

Predicting Coronary Heart Disease Using An Improved Light Gbm Model

¹Gomathi T, ²Nivedita K, ²Sindhu M, ²Soundarya P

¹ Associate professor, Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur-621212.

² Student of Biomedical Department, Dhanalakshmi Srinivasan Engineering College, Perambalur.

Abstract- Coronary Heart Disease (CHD) remains one of the leading causes of mortality worldwide, necessitating early and accurate prediction methods to improve patient outcomes. This paper proposes an efficient predictive framework using an improved Light Gradient Boosting Machine (LightGBM) algorithm for the early detection of CHD. The proposed model integrates advanced preprocessing techniques, including data cleaning, normalization, and feature selection, to enhance data quality and relevance. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is employed, thereby improving model robustness. Hyperparameter tuning is also performed to optimize model performance and reduce overfitting. The system is trained and evaluated using clinical datasets containing key attributes such as age, blood pressure, cholesterol levels, and lifestyle factors. Experimental results demonstrate that the improved LightGBM model achieves higher accuracy, precision, and recall compared to traditional machine learning approaches. Additionally, the model identifies significant risk factors contributing to CHD, supporting clinical decision-making. The proposed approach provides a reliable, scalable, and efficient solution for early CHD prediction, with potential applications in healthcare systems for preventive diagnosis and risk assessment. The improved LightGBM model not only enhances predictive accuracy but also reduces computational complexity, making it suitable for large-scale medical datasets. Furthermore, the interpretability of the model, achieved through feature importance analysis, enables healthcare professionals to better understand contributing risk factors and take proactive preventive measures. This approach bridges the gap between data-driven insights and clinical practice, ultimately contributing to improved patient care, early intervention, and reduced mortality associated with coronary heart disease.

Keywords: Coronary Heart Disease (CHD), LightGBM (Light Gradient Boosting Machine), machine learning, predictive modeling, healthcare analytics, feature selection, hyperparameter tuning, SMOTE, risk prediction, and clinical decision support system.

I. INTRODUCTION

Coronary Heart Disease (CHD) is a critical cardiovascular disorder characterized by the obstruction or narrowing of coronary arteries due to atherosclerotic plaque accumulation, leading to reduced myocardial blood supply. It is a primary contributor to global morbidity and mortality, emphasizing the need for early, accurate, and automated diagnostic systems. Traditional diagnostic approaches and statistical prediction models often exhibit limited performance when

handling high-dimensional, nonlinear, and imbalanced medical data.

Recent advancements in Artificial Intelligence (AI), machine learning, and medical image processing have enabled the development of intelligent systems capable of assisting clinicians in disease detection and prognosis. Image-based diagnostic modalities such as echocardiography and coronary angiography provide valuable structural and functional information; however, manual interpretation is time-intensive and prone to inter-observer variability. Automated image analysis

combined with machine learning classifiers offers a scalable and objective solution to these challenges. In this project, a hybrid framework integrating image processing techniques with an optimized Light Gradient Boosting Machine (Light GBM) model is proposed for the prediction of coronary heart disease. Initially, image pre-processing and enhancement techniques are applied to improve image quality and suppress noise. K-means clustering-based segmentation is then utilized to isolate coronary artery regions of interest. From the segmented images, relevant statistical and handcrafted features are extracted and supplied to the Light GBM classifier for binary classification into CHD and normal classes.

To enhance predictive performance and address class imbalance, the Light GBM model hyperparameters are optimized using the OPTUNA optimization framework, and an improved focal loss function is incorporated to emphasize hard-to-classify samples. The proposed system demonstrates improved accuracy, robustness, and generalization capability compared to conventional methods. This framework highlights the effectiveness of combining classical image processing with advanced machine learning techniques for reliable and early CHD prediction, supporting clinical decision-making and improving patient outcomes.

II. METHOD

1. Image acquisition:

Echocardiogram and angiography images are collected from the medical image dataset and clinical repositories. These images serve as the primary input for the system. The acquired images may vary in resolution, contrast, and noise levels, making further preprocessing essential for accurate analysis.

