Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, No.2, August : 2024

MACHINE LEARNING WITH OPTIMIZATIONS FOR HEART DISEASE PREDICTION USING ECG DATA

Rubeena Rab, Assistant Professor, Dept of CS&AI, MJCET, O.U, Hyderabad, T.S, India. rubeena.rab@mjcollege.ac.in
Dr. G. Sailaja, Associate Professor, Dept of Mechanical Engineering, MJCET, O.U, T.S, India. sailajasinha@mjcollege.ac.in
M. Karteek Kumar Reddy, Assistant Professor, Dept of I.T, MJCET, O.U, Hyderabad, T.S, India. kartheek.kumar@mjcollege.ac.in
Asad Hussain Syed, Assistant Professor, Dept of CS&AI, MJCET, O.U, Hyderabad, T.S, India. asadhussain@mjcollege.ac.in
Zeba Ruqshanh, Assistant Professor, Dept of CSE, Bhaskar Engineering College, JNTUH, T.S, India. zebaruqshanh892@gmail.com

ABSTRACT
Heart diseases are the ones causing most of the health issues and death to human beings. With the emergence of Artificial Intelligence (AI), it is possible to exploit learning-based approaches for efficient prediction of heart diseases. The existing literature on heart disease prediction using ECG data revealed that machine learning models need optimization towards leveraging prediction performance. Motivated by this fact, in this paper, we proposed a machine learning framework for automatic detection of heart diseases from ECG data. The framework has provision for both the training and testing phases involving optimizations with feature selection towards improving quality and training phase. We proposed an algorithm known as Learning based Optimized Method for Heart Disease Prediction (LbOM-HDP). This algorithm is trained with benchmark dataset and then it takes ECG signal of a patient to predict heart disease. The experimental study made with MIT-BIH dataset showed that the proposed framework could improve heart disease prediction accuracy due to optimizations in terms of preprocessing and feature engineering. The proposed algorithm with Generalized Linear Model (GLM) showed highest accuracy 95%.

Keywords –
Heart Disease Prediction, Machine Learning, Feature Engineering, Learning Based Approach, Artificial Intelligence

1. INTRODUCTION
According to World Health Organization (WHO), there are number of heart related ailments that are causing health issues to human beings. The traditional approaches in healthcare services could provide required procedures towards treating patients. With respect to diagnosis of various diseases, of late, Artificial Intelligence (AI) is widely used. WHO recommends responsible usage of AI towards disease diagnosis in healthcare domain. There are number of machine learning models being used to solve the problems in real world applications. Machine Learning (ML) models are also widely used in heart disease prediction with clinical data which has symptoms of patients. However, usage of ECG data for analyzing heart tissues is to be given paramount importance. Towards this end there are many existing contributions found in the literature.
Li et al. [4] suggested to use PCG and ECG data to forecast cardiovascular diseases (CVDs) using a unique multi-modal approach. With an AUC of 0.936, the approach performs better than single-modal approaches. Bellfield et al. [8] explored machine learning modeling in the prediction of myocardial infarction. In the event that Signal ECGs are not accessible, Image ECGs might be utilized. Naeem et al. [9] with over 85% accuracy, a unique approach employing ANN and sensors recognizes persons based on their fragrance patterns. Ullah et al. [10] used machine learning and FCBF for feature selection, a new framework for CVD diagnosis achieves 78% accuracy with Extra Tree and Random

Forest models. From the literature, it was observed that ML models are capable of learning from ECG signals. Another important observation is that ML models lead to deteriorated performance if used directly without optimizations. Towards this end, we proposed a ML framework with optimizations towards leveraging performance in heart disease prediction. Our contributions in this paper are as follows.

1. We proposed a machine learning framework for automatic detection of heart diseases from ECG data.
2. We proposed an algorithm known as Learning based Optimized Method for Heart Disease Prediction (LbOM-HDP).
3. We developed an application to evaluate our framework and models in automatic detection of heart diseases from ECG signals.

The remainder of the paper is structured as follows. Section 2 reviews literature on prior works pertaining to heart disease prediction. Section 3 presents the proposed ML framework, its mechanisms and algorithm towards detection of heart diseases. Section 4 provides results of our experiments with ECG data. Section 5 concludes our work besides providing scope for future research.

2. RELATED WORK

The section reviews prior walks on machine learning models used for detection of heart diseases using ECG data. Bertsimas et al. [1] identified cardiac abnormalities, which are a major cause of mortality by Machine Learning on ECGs. A new approach predicts seven categories in real time, facilitating early diagnosis. Katarya and Meena [2] increased in heart disease instances is a result of hectic lifestyles. An effective way to anticipate and analyze cardiac disease is via machine learning. Alarsan and Younes [3] approached for classifying ECG data is presented, which achieves high accuracy for binary classification using Gradient-Boosted Trees and huge accuracy using Random Forests. Li et al. [4] suggested to use PCG and ECG data to forecast cardiovascular diseases (CVDs) using a unique multi-modal approach. With an AUC of 0.936, the approach performs better than single-modal approaches. Kahn et al. [5] discussed machine learning approaches that are often used to predict heart failure, including ensemble classifiers, SVM, and neural networks.

