Enhancing Heart Disease Diagnosis Using Deep Learning and ML Algorithms

Mrs.Seethalakshmi[1], Ms.Pavithra[2], Mrs.Agnes[3], Ms.Varshini[4]
[1]*M.E, Department of CSE, TJS Engineering College, Peruvoyal*
[2,3,4]*Assistant professor, TJS Engineering College, Peruvoyal*

Abstract - One of the primary causes of mortality worldwide is heart disease, which emphasizes the importance of early and precise diagnostic techniques for better patient outcomes. The present research investigates the use of different deep learning and machine learning algorithms for the prediction of heart disease. Deep learning models, such as convolutional neural networks like ResNet and VGG19, are compared with traditional machine learning methods like the XGBoost support vector machine (SVM), random forest (RF), and linear regression. Using cross-validation approaches, each algorithm is painstakingly trained and refined to maximize performance parameters like precision, recall, precision, and F1-score. In order to improve prediction capabilities, ensemble approaches are investigated as a means of utilizing the advantages of numerous models. The findings show that deep learning models—in particular, CNNs—perform better than conventional machine learning algorithms at identifying intricate relationships in the data, leading to increased sensitivity and accuracy in the prediction of heart disease. The research yields valuable insights that enhance diagnostic approaches in cardiovascular health. These insights provide healthcare practitioners with strong instruments to perform early risk assessment and customize patient care plans. The results highlight how machine learning and deep learning approaches have the potential to transform cardiovascular healthcare by giving physicians dependable decision-support tools. In order to guarantee generalizability and usability in clinical settings, the next research topics include validation across a variety of patient populations and additional investigation of interpretability techniques for deep learning models.

Keywords: ***VGG19, Deep Learning, Convolutional Neural Network (CNN), Medical Imaging, Image Preprocessing***.

I. INTRODUCTION

Heart disease continues to be a major global public health concern, contributing to high rates of morbidity and mortality. For prompt intervention and better patient outcomes, heart disease must be identified early and accurately predicted. Recent developments in deep learning (DL) and machine learning (ML) have demonstrated encouraging outcomes in a range of healthcare applications, such as prognosis and disease diagnostics. Large-scale datasets, including a variety of patient data, may be analyzed by these algorithms, giving medical professionals the ability to tailor treatment plans and make data-driven decisions. The primary objective of this research is to create reliable cardiac disease prediction models by utilizing ML and DL approaches.

This research searches for risk factors and predicted patterns related to heart disease using extensive datasets that include clinical measures, medical history, and demographic information. Conventional machine learning algorithms like SVM, XGBoost, Random Forests, and Linear Regression provide well-proven approaches for predictive modelling, whereas deep learning models such as CNNs (e.g., ResNet, VGG19) are excellent at identifying complex patterns in data by means of hierarchical feature learning. By incorporating these approaches into medical procedures, it is possible to improve the precision of diagnoses, assist in the early identification of heart disease risks, and maximize the use of resources in healthcare systems.

This work compares and contrasts several machine learning (ML) and deep learning (DL) techniques, emphasizing the advantages and disadvantages of each in the prediction of heart disease. By improving models interpretability, scalability, and practical use in clinical settings, the findings further the field of analytics for prediction in cardiovascular health and establish a foundation for future investigations. This work aims to equip medical professionals with advanced tools to lessen the impact of heart disease and enhance patient outcomes by investigating the relationship between ML/DL techniques and cardiovascular health.

II. LITERATURE REVIEW

et.al Roy, J. K., Roy, T. S., and Mandal N. Worldwide, the prevalence of heart-related illnesses is rising quickly. Early diagnosis and detection are greatly aided by artificial intelligence using computer techniques. The goal of this research has been to identify the optimal classifiers for various valvular heart issues utilizing machine learning algorithms built in Python 3.8 using well-known CNN-based deep learning models. This work develops a CNN-based Xception as a network system for rheumatic cardiac sound evaluation for the first time. The attributes in the time domain are root mean square, energy, power, zero crossing speed, skewness, kurtosis, and total harmonic distortion. The research was done using the standard heartbeat sound data repository, which contains heart sounds of both healthy and sick people. In the end, a comparison of all the assessed outcomes revealed that the Random Forest and SVM methods are the most successful machine learning techniques. The most effective deep learning technique is the suggested modified CNN-based Xception model.

