening.
5. K-Nearest Neighbors (KNN): Compares the patient's condition with similar conditions in the data

set to classify the patient's condition. Easy to implement but computationally intensive for large data sets.

6. Recursive Feature Elimination (RFE): Picks the most useful features by iteratively dropping the least important ones. Assists in enhancing model accuracy and interpretability by highlighting important indicators such as cholesterol and blood pressure.

7. Ensemble Voting Method: Aggregates several models (e.g., DNN and Logistic Regression) to make it more reliable and reduce errors, thus making it very effective for medical diagnosis.

6. MATHEMATICAL MODEL

1. Logistic Regression

Logistic Function: The logistic regression model predicts the probability of a heart attack occurring given the features X .

$$P(Y = 1|X) = \frac{1}{(1 + e^{-z})}, \quad z = \beta^0 + \sum \beta_i X_i$$

2. Deep Neural Network (DNN)

Applies several layers with weights W and biases b .

Output probability: $P(Y = 1 | X) = \sigma(Z^L)$

Trained with binary cross-entropy loss.

3. Support Vector Machine (SVM)

Classification tasks, SVM seeks to find the hyperplane that maximizes the margin between classes.

Decision Function:

$$\text{Minimize } \left(\frac{1}{2}\right) \|W\|^2, \text{ subject to } y^i(W \cdot X^{(i)} + b) \geq 1$$

7. PREDICTIVE RESULTS

The Ensemble Voting Method (DNN + Logistic Regression) yielded the best accuracy, minimizing bias and enhancing reliability.

- DNN: Identified subtle data interactions, providing robustness.

- Logistic Regression: Supplied probability scores, which helped inform clinical decisions.
- SVM: Efficient with easily separable classes.
- Naïve Bayes: Quick and reliable for preliminary screening.
- KNN: Performed well on small data sets but at a high computational cost.

Diagram 01: Architecture of the proposed system

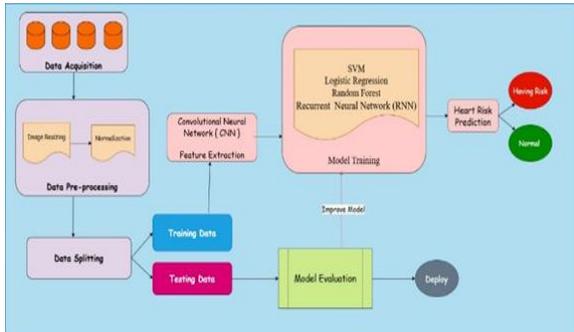


Diagram 02: Flowchart of the model

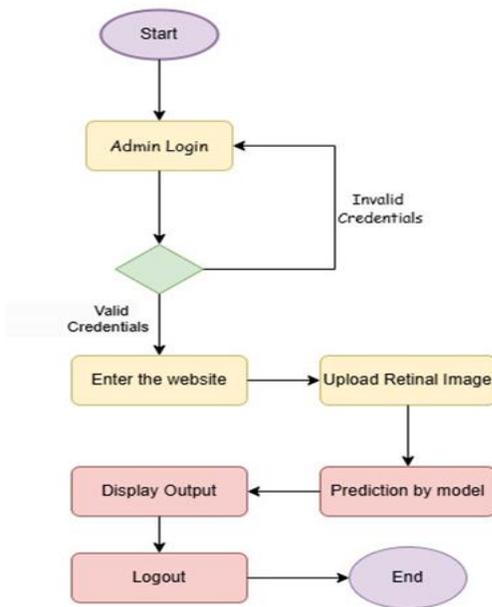


Figure no. 02 Flowchart

Figure no. 01. Architecture Diagram of Proposed System

9. CONCLUSION

The application of retinal images for the prediction of heart disease is a step forward in non-invasive diagnosis, with the goal of alleviating the burden of cardiovascular diseases (CVDs) globally. Machine

learning algorithms can predict retinal vasculature characteristics like vessel tortuosity, arteriovenous ratios, and microaneurysms to offer early risk predictions, making preventive care more accessible. Deep learning algorithms, especially CNNs, have proven effective in detecting intricate patterns in retinal images, enabling accurate risk assessment

8. DIAGRAMS

without invasive methods.

Yet, challenges remain, such as model interpretability, data quality, and generalizability across heterogeneous populations. Integration of AI into clinical environments is important to increase medical professionals' trust in AI-driven diagnostics. Furthermore, telemedicine advances can extend cardiovascular risk assessments to underserved areas. Future studies should aim to enhance model robustness by integrating multi-modal data and improving algorithm performance to process noisy or low-quality images. Early diagnosis through retinal image analysis is a milestone of AI-powered medical innovation.

10. REFERENCES

- [1] A. A. Almazroi, E. A. Aldahri, S. Bashir and S. Ashfaq, "A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning," in *IEEE Access*, vol. 11, pp. 61646-61659, 2023, doi: 10.1109/ACCESS.2023.3285247.
- [2] Kumar, K. U. Singh and M. Kumar, "A Clinical Data Analysis Based Diagnostic Systems for Heart Disease Prediction Using Ensemble Method," in *Big Data Mining and Analytics*, vol. 6, no. 4, pp. 513-525, December 2023, doi:10.26599/BDMA.2022.9020052.
- [3] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," in *IEEE Access*, vol. 11, pp. 121574-121591, 2023, doi: 10.1109/ACCESS.2023.3328909.
- [4] S. Mondal, R. Maity, Y. Omo, S. Ghosh and A. Nag, "An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches," in *IEEE Access*, vol.12, pp.7255-

- 7270, 2024,doi:
10.1109/ACCESS.2024.3350996.
- [5] G. N. Ahmad, H. Fatima, S. Ullah, A. Salah Saidi and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV," in *IEEE Access*, vol. 10, pp. 80151-80173, 2022, doi: 10.1109/ACCESS.2022.3165792.
- [6] A. Noor, N. Javaid, N. Alrajeh, B. Mansoor, A. Khaqan and S. H. Bouk, "Heart Disease Prediction Using Stacking Model With Balancing Techniques and Dimensionality Reduction," in *IEEE Access*, vol. 11, pp. 116026-116045, 2023, doi: 10.1109/ACCESS.2023.3325681.
- [7] S. Ghorashi et al., "Leveraging Regression Analysis to Predict Overlapping Symptoms of Cardiovascular Diseases," in *IEEE Access*, vol. 11, pp. 60254-60266, 2023, doi: 10.1109/ACCESS.2023.3286311.
- [8] M. C. Badescu, O. V. Badulescu, L. I. Butnariu, I. B. Bojan, M. C. Vladeanu, N. Dima, et al., "Cardiovascular risk factors in patients with congenital hemophilia: A focus on hypertension", *Diagnostics*, vol. 12, no. 12, pp. 2937, Nov. 2022.
- [9] C. Xiao, Y. Li and Y. Jiang, "Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm," in *IEEE Access*, vol. 8, pp. 140108-140121, 2020, doi: 10.1109/ACCESS.2020.3010800
- [10] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in *IEEE Access*, vol. 8, pp. 107562-107582, 2020, doi: 10.1109/ACCESS.2020.30011