Ganesan and kumar [6] used sensors and the UCI dataset, a cloud-based and Internet of Things model is created to predict cardiac illness. In terms of accuracy and performance metrics, J48 classifier performs better than MLP, SVM, and LR. Indrakumari et al. [7] analysed method of EDA comes before statistical modeling. Analytics in healthcare helps with preventative care. Heart disease risk variables are inferred from 209 data using K-means algorithm. Bellfield et al. [8] suited for machine learning modeling in the prediction of myocardial infarction. In the event that Signal ECGs are not accessible, Image ECGs might be utilized. Naeem et al. [9] with over 85% accuracy, a unique approach employing ANN and sensors recognizes persons based on their fragrance patterns. Ullah et al. [10] used machine learning and FCBF for feature selection, a new framework for CVD diagnosis achieves 78% accuracy with Extra Tree and Random Forest models.

Qadri et al. [11] focused on utilizing machine learning to detect heart failure early. Compared to prior approaches, a revolutionary PCHF feature engineering strategy achieved accuracy. Ozcan and Peker [12] with an accuracy rate of 87%, the CART algorithm makes heart disease prediction easier to make. Oldpeak and ST Slope are important characteristics. Additional data types and big data analytics might be included in future improvements. Aggarwal and Kumar [13] investigated HRV in people with normal sinus rhythm and congestive heart failure. HRV is essential for predicting cardiac illness. To improve accuracy, feature selection techniques including filtering, wrapper, and embedding approaches are used. With 95.4% accuracy, filtering proved to be the most effective technique when using machine learning classifiers. Ahmad et al. [14] achieved great accuracy in heart disease prediction by combining SFS with LDA, RF, GBC, DT, SVM, and KNN algorithms. Rani et al. [15] provided a hybrid heart disease detection system that has an accuracy of 86.6% on the Cleveland dataset to help in early diagnosis.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, No.2, August : 2024

Sujatha and Lakshmi [16] tackled by data analytics in healthcare, which supports illness prediction and decision-making. With an accuracy rate of 83.52%, Random Forest is quite good at predicting cardiac illness. Farzana and Veeraiah [17] used 13 characteristics, such as age and blood pressure, machine learning algorithms help estimate the risk of heart disease. The framework ensures efficacy by predicting the stage of the disease and recommending preventive measures. Hagan et al. [18] concentrated on leveraging UCI and Kaggle datasets to use machine learning methods, such as SVM, MLP, and decision trees, to predict cardiovascular illness. With an accuracy ranging from 72% to 96%, SVM performs better than MLP. Decision trees are efficient because they obtain the most accuracy with the least amount of CPU time. Maini et al. [19] promised for cost-effective CVD prediction with ML algorithms. Five machine learning methods were evaluated and made accessible online using 1670 medical records. The relevance of machine learning (ML) in Indian healthcare is highlighted by this study, which also recommends more comprehensive data collecting and multicentric research to enhance preventative care. Ghosh et al. [20] although early diagnosis can reduce mortality, cardiovascular disease (CVD) remains a serious public health problem. An accuracy of huge was attained by a mixed classifier model. From the literature, it was observed that machine learning models need optimization towards leveraging prediction performance.

3. PROPOSED FRAMEWORK

We proposed a ML framework which is meant for using learning-based approaches with optimizations towards improving heart disease detection performance in ECG signals. The framework is equipped with optimizations in the form of preprocessing and feature engineering. ECG data needs preprocessing towards improving the quality of samples. From the ECG data it is possible to extract features. This phenomenon is known as feature extraction. However, all the features may not be able to contribute towards class label prediction. Towards this end, we used a feature engineering method which automatically selects features that are contributing towards disease diagnosis.

