

HEART DISEASE PREDICTION USING DEEP LEARNING

Mansi Singh Parihar - Computer Science & Engineering Department,
MIT ADT University Pune, India

mansiparihar1999@gmail.com

Prof. Dr. Reena Gunjan - Professor, Computer Science & Engineering Department,
MIT ADT University Pune, India

reena.gunjan@mituniversity.edu.in

Abstract— In this work, a unique approach to the early and accurate identification of heart disease—a leading cause of mortality worldwide—is discussed. Recognizing the urgent need to address this health concern, the research highlights the utilization of cutting-edge medical technology as the main diagnostic tool, namely Electrocardiogram (ECG) data. In an effort to increase detection accuracy and prognostic capacity for cardiovascular illness, the study looks into a variety of machine learning (ML) and deep learning (DL) models. The suggested technique intends to improve the diagnosis of heart illness by using ECG signals, which offer important information regarding heart function. The paper's creative methodology advances the development of effective and trustworthy instruments for the early detection of cardiovascular disorders. Through the integration of cutting-edge machine learning and deep learning techniques, the research hopes to have a major influence on healthcare and, in the end, save lives by treating patients at risk of heart-related diseases in a timely manner. The potential of technology to lower the worldwide burden of cardiovascular illnesses is highlighted by this work, which marks a significant improvement in medical diagnosis.

Keywords—*Heart Disease, Cardiovascular Disease, Early Detection, Deep Learning, Machine Learning, Medical Diagnostics, Risk Prediction, Big Data, AI, Feature Extraction, Neural Networks, Predictive Models, and Healthcare Technology.*

I. INTRODUCTION

The Even though heart disease is the world's top cause of mortality, many lives may be spared if it were detected accurately and early on. Medical examinations, cardiac sounds, computed tomography (CT) pictures, electrocardiogram (ECG) signals, and other testing are the main ways to diagnose heart disease. The utilization of ECG signals in the diagnosis and treatment of heart illness is one of the most significant of these.

A large body of research and review papers on the use of several machine learning (ML) and deep learning (DL) models to the categorization of cardiac disease have recently been published. Nevertheless, it has been noted that imbalanced heart disease data has a reduced detection accuracy. Thus, the use of deep learning techniques is being investigated as a potential remedy to raise the forecast accuracy of heart disease.

With heart disease fast overtaking all other causes of death globally, reliable risk prediction is essential to patient-centered therapy. Significant risks and expensive medical

expenses are associated with invasive coronary angiography, which is frequently performed to identify coronary artery stenosis in patients with suspected coronary heart disease (CHD). Therefore, it is crucial to create accurate and non-invasive techniques for cardiac disease prediction in order to enhance patient outcomes and save healthcare costs. Deep learning in particular, which is inspired by the tens of millions of neurons and over 100,000 connections seen in the human brain, provides potential answers for this kind of application.

Advances in medical technology and the emergence of the big data age in the biomedical field—fueled by artificial intelligence—have propelled the rise of computational medicine. The enormous volumes of biological data that are available must be mined for important insights in order to further precision medicine. Machine learning methods were initially used to extract biological data and find pertinent features. This is a labor-intensive process that takes a lot of time and money because it primarily depends on feature engineering and the knowledge of experts. But traditional feature engineering is no longer necessary since deep learning, a sophisticated subfield of machine learning, can automatically extract complex and reliable features from unprocessed data.

Deep learning is said to provide major benefits for maximizing the use of biological data and enhancing medical results, especially in the areas of genetics, medication development, medical imaging, and electronic health records. As a result, deep learning is having a lot of potential applications in the medical profession and is growing in significance. Deep learning's use in computational health still faces a number of difficulties, including interpretability issues, data heterogeneity, and a lack of sufficient data. To improve deep learning's efficacy in medicine, several issues are examined and resolved. The ultimate goal of the research is to digitize the time-consuming and expensive process of medical diagnosis that is now performed in hospitals by identifying deep learning approaches that can properly anticipate heart problems.

II. LITERATURE REVIEW

1. A clinical decision support system for predicting coronary artery stenosis in patients with suspected coronary heart disease, 2022 In patients suspected of having coronary heart disease, a research offered a clinical decision support system that employs machine learning to predict coronary artery stenosis noninvasively. The goal of this technology is to lower the expenses and hazards related to intrusive angiography. The system utilizes XGBoost and Random Forest models to precisely identify individuals with varying

degrees of stenosis. It then provides tailored intervention recommendations and may reduce the need for invasive operations, ultimately leading to improved patient outcomes.