et.al Yaganteeswarudu, A. A large number of machine learning models now in use for medical analysis focus on a single disease at a time. For example, separate analyses could be conducted for diabetes, cancer, or skin conditions. The single analysis cannot forecast beyond one disease under any common framework. This article suggests a Flask API-based system for multiple disease prediction. This article examines the analysis of diabetes, diabetic retinopathy, heart disease, and breast cancer. Later on, additional illnesses could be included, such as skin conditions, fever analysis, and numerous other illnesses. Many sickness analyses were implemented using the Flask API, Tensor Flow, and machine learning approaches. Python pickling is used to save the model behavior, while Python unpickling is used to load the pickle file whenever it is needed. The significance of this article's study lies in its inclusion of all the factors that contribute to the disease's development, making it feasible to identify the disease's full range of impacts. The Flask API has been created. The disease's name and parameters must be sent by the user in order to access this API. The Flask API will call the relevant model and provide the patient's status back. The significance of this research lies in its examination of the most common diseases, which enables continuous patient monitoring and early patient warning to reduce the death rate.

et.al Patel, S., Shah, D., and Bharti, S. K. Heart disease, sometimes called cardiovascular disease, has been leading the cause of death globally over the last several decades. It links a number of heart disease risk factors and emphasizes the urgency of finding timely, accurate, and reasonable ways to diagnose the condition early on and begin treating it. One popular method for handling massive amounts of information in the healthcare industry is data mining. Researchers analyze vast amounts of intricate medical data using a variety of data extraction and machine learning techniques, assisting medical personnel in the prediction of cardiac disease. It makes use of the current dataset from the UCI heart disease patient repository's Cleveland database. There are 76 attributes and 303 instances in the collection. Just 14 out of the 76 attributes are taken into account during testing, which is crucial to proving the effectiveness of various algorithms. The purpose of this research is to estimate the patients' risk of acquiring heart disease. The findings show that the K-nearest neighbor yields the highest accuracy score.

et.al Meena, S. K., and Katarya, R. These days, many are too preoccupied with going about their daily lives, working, and other commitments to pay attention to their health. People's health is becoming more neglected as a result of their busy lives and lack of awareness of it. In addition, the majority of people have a condition like heart disease. According to data provided by the International Health Organization (WHO), heart-related diseases account for over 31% of all fatalities worldwide. Thus, the ability to forecast whether cardiac disease will develop or not becomes crucial for the medical community. However, the volume of data that hospitals and the medical industry receive might make analysis challenging at times. For medical professionals, using machine learning algorithms for data handling and prediction can become highly efficient. Therefore, we have covered cardiovascular disease, including its risk factors as well as machine learning approaches, in this study. We forecast heart disease using these machine learning techniques, and we have given a comparative study of the machine learning algorithms employed in the prediction experiment. This research's goal or objective is entirely focused on to predict cardiac disease.

et.al Kavitha, M., Sai, Y. R., Dinesh, R., Gnaneswar, G., & Suraj, R. S. Heart disease has become a serious health concern for many individuals and is responsible for a notable death rate worldwide. Many lives can be saved by early detection of heart illness; nevertheless, routine clinical data analysis poses a significant difficulty in the detection of cardiovascular disorders such as heart attacks and coronary artery diseases. Machine learning (ML) has the potential to improve decision-making and forecast accuracy. When it comes to applying machine learning techniques, the medical field is leading the way. The proposed work proposes a novel machine learning technique for the prediction of cardiac disease. The suggested study used the Ohio cardiovascular disease dataset and performed data mining techniques, including regression and classification. Machine learning techniques There is use of random forests and decision trees. A novel approach to the machine training model is developed. The solution uses three artificial intelligence algorithms: 1. decision tree, 2. random forest (RF), and 3. hybrid framework, which is a combination of decision tree and random forest. To forecast cardiac illness, the interface design employed a hybrid model that included a decision tree with a random forest to determine the user's output parameter.

III. RESEARCH METHODOLOGY

A variety of algorithms for deep learning and machine learning are used in the study technique to create and assess heart disease forecasting models through a number of crucial processes. First, the dataset is generated and gathered that includes clinical measures, medical history, and demographic data. This entails encoding categorical variables, normalizing numerical features, and performing stringent data cleaning procedures to address missing values and outliers. The strongest predictive variables are then found using feature selection approaches, which may make use of correlation evaluation and feature significance measurements using models based on trees.

A.DATA MANIPULATION

This section of the document loads the data, verifies its cleanliness, trims it, and outputs a dataset that is simple to study. Verify the file's procedures and give an explanation for any cleanup choices made.

B.DATA GATHERING

The information about the atmosphere is gathered from the UCI ML preserve, which provides access to multiple datasets for machine learning applications covering nearly every domain, such as intrusion detection and illness prediction.

C.PRIOR OF PREPROCESSING

It's possible that the data we obtain from various sources contains repeated data, missing numbers, and inconsistent data. To get precise prediction results, the data set need to be cleansed. Missing values need to be filled up using mean values or another approach, or they can be deleted. Additionally, redundant data needs to be purged to prevent bias in the outcomes. To achieve good prediction accuracy, extreme or outlier values in some datasets may also need to be eliminated.