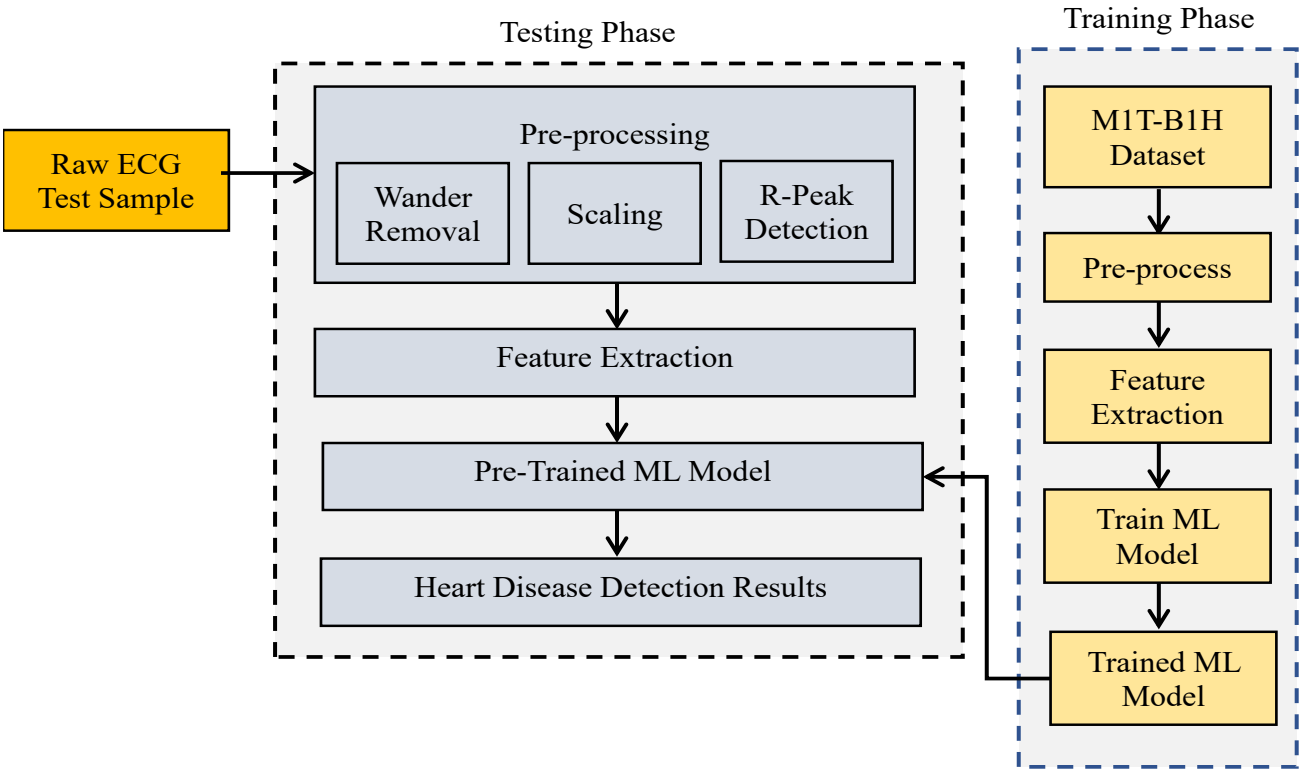

Figure 1: Proposed machine learning based framework

The preprocessing includes different mechanisms like wander removal, scaling and r-peak detection. These preprocessing mechanisms help in improving the given raw ECG sample. Afterwards, the

feature selection process exploits entropy and gain based feature selection method towards identifying features that can discern class labels. The framework has both training phase and testing phase towards supervised classification. In the training phase ML models are trained and also optimized towards detection of heart diseases. The trained models are used in the testing phase in order to analyze ECG data and provide the probability of heart disease. The framework supports multiclass classification revealing different classes of heart disease. Since we used learning based approach, metrics derived from confusion matrix, shown in Figure 2, are used for evaluation our methodology.

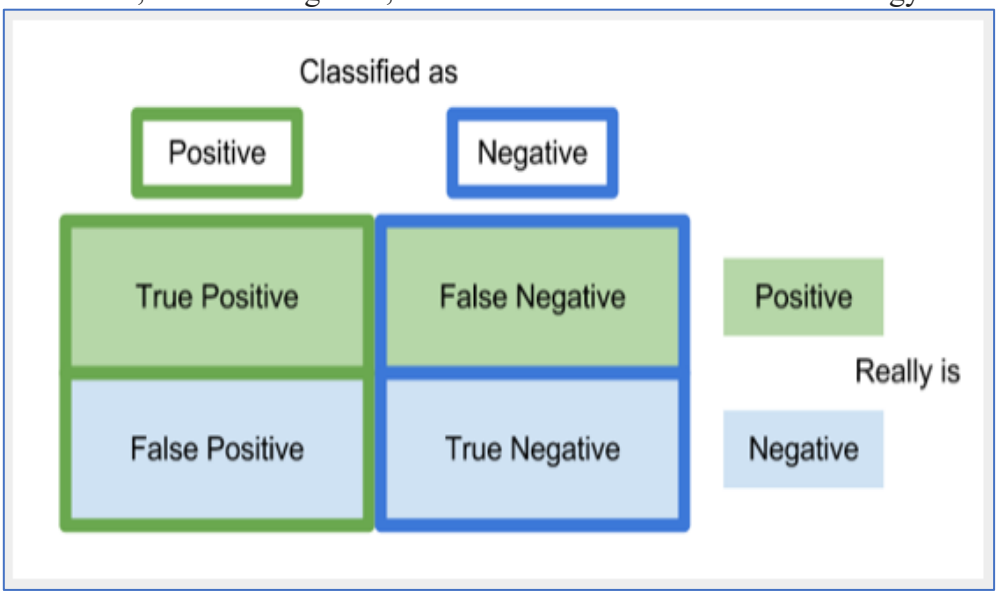

Figure 2: Confusion matrix

Based on confusion matrix, the predicted labels of our method are compared with ground truth to arrive at performance statistics. Eq. 1 to Eq. 4 express different metrics used in the performance evaluation.

$$\text{Precision (p)} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall (r)} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{F1-score} = 2 * \frac{(p*r)}{(p+r)} \quad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

The measures used for performance evaluation result in a value that lies between 0 and 1. These metrics are widely used in machine learning research. We proposed an algorithm known as Learning based Optimized Method for Heart Disease Prediction (LbOM-HDP). This algorithm is trained with benchmark dataset and then it takes ECG signal of a patient to predict heart disease.