2. Image Preprocessing:

In this stage, the acquired images undergo preprocessing to enhance image quality and remove unwanted noise. Techniques such as grayscale conversion, image resizing, noise reduction using filtering methods, and contrast enhancement are applied. Preprocessing ensures uniformity in image

dimensions and improves the visibility of coronary artery structures, which is crucial for reliable segmentation and feature extraction.

3. Image Segmentation

K-means clustering is employed to segment the preprocessed images and isolate the coronary artery regions of interest. This segmentation process helps in separating relevant anatomical structures from the background, reducing unnecessary information and improving the accuracy of subsequent feature extraction.

4. Feature Extraction

After segmentation, meaningful features are extracted from the coronary artery regions. These include statistical features, texture-based features, and intensity-related attributes that effectively represent the characteristics of normal and diseased arteries. The extracted features form the input feature set for the classification model.

5. Dataset Preparation

The extracted feature dataset is divided into training and testing sets. This separation ensures unbiased evaluation of the model's performance. Proper dataset splitting helps in validating the generalization capability of the proposed system.

6. Light GBM Model Initialization

A Light Gradient Boosting Machine (Light GBM) classifier is initialized for predicting coronary heart disease. Light GBM is selected due to its high efficiency, fast training capability, and ability to handle large-scale datasets with complex feature interactions.

7. Hyperparameter Optimization

To enhance the performance of the Light GBM model, hyperparameters are optimized using the OPTUNA optimization framework. OPTUNA systematically searches for the best parameter combinations, reducing overfitting and improving model accuracy and robustness.

8. Handling Class Imbalance

An improved focal loss function is incorporated into the model training process to address class

imbalance in the dataset. This loss function gives higher importance to difficult-to-classify CHD cases, thereby improving sensitivity and reducing false negatives.

9. Classification

The optimized Light GBM model performs binary classification, categorizing the input images into coronary artery disease (CAD) and Normal classes. The trained model predicts disease presence based on learned patterns from the extracted features.

10. Performance Evaluation

The performance of the proposed system is evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC. These metrics provide a comprehensive assessment of the model's effectiveness and reliability for clinical decision support.

III. RESULT

The proposed framework generates a detailed series of visual outputs and performance metrics to classify Coronary Artery Disease (CAD) and map its severity. Below is the results summary and the corresponding output flow diagram.

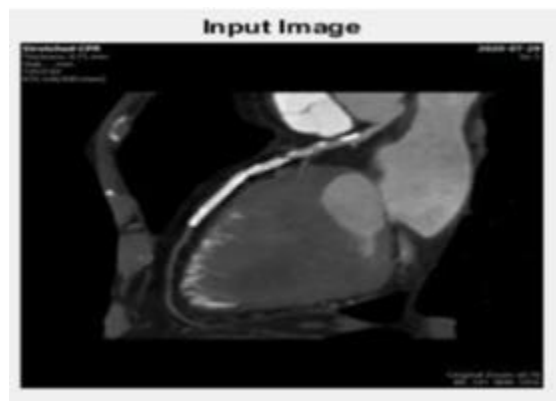


Fig : input image

Registration Results:

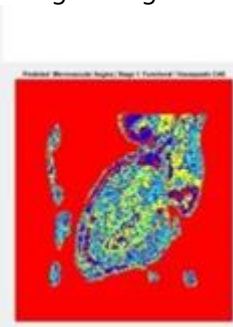
The registration and pre-processing stage ensures that the input medical images are standardized and optimized for accurate analysis. Using CLAHE (Contrast Limited Adaptive Histogram Equalization), the image contrast is enhanced to highlight subtle coronary structures, while skull/background removal

isolates the heart region from non-essential anatomical data. This alignment and enhancement process minimizes noise and spatial variances, providing a clean, normalized grayscale input that allows the subsequent AMSOM-FCM segmentation and Inception-V3 model to achieve higher diagnostic precision and reliability in identifying CAD severity.