2. Classifier identification using deep learning and machine learning algorithms for the detection of valvular heart diseases, 2022 Using a dataset of 1,000 PCG recordings from different heart diseases, the research focuses on evaluating heart sounds to diagnose heart-related problems. The goal of the project is to increase classification accuracy for early heart disease screening by developing a cost-effective classifier using a modified Xception model inside a deep learning neural network.

3. Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease. (2022) The study developed deep learning algorithms using electrocardiography (ECG) to detect moderate or severe aortic stenosis (AS), aortic regurgitation (AR), and mitral regurgitation (MR), either individually or in combination. Utilizing a dataset comprising 77,163 patients, the deep learning model demonstrated high accuracy, with AU-ROC values of 0.88 for AS, 0.77 for AR, 0.83 for MR, and 0.84 for any of AS, AR, or MR. Additionally, the model exhibited strong predictive values, with a positive predictive value of 20% and a negative predictive value of 97.6%, making it effective for valvular heart disease screening.

4. FetalNet: Low-light fetal echocardiography enhancement and dense convolutional network classifier for improving heart defect prediction The work addresses the problem of effectively identifying embryonic cardiac abnormalities by ultrasonography, which is frequently hindered by low picture quality. It suggests FetalNet, a deep learning-based technique that combines a convolutional neural network (CNN) classifier with low-light image enhancement (LLIE). FetalNet highlights the potential to improve prenatal screening and diagnosis by significantly improving picture quality and demonstrating encouraging results in predicting fetal heart abnormalities.

5. A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques The survey report addresses the growing incidence of cardiovascular and peripheral vascular diseases (CVD) and highlights the critical need for cutting-edge methods such as data mining and artificial intelligence for early diagnosis and prediction. In an effort to support physicians in making decisions and lower the death rates linked to CVD, it evaluates many models used in classification, data mining, machine learning, and deep learning for CVD prediction.

6. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP The researchers used a machine learning model called Extreme Gradient Boosting (XGBoost) to forecast the 3-year all-cause death rate of patients suffering from heart failure as a result of coronary heart disease. This model effectively differentiated risk, demonstrating that a greater ML score was linked to a

considerably larger hazard of events. Predictions were influenced by important characteristics such as age, biomarkers, employment, and medication usage.

7. Heart disease detection using deep learning methods from imbalanced ECG samples The study addresses unbalanced data in the context of heart disease (HD) identification utilizing electrocardiogram (ECG) signals by putting forth a GAN-LSTM ensemble model. The GAN-LSTM model offers a promising technique for HD identification with potential applications in healthcare. Simulations using the MIT-BIH and PTB-ECG datasets show that this model achieves higher accuracy, F1-score, and AUC over previous ML and DL models.

8. Evolutionary algorithm-based convolutional neural network for predicting heart diseases In this research, a novel method for predicting heart disease utilizing the jSO algorithm to dynamically optimize CNN hyperparameters is presented: CNN-jSO. CNNjSO surpasses other models in testing on combined PhysioNet and Kaggle datasets, obtaining testing accuracy of 94.12% and training accuracy of 97.76%. With the capacity to fine-tune model architecture via jSO and the strength of CNN feature extraction, this method offers a promising approach to efficient and precise heart disease prediction.

9. Random forest swarm optimization-based for heart diseases diagnosis In order to forecast cardiac illness, the research suggests a unique method that combines Random Forest with Multi-Objective Particle Swarm Optimization (MOPSO). MOPSO is used to provide a variety of training sets, maximizing diversity and accuracy at the same time. Using this technique, Random Forest performs better by figuring out the almost ideal number of classifiers. The suggested method demonstrates improved predictive accuracy through comparisons with solo and ensemble classifiers on six heart datasets, providing a potential strategy for varied and accurate heart disease prediction.

10. Prognosis analysis of thick data: Clustering heart diseases risk groups case study. The study highlights the value of qualitative analytics in the interpretation of clinical data, especially when it comes to risk prediction for heart disease. In addition to highlighting the importance of patient attributes including lifestyle and socioeconomic status, it suggests using the Fuzzy C-Means algorithm to provide prognostic predictions for illnesses related to Cardiovascular Disease (CVD). By using Fuzzy C-Means clustering to identify risk subgroups across three UCI CVD datasets with an average accuracy of 75%, the study seeks to solve the difficulties associated with working with tiny clinical datasets.