Algorithm: Learning based Optimized Method for Heart Disease Prediction (LbOM-HDP)
Input: MIT-BIH Dataset D, models M
Output: Heart disease classification results R, performance statistics P
1. Begin
2. D'←Preprocess(D)
3. (T1, T2)←DataSplit(D')
4. F1←FeatureSelection(T1)
5. F2←FeatureSelection(T2)
6. For each model m in M
7. Compile m
8. Train m using T1, F1
9. R←Test(m, T2, F2)
10. P←Evaluate(R, ground truth)
11. Display R

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, No.2, August : 2024

12. Display P
13. End

Algorithm 1: Learning based Optimized Method for Heart Disease Prediction (LbOM-HDP)

As presented in algorithm 1, it has provision for feature selection from given training and test data towards improving performance of ML models. Besides, the algorithm has an iterative process in which each model is trained and then evaluated with the test data. The algorithm provides multiclass classification results along with performance statistics based on the functionality of models.

4. EXPERIMENTAL RESULTS

The ML framework is evaluated with ECG dataset [21]. The ML models are evaluated with and without feature engineering process. From the empirical study it was observed that featuring engineering and preprocessing could contribute to the enhancement of quality in the training process leading to improved prediction performance.

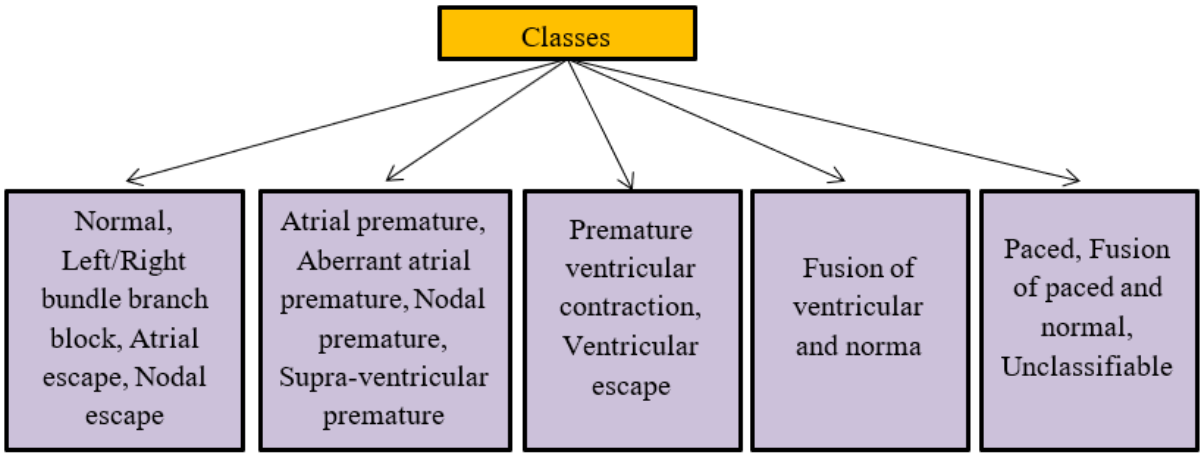

Figure 3: Shows different classes involved in the multi-class classification

As presented in Figure 3, there are different classes pertaining to heart diseases as reflected in the dataset used for empirical study.