Fig: CLAHEIMAGE



Fig: Background



removed image

Segmented Results:

The segmentation phase produces several key visual and quantitative results to ensure the accuracy of the CAD diagnosis.

Visual Results

The hybrid AMSOM-FCM (Adaptive Self- Organizing Map combined with Fuzzy C- Means) segmentation transforms the pre- processed input into a detailed map of the affected areas:

AMSOM-FCM Segmented Image: This is the primary output of the segmentation stage, which precisely isolates the coronary vessels and pathological lesions.

Background Removed Image: A prerequisite result that displays the cardiac region of interest after non-essential anatomical structures (like the skull or surrounding tissue) are suppressed.

Quantitative Validation

To ensure the reliability of the segmented regions, the framework calculates several image quality metrics:

MSE (Mean Squared Error): Measures the average squared difference between the segmented and original structural values.

PSNR (Peak Signal-to-Noise Ratio): Evaluates the quality of the segmented image relative to the noise introduced during the process.

RMSE (Root Mean Squared Error): Provides a statistical measure of how closely the segmented output matches the groundtruth.

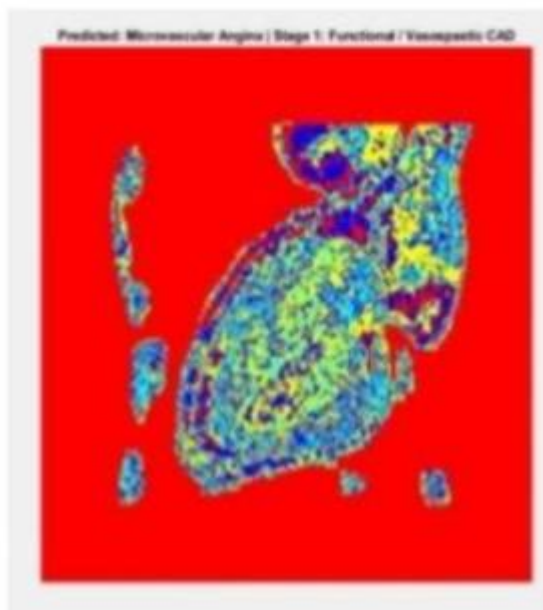


Fig : AMSOM-FCH Segmented Images.

3. E. L. Rommand I. F. Tsigelny, 'Artificial intelligence in drug treatment,' *Annu. Rev. Pharmacol. Toxicol.*, vol.60, no. 1, pp. 353–369, Jan. 2020, doi: 10.1146/annurev-pharmtox-010919-023746.
4. L. Lo Vercio, K. Amador, J. J. Bannister, S. Crites, A. Gutierrez, M. E. MacDonald, J. Moore, P. Mouches, D. Rajashekar, S. Schimert, N. Subbanna, Tuladhar, N. Wang, M. Wilms, A. Winder, and N. D. Forkert, 'Supervised machine learning tools: A tutorial for clinicians,' *J. Neural Eng.*, vol. 17, no. 6, Dec. 2020, Art. No. 062001, doi:10.1088/1741-2552/abbff2.
5. S. Rauschert, K. Raubenheimer, P. E. Melton, and R. C. Huang Machine learning and clinical epigenetics: A review of challenges for diagnosis and classification,' *Clin. Epigenetics*, vol. 12, no. 1, p. 51, Apr. 2020, doi:10.1186/s13148-020-00842-4.

REFERENCE

1. N.Katta, T. Loethen, C. J. Lavie, and M. A. Alpert, 'Obesity and coronary heart disease: Epidemiology, pathology and coronary artery imaging, *Current Problems Cardio.*, vol. 46, no. 3, Mar. 2021, Art. No. 100655, doi1016/j.cpcardiol.2020.100655. VOLUME 11, 2023 H. Yang et al.: Predicting Coronary Heart Disease Using an Improved LightGBM Model.
2. G. S. Handelman, H. K. Kok, R. V. Chandra, A.H. Razavi, M. J. H. Asadi, EDoctor: Machine learning and the