III. TAXONOMY

I. Predictive Models for Coronary Artery Stenosis:

The paper "A clinical decision support system for predicting coronary artery stenosis in patients with suspected coronary heart disease," published in 2022, presents the groundbreaking clinical decision support system designed to predict coronary artery stenosis in patients presenting with suspected coronary heart disease. The dataset that the system uses consists of demographic information, imaging tests, and therapy measures. A sophisticated Multi-Classification prediction component serves as its main algorithm [1]. While the results are encouraging, the study points out that there is room for improvement in this innovative prediction approach as certain patient projections may not be entirely correct.

II. Detection of Valvular Heart Diseases: The 2022

article "Classifier identification using deep learning and machine learning algorithms for the detection of valvular heart diseases" shows a significant improvement in the identification of valvular heart ailments. The study's primary source of training and validation data was heart sound recordings. Using a CNN-based deep learning neural network, the research aimed to discover classifiers capable of reliably and effectively identifying valvular heart diseases [2]. Even though the study's results are optimistic, it emphasizes a learning phase limitation and suggests that it takes a considerable amount of time. This work marks a significant development in the application of state-of-the-art machine learning techniques, particularly deep learning, for the accurate and timely identification of valvular heart diseases, notwithstanding optimization issues.

III. Deep Learning for Left-Sided Valvular Heart

Disease Detection: The 2022 paper "Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease" describes a novel technique for using deep learning algorithms to identify left-sided valvular heart disease. Echocardiography data served as the study's main training and validation dataset. Using a convolutional neural network (CNN) for deep learning analysis, the research aimed to develop a model that could accurately identify symptoms of left-sided valvular heart disease from electrocardiographic data [3]. The research indicates that because of a flaw in the screening process, the results could not fairly represent clinical situations in the real world. This work

constitutes a significant advancement in deep learning techniques, with room for further validation and enhancement to increase its effectiveness.

IV. Fetal Echocardiography Enhancement for Heart

Defect Prediction: The study "FetalNet: Low-light fetal echocardiography enhancement and dense convolutional network classifier for improving heart defect prediction" released in 2022 describes a novel way to improving fetal echocardiography for heart defect detection. The study used a collection of echocardiogram pictures from GitHub, with convolutional neural networks (CNNs) as the fundamental computational framework [4]. The study's goal was to create "FetalNet," a dense convolutional network classifier, by focusing on low-light circumstances common in prenatal imaging. Despite its gains, the study identifies a restriction in the model's usefulness, citing its reliance on data quality and the unique features found in fetal echocardiography images. This study offers a significant step forward in using deep learning techniques to improve the accuracy of cardiac defect prediction.

V. Comparative Study of Cardio-Vascular Diseases

Prediction: An extensive analysis of the effectiveness of machine learning and deep learning approaches in the prediction of cardiovascular disease (CVD) can be found in the 2022 publication "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques". With random forest swarm optimization serving as the main technique, the study's analysis was based on a wide variety of six cardiac datasets [5]. Although the research offers valuable insights into the efficacy of different approaches, it acknowledges a limitation in its generalizability, namely that the results are derived from a limited number of datasets. The study also highlights how ambiguous the model is when used in actual clinical situations. This study provides a helpful machine learning comparison.

VI. Predicting 3-Year All-Cause Mortality in Heart

Failure Patients: A study aiming at predicting 3-year all-cause mortality in patients with heart failure related to coronary heart disease is described in the 2021 paper "Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP". The study analyzed patient data from case reports using a range of machine learning techniques, such as LR, KNN, SVM, NB, and MLP [6]. A major limitation of the study is that it only included patients from certain hospitals and geographic areas. This highlights the need

for more data to be included in the study in order to increase the generalizability of the prediction model. This work demonstrates the potential of machine learning techniques in clinical settings and represents a significant advancement towards interpretable death prediction in heart failure patients.