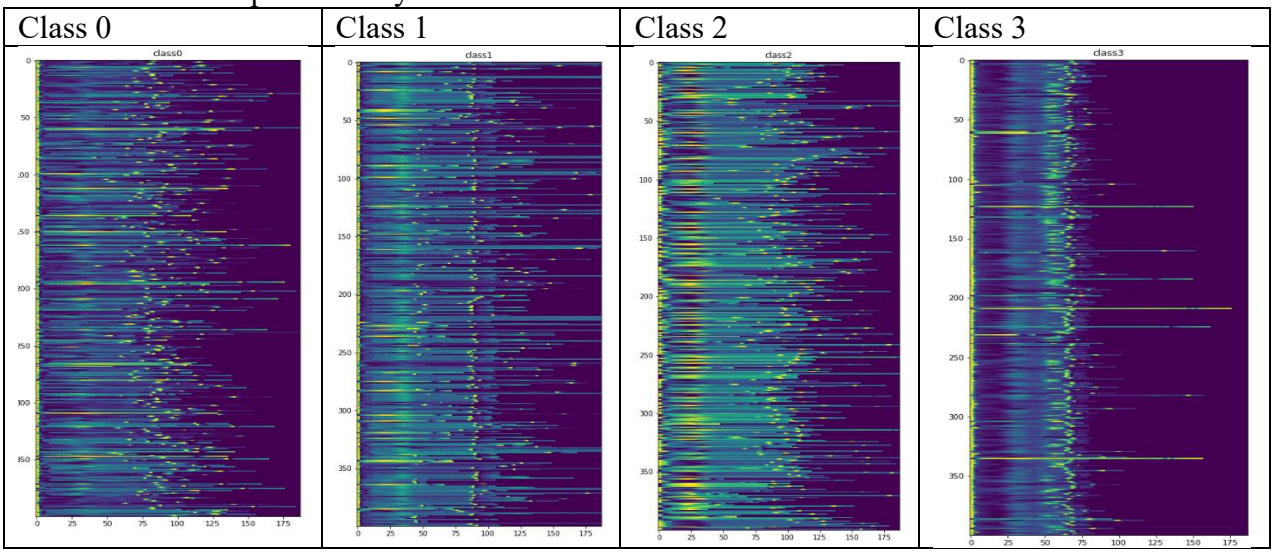

Figure 4: Shows a visualization of different classes

The histogram visualization of different classes associated with an ECG signal is provided for understanding different classes as in Figure 4.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, No.2, August : 2024

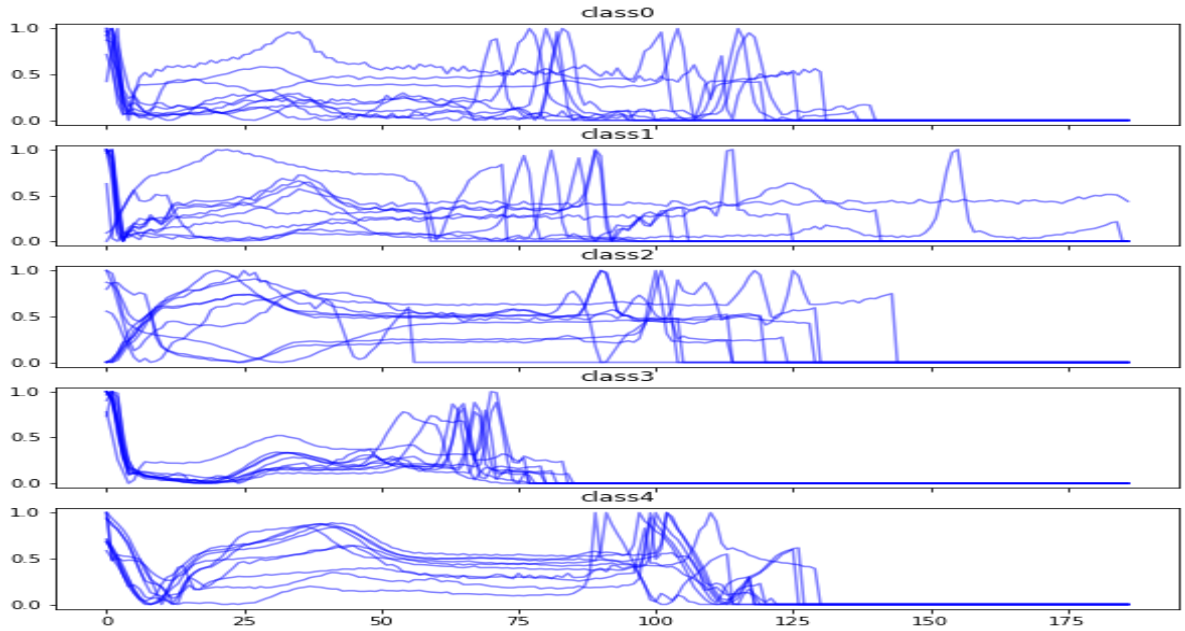

Figure 5: Data distribution dynamics in terms of the five classes

As presented in Figure 5, it is observed that there are five different classes of heart diseases for which data distribution dynamics is visualized.

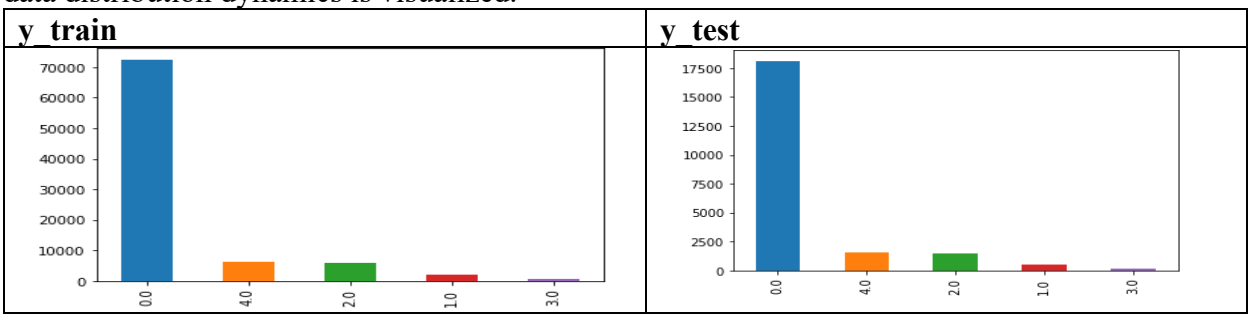

Figure 6: Data dynamics in terms of training and testing

As presented in Figure 6, it is found that training data and test data are visualized in terms of their data distribution.

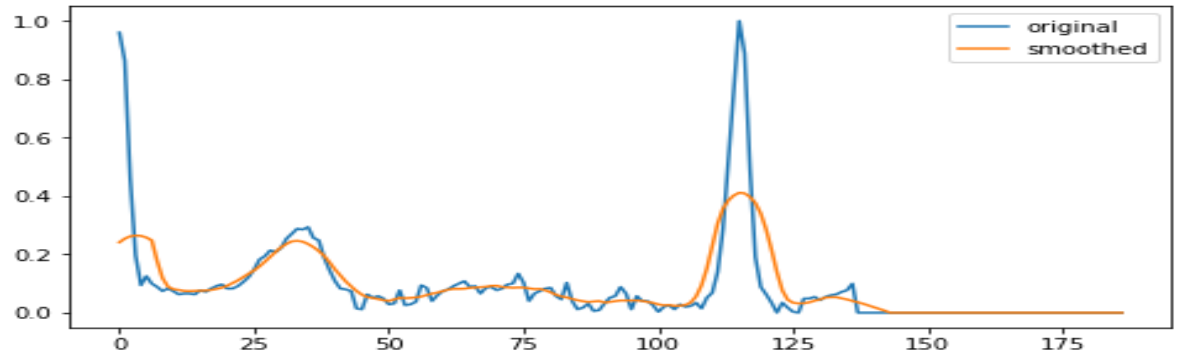

Figure 8: Shows the ECG signal comparison between original and preprocessed one

As presented in Figure 8, the original ECG signal is compared with the preprocessed ECG signal showing the difference between them.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, No.2, August : 2024

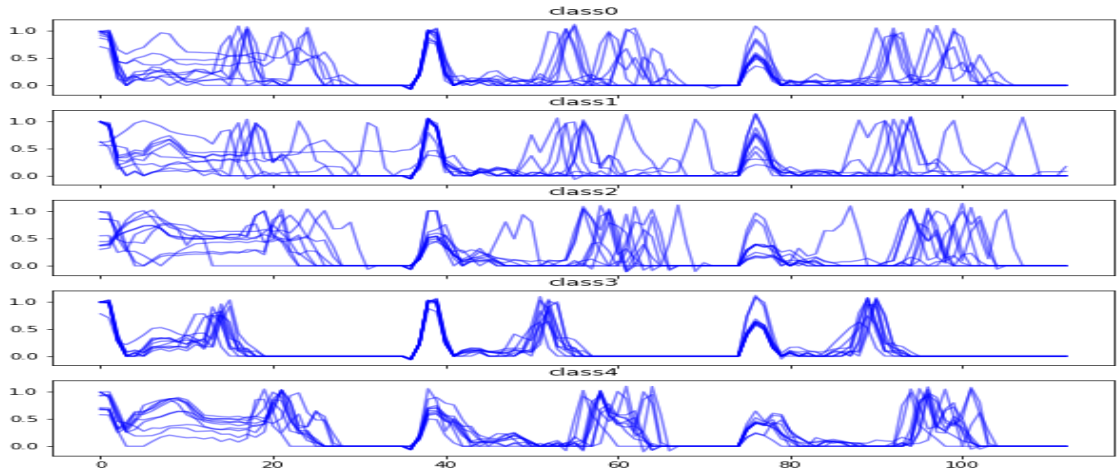

Figure 9: Shows ECG preprocessed samples of different classes

As presented in Figure 9, there are five classes of ECG samples that are preprocessed for improvement as visualized.