VII. Heart Disease Detection using Deep Learning Methods: An investigation into the application of deep learning algorithms for the diagnosis of heart illness using unequal electrocardiogram (ECG) data was published in the paper "Heart disease detection using deep learning methods from imbalanced ECG samples" in 2021. The study employed many deep learning methods, including MLP, LSTM, GAN, and GAN-LSTM ensemble models, using ECG datasets from patients with heart disease [7]. The study acknowledges a limit in the possibility for further increase in performance accuracy, despite the encouraging results. The effectiveness of deep learning techniques in identifying heart conditions from ECG recordings is demonstrated in this work, opening the door for further advancements in the precision and dependability of diagnostic instruments.

VIII. Evolutionary Algorithm-Based CNN for Heart Disease Prediction: The 2021 paper "Evolutionary algorithm-based convolutional neural network for predicting heart diseases" outlines a unique method for heart disease prediction that makes use of an enhanced convolutional neural network (CNN) for evolutionary algorithms. In order to create a more precise prediction model, the study employed datasets from Kaggle's heartbeat sounds and PhysioNet heart sounds [8]. The study admits that in order to improve the accuracy of the model, greater process speeds and testing on several datasets are necessary. This work emphasizes how evolutionary algorithms and CNNs may be used to improve the field of cardiac illness prediction, but it also emphasizes how further optimizations and validations are required before any real-world clinical applications can be made.

IX. Random Forest Swarm Optimization for Heart Diseases Diagnosis: The random forest method and multi-objective particle swarm optimization (MOPSO) are combined in a unique way to diagnose heart illness, as described in the 2021 paper "Random forest swarm optimization-based for heart diseases diagnosis". The study aimed to improve the identification accuracy of heart illness using six distinct cardiac datasets [9]. The research did find a limit on the amount of production time required to train decision trees in this particular design. While

highlighting the significance of adjusting training periods for realistic clinical deployment, this study demonstrates how combining optimization approaches with random forest algorithms may increase diagnosis accuracy in cardiac diseases.

X. Prognosis Analysis of Heart Diseases Risk Groups: A specific investigation into clustering heart disease risk groups using the Fuzzy C-Means (FCM) approach is described in the 2021 paper "Prognosis analysis of thick data: Clustering heart diseases risk groups case study". One tiny dataset, the Cleveland Dataset, was examined in the study [10]. The study's shortcoming, though, was the tiny dataset, which could have limited how far the findings could be applied. Moreover, the research failed to include extraneous variables that can impact the prognosis of heart disease. Notwithstanding these drawbacks, the study offers helpful information on the clustering of heart disease risk groups, indicating the appropriateness of the FCM method for this kind of analysis. Our understanding and application may be enhanced by greater study using bigger and more varied datasets and taking into account outside factors.

IV. PROPOSED ARCHITECTURE

The study's objectives are to develop a system that can determine whether a patient has heart disease and how severe it is, all while studying how computers can comprehend human heart disease and its symptoms and developing machine learning algorithms that become better with practice. The Python language will be used in this endeavor. The research will be conducted on Google Colab, an online collaborative neural network work environment. The first step is to collect a large amount of data on heart disease and related characteristics, such as age, years of smoking, kind of chest pain, LDL cholesterol, height, and family history. For training and evaluating our computer model, these parameters are essential.

Moreover, using front-end website construction, we may effortlessly share the content we've discovered on our homepage. Following the model's training, an independent set of test messages will be used to assess the model's performance. We wish to test its ability to accurately detect hereditary disorders. To assess the quality of our model, we will also contrast it with alternative approaches. Our strategy focuses on using both machine learning and deep learning algorithms to identify and diagnose cardiac disease. Actually, these techniques work well for automating the entire process of patient monitoring and cardiac disease identification.

A. Dataset

This dataset has been prepared from a private hospital in Middle East. The data of 22 attributes have been collected manually from each patient who has been seen by the

cardiologists. The name attribute has been removed after the stored in an excel sheet correctly.

B. Material/Tools Required

- PYTHON
- DEEP LEARNING
- MACHINE LEARNING
- DATA ANALYTICS
- NUMPY, PANDAS, KERAS
- GOOGLE COLAB

V. ALGORITHM

Machine Learning Implementation

Decision Tree Classifier:

Accuracy: 74%

Explanation: This model makes decisions by splitting the data into branches based on features. It asks yes/no questions about the data until it reaches a conclusion. The "leaves" of the tree represent the final outcomes or predictions.

Random Forest Classifier:

Accuracy: 72%

Explanation: This model consists of a collection of many decision trees. Each tree in the forest is trained using a different subset of the data, resulting in an impressive sight. In order to arrive at a final forecast, it compiles the data from each tree.