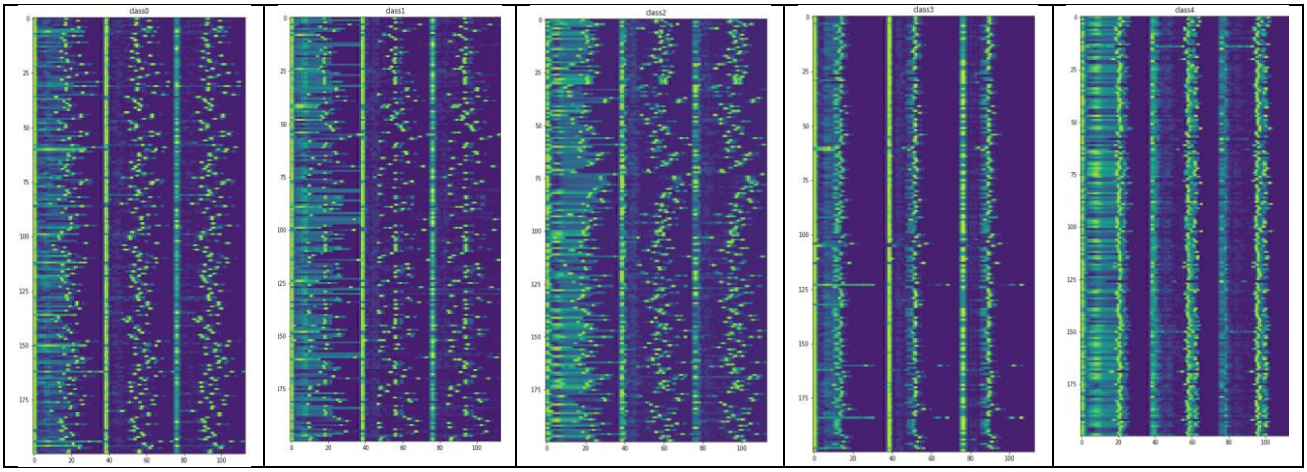

Figure 10: Visualization of different heart disease classes

As president in Figure 10, there are five heart disease classes for which ECG signals are processed and the data analytics results are visualized.

Models	Precision	Recall	F1-Score	Accuracy
LR without Optimizations	0.84	0.75	0.79	0.87
LR with Optimizations	0.93	0.88	0.9	0.89
GLM without Optimizations	0.95	0.92	0.93	0.92
GLM with Optimizations	0.98	0.94	0.96	0.95

Table 1: Results of models with and without optimizations

As presented in Table 1, it is observed that the ML models with and without optimizations reflected different level of performance.

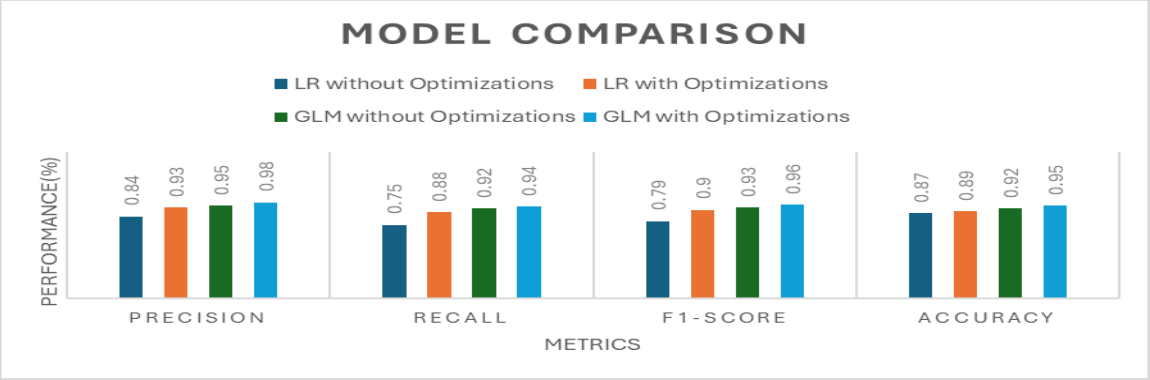

Figure 11: Performance comparison with and without optimizations

The results of experiments with and without optimizations are presented in Figure 11. The Logistic Regression (LR) model could achieve 84% precision, 75% recall, 79% F1-score and 87% accuracy. These observations are without optimizations. With optimizations the LR model could achieve 93% precision, 88% recall, 90% F1-score and 89% accuracy. Without optimizations the GLM model could achieve 95% precision, 92% recall, 93% F1-score and 92% accuracy. With optimizations GLM model could achieve 98% precision, 94% recall, 96% F1-score and 95% accuracy.

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a machine learning framework for automatic detection of heart diseases from ECG data. The framework has provision for both the training and testing phases involving optimizations with feature selection towards improving quality and training phase. We proposed an algorithm known as Learning based Optimized Method for Heart Disease Prediction (LbOM-HDP). This algorithm is trained with benchmark dataset and then it takes ECG signal of a patient to predict heart disease. The experimental study made with MIT-BIH dataset showed that the proposed framework could improve heart disease prediction accuracy due to optimizations in terms of preprocessing and feature engineering. The proposed algorithm with Generalized Linear Model (GLM) showed highest accuracy 95%. Here are possible directions for future scope of the research. The proposed framework can be extended with deep learning models, hyperparametric tuning and also Generative Adversarial Network (GAN) architecture.

REFERENCES

[1] Dimitris Bertsimas; Luca Mingardi and Bartolomeo Stellato; (2021). Machine Learning for Real-Time Heart Disease Prediction. IEEE Journal of Biomedical and Health Informatics. http://doi:10.1109/jbhi.2021.3066347

[2] Rahul Katarya and Sunit Kumar Meena. (2020). Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. *Springer*, pp.1-11. http://doi:10.1007/s12553-020-00505-7

[3] Alarsan, Fajr Ibrahem and Younes, Mamoon (2019). Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. Journal of Big Data, 6(1), 81–. http://doi:10.1186/s40537-019-0244-x

[4] Li, Pengpai; Hu, Yongmei and Liu, Zhi-Ping (2021). Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods. Biomedical Signal Processing and Control, 66, 102474–. http://doi:10.1016/j.bspc.2021.102474

[5] Khan, Younas; Qamar, Usman; Yousaf, Nazish and Khan, Aimal (2019). Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC '19 - Machine Learning Techniques for Heart Disease Datasets. 27–35. http://doi:10.1145/3318299.3318343

[6] Ganesan, M. and Sivakumar, N. (2019). IEEE International Conference on System, Computation, Automation and Networking (ICSCAN) - IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. 1–5. http://doi:10.1109/ICSCAN.2019.8878850

[7] Indrakumari, R.; Poongodi, T. and Jena, Soumya Ranjan (2020). Heart Disease Prediction using Exploratory Data Analysis. Procedia Computer Science, 173, 130–139. http://doi:10.1016/j.procs.2020.06.017
[8] Ryan A.A. Bellfield, Sandra Ortega-Martorell, Gregory Y.H. Lip, David Oxborough and Ivan Olier. (2024). Impact of ECG data format on the performance of machine learning models for the prediction of myocardial infarction. *Elsevier*. 84,pp.17-26. https://doi.org/10.1016/j.jelectrocard.2024.03.005
[9] AWAD BIN NAEEM, BISWARANJAN SENAPATI, DIPEN BHUVA, ABDELHAMID ZAIDI , ABHISHEK BHUVA, MD. SAKIUL ISLAM SUDMAN AND AYMAN E. M. AHMED. (2024). Heart Disease Detection Using Feature Extraction and Artificial Neural Networks: A Sensor-Based Approach. *IEEE*. 12, pp.37349 - 37362. http://DOI:10.1109/ACCESS.2024.3373646
[10] TAHSEEN ULLAH, SYED IRFAN ULLAH, KHALIL ULLAH, MUHAMMAD ISHAQ, AHMAD KHAN, YAZEED YASIN GHADI, AND ABDULMOHSEN ALGARNI. (2024). Machine Learning-Based Cardiovascular Disease Detection Using Optimal Feature Selection. *IEEE*. 12, pp.16431 - 16446. http://DOI:10.1109/ACCESS.2024.3359910
[11] AZAM MEHMOOD QADRI, ALI RAZA, KASHIF MUNIR AND MUBARAK S. ALMUTAIRI. (202). Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning. *IEEE*. 11, pp.56214 - 56224. http://**DOI:**10.1109/ACCESS.2023.3281484
[12] Mert Ozcan and Serhat Peker. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Elsevier*. 3, pp.1-9. https://doi.org/10.1016/j.health.2022.100130
[13] Ritu Aggarwal and Suneet Kumar. (2022). HRV based feature selection for congestive heart failure and normal sinus rhythm for meticulous presaging of heart disease using machine learning. *Elsevier*. 24p.1-15. https://doi.org/10.1016/j.measen.2022.100573
[14] GHULAB NABI AHMAD, SHAFIULLAH, ABDULLAH ALGETHAMI, HIRA FATIMA, AND SYED MD. HUMAYUN AKHTER. (2022). Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique With and Without Sequential Feature Selection. *IEEE*. 10, pp.23808 - 23828. http://DOI:10.1109/ACCESS.2022.3153047
[15] Pooja Rani; Rajneesh Kumar; Nada M. O. Sid Ahmed and Anurag Jain; (2021). A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments. http://doi:10.1007/s40860-021-00133-6
[16] P. Sujatha and K. Mahalakshmi; (2020). Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease . 2020 IEEE International Conference for Innovation in Technology (INOCON),. http://doi:10.1109/inocon50539.2020.9298354
[17] Shaik Farzana and Duggineni Veeraiah; (2020). Dynamic Heart Disease Prediction using Multi-Machine Learning Techniques . 2020 5th International Conference on Computing, Communication and Security (ICCCS). http://doi:10.1109/ICCCS49678.2020.9277165
[18] Rachael Hagan; Charles J. Gillan and Fiona Mallett; (2021). Comparison of machine learning methods for the classification of cardiovascular disease . Informatics in Medicine Unlocked. http://doi:10.1016/j.imu.2021.100606
[19] Ekta Maini; Bondu Venkateswarlu; Baljeet Maini and Dheeraj Marwaha; (2021). Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India . Medical Journal Armed Forces India. http://doi:10.1016/j.mjafi.2020.10.013
[20] Pronab Ghosh; Sami Azam; Mirjam Jonkman; Asif Karim; F. M. Javed Mehedi Shamrat; Eva Ignatious; Shahana Shultana; Abhijith Reddy Beeravolu and Friso De Boer; (2021). Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques . IEEE Access. http://doi:10.1109/access.2021.3053759
[21] MIT-BIH Dataset. Retrieved from https://www.physionet.org/content/mitdb/1.0.0/.