AdaBoost Classifier:

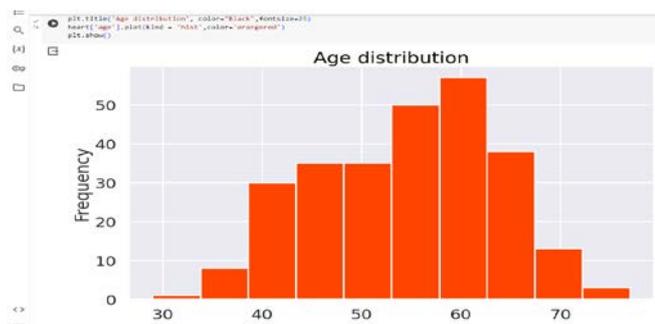
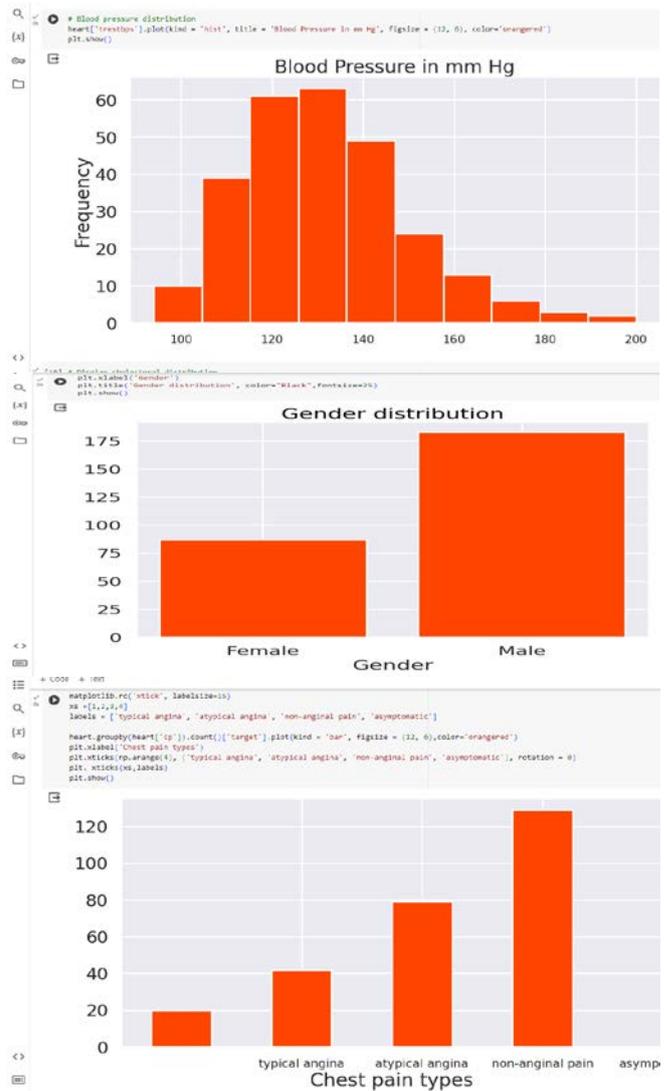
Accuracy: 74%

Explanation: AdaBoost is an acronym for Adaptive Boosting. It creates a powerful learner by combining several weak learners, which are often basic models. It trains each model in turn, assigning greater weight to the data points that the preceding model mispredicted.

Gradient Boosting Classifier:

Accuracy: 74%

Explanation: Gradient Boosting combines numerous weak learners in a manner akin to AdaBoost. It accomplishes this, albeit, in a rather different manner. It constructs trees one after the other, fixing mistakes in each one as it goes. The total of all the trees' forecasts makes up the final forecast.



Neural

Network

Overview

Their design and operation are modeled by the actual neural networks found in the brain. Neural Network Structure Neurons, which are linked nodes arranged in layers, make up neural networks. An input layer, one or more hidden layers, and an output layer are the standard layers. Neural Network Training Neural networks modify the weights and biases while they are being trained. One important approach for computing gradients and updating weights and biases is backpropagation. Deep Convolutional Networks Multiple hidden layer neural networks are referred to as deep learning networks. Applications are found in many different fields, including speech recognition, natural language processing, and picture recognition. Important Industries: banking, medical, etc.

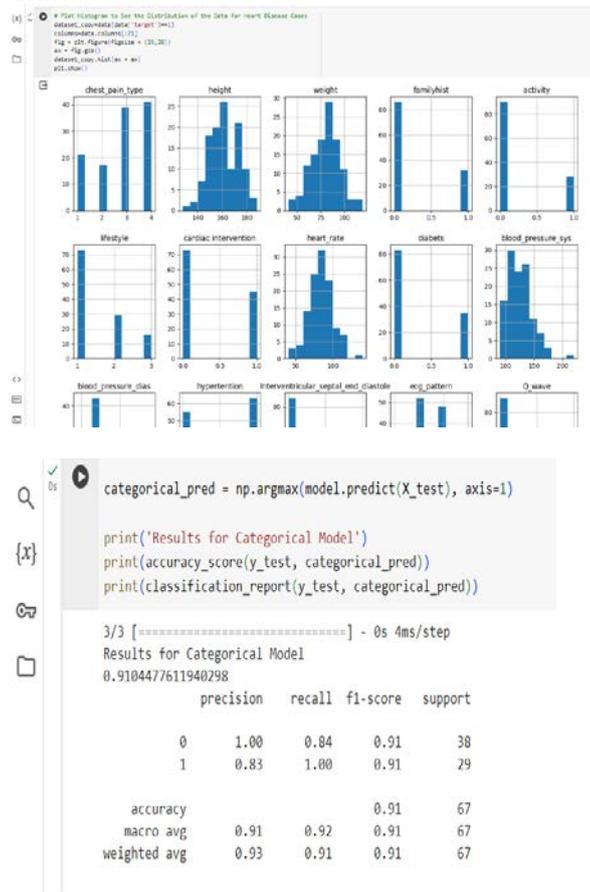
Application of Neural Networks

Here, a sequential neural network model with several dense layers is being deployed. The ReLU activation function is used by each Dense layer, which employs a fixed number of units (neurons) for the hidden layers. In order to avoid

overfitting, dropout layers are also included in the model.

1. There is a 20% dropout after the first layer, which has 64 units.
2. There are 32 units in the second layer, and then another 20% dropout.
3. There are 16 units in the third stratum, and the dropout rate is 20%.
4. Eight units with a softmax activation for categorization make up the fourth layer.
5. A dense layer with two units and a sigmoid activation for binary classification makes up the last layer.

Utilizing categorical cross-entropy loss, the model is trained.

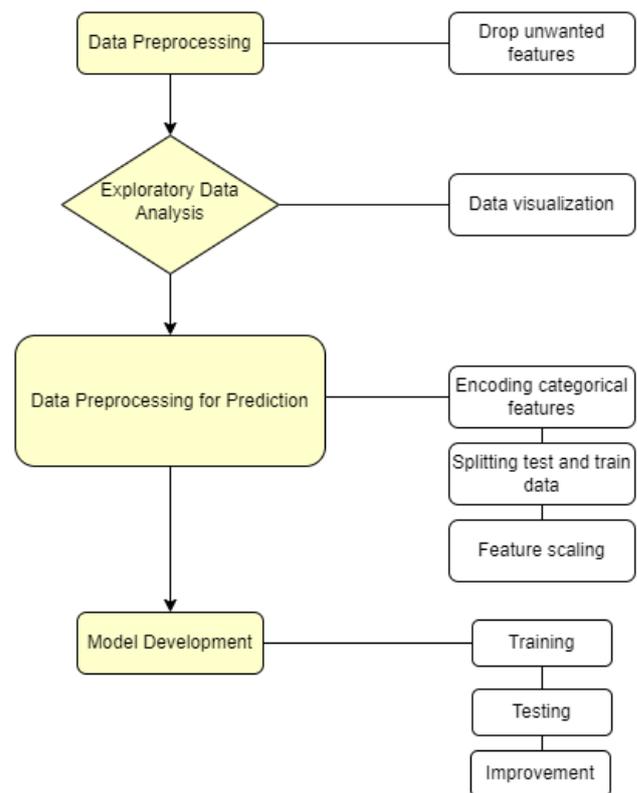


VI. CONCLUSION

This study paper highlights the need of correctly and rapidly diagnosing heart disease, as it is a leading cause of mortality globally. In order to further existing efforts to improve diagnostic precision, the study combines machine learning and deep learning techniques, with a focus on the challenges caused by unbalanced data in electrocardiogram (ECG) signals. The literature review highlights the growing body of research that classifies cardiac disease using several models, which encourages more study into the potential of deep learning to overcome the accuracy limitations caused by unbalanced datasets. The study demonstrates how deep learning, inspired by the complexity of the human brain, may automate feature extraction with minimal feature engineering work.

Big data and artificial intelligence, the factors underpinning the creation of computational medicine, are predicted to aid in the advancement of precision medicine. Despite the substantial promise that deep learning has for medical applications, the study acknowledges a number of challenges, such as data heterogeneity, interpretability problems, and insufficient data. The primary goal is to discover trustworthy deep learning methods to predict cardiac diseases. This will eventually assist hospitals in decreasing the amount of work associated with their labor-intensive diagnostic processes and further the discussion about how cutting-edge technology enhances healthcare.

A. Figures and Tables



REFERENCES

- [1] Cai L, Zhao E, Niu H, Liu Y, Zhang T, Liu D, Zhang Z, Li J, Qiao P, Lv H, Ren P, Zheng W and Wang Z. (2023). A machine learning approach to predict cerebral perfusion status based on internal carotid artery blood flow. *Computers in Biology and Medicine*. 164:C. Online publication date: 1-Sep-2023.
- [2] Roy, Tanmay & Roy, Dr. Joyanta & Mandal, Nirupama. (2022). Classifier identification using Deep Learning and Machine Learning Algorithms for the detection of Valvular Heart diseases. *Biomedical Engineering Advances*. 3. 100035. 10.1016/j.bea.2022.100035.

- [3] Elias P, Poterucha TJ, Rajaram V, Moller LM, Rodriguez V, Bhave S, Hahn RT, Tison G, Abreau SA, Barrios J, Torres JN, Hughes JW, Perez MV, Finer J, Kodali S, Khalique O, Hamid N, Schwartz A, Homma S, Kumaraiah D, Cohen DJ, Maurer MS, Einstein AJ, Nazif T, Leon MB, Perotte AJ. Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease. *J Am Coll Cardiol*. 2022 Aug 9;80(6):613-626. doi: 10.1016/j.jacc.2022.05.029. PMID: 35926935.
- [4] Sutarno Sutarno, Siti Nurmaini, Radiyati Umi Partan, Ade Iriani Sapitri, Bambang Tutuko, Muhammad Naufal Rachmatullah, Annisa Darmawahyuni, Firdaus Firdaus, Nuswil Bernolian, Deny Sulistiyo, "FetalNet: Low-light fetal echocardiography enhancement and dense convolutional network classifier for improving heart defect prediction, " ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2022.101136>.
- [5] M. Swathy, K. Saruladha, A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques, ISSN 2405-9595, <https://doi.org/10.1016/j.ict.2021.08.021>.
- [6] Ke Wang, Jing Tian, Chu Zheng, Hong Yang, Jia Ren, Yanling Liu, Qinghua Han, Yanbo Zhang, Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP, *Computers in Biology and Medicine*, ISSN 0010-4825, <https://doi.org/10.1016/j.combiomed.2021.104813>.
- [7] Adyasha Rath, Debahuti Mishra, Ganapati Panda, Suresh Chandra Satapathy, Heart disease detection using deep learning methods from imbalanced ECG samples, *Biomedical Signal Processing and Control*, Volume 68, 2021, 102820, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2021.102820>.
- [8] Ali A. Samir, Abdullah R. Rashwan, Karam M. Sallam, Ripon K. Chakraborty, Michael J. Ryan, Amr A. Abohany, Evolutionary algorithm-based convolutional neural network for predicting heart diseases, *Computers & Industrial Engineering*, Volume 161, 2021, 107651, ISSN 0360-8352, <https://doi.org/10.1016/j.cie.2021.107651>.
- [9] Asadi S, Roshan S, Kattan MW. Random forest swarm optimization-based for heart diseases diagnosis. *J Biomed Inform*. 2021 Mar;115:103690. doi: 10.1016/j.jbi.2021.103690.
- [10] J. Fiaidhi, S. Mohammed, Prognosis analysis of thick data: Clustering heart diseases risk groups case study, *Computers & Electrical Engineering*, Volume 92, 2021, 107187, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2021.107187>.