IV.SYSTEMARCHITECTURE

This architecture incorporates a Flask-based web application with the VGG19, a convolutional neural network, or CNN, model for cardiovascular disease prediction. Preprocessing medical pictures is required, and the trained model is used to provide real-time diagnostic results via an intuitive user interface. This configuration guarantees effective clinical decision-making and improves the delivery of healthcare by offering readily available and precise forecasts derived directly from medical data input.

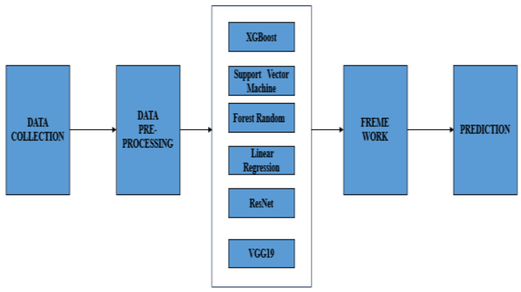


Fig1: System architecture

Benefits: Using the VGG19 CNN model facilitates highly accurate and efficient heart disease prediction, enabling real-time diagnostics and improving healthcare delivery through its integration into web applications.

V. MODULES

A. Data collection

B. Data Preprocessing

C. Model Implementation

D. Framework

E. Prediction

A.DATA COLLECTION

The processes of gathering data for this research include obtaining a variety of ECG (electrocardiogram) datasets from reputable sources or by working with healthcare organizations. These datasets will include enough ECG recordings of both healthy and pathological cardiovascular states to provide thorough coverage of all possible diagnostic scenarios. Verifying the collected data's quality and accuracy is essential to make sure it is suitable for training the prediction model.

B.DATA PREPROCESSING

Cleaning acquired ECG data to eliminate noise and artifacts is known as data preparation, and it guarantees that only pertinent signals are left for analysis. After that, the ECG records are converted into standardized formats—such as pictures or numerical representations—that are appropriate for deep learning models. In order to preserve uniform feature scales and improve model performance, data normalization is used. To enable efficient model training and evaluation, the preprocessed data is then divided into testing, validation, and training sets. Coronary artery disease patients are identified by annotations made by medical professionals. The model's accuracy in predicting cardiac conditions is ensured by the existence or absence of critical labeled data, which is necessary for supervised learning.

C. MODEL IMPLEMENTATION

In this work, we combined deep learning and machine learning models to create a reliable cardiac disease prediction system. To create a reliable prediction baseline, we used machine learning (ML) techniques such as XGBoost, supported vector machines (SVM), random forests (RF), and linear regression. In addition, we used convolutional neural network, or CNN, architectures, namely ResNet and VGG19, to use the capability of deep learning in order to identify complex patterns in ECG data. Prominent for its ability to the extract deep features, the VGG-19 model performed better in terms of heart disease prediction. Real-time prediction and user-friendly interactions were made possible by seamless integration of the whole model suite into a Flask web application.

During training, important metrics, including precision, recall, accuracy, and the F1-score, are continuously assessed to make sure the model can effectively learn from the input and generalize. The trained model is then put through a rigorous evaluation process on the validation dataset in order to see how accurate and reliable it is at predicting cardiovascular disease in a variety of patient profiles. Validating the model's dependability and suitability for use to clinical settings where prompt and accurate diagnosis can have a substantial impact on patient care and outcomes is the primary goal of this review phase.

D. FRAMEWORK

Building a safe, adaptable, and user-friendly platform is the main goal when developing the Flask web application for delivering the trained model. Creating an easy-to-use interface that allows users to quickly enter patient data or upload ECG recordings is part of this process. Real-time heart disease prediction was made possible through the integration of the trained model with the Flask application. This allows for instant diagnostic insights, depending on user input. The Flask application can be deployed on a stable web server to guarantee internet connectivity and its ability to process several requests at once. This makes the application responsive and dependable for the delivery of healthcare.

E. PREDICTION

In the prediction phase for cardiovascular disease using ECG images, segmentation techniques are applied to isolate key features such as the QRS complex and ST segment. This involves employing methods like thresholding and edge detection to enhance signal clarity and extract relevant regions from the ECG recordings. Following segmentation, discriminative features are extracted using techniques such as wavelet transforms, Fourier analysis, and time-domain features. These methods capture essential characteristics that differentiate between normal and abnormal cardiac conditions, facilitating accurate diagnosis. After that, ECG images are classified into diagnostic groupings using classification algorithms such as support vector machines (SVM), random forests, or deep neural networks. Model performance is optimized by hyperparameter tweaking, and the classifier's capacity to predict cardiovascular illnesses from the retrieved features is evaluated using measures like accuracy and F1-score. Effective healthcare decision-making is supported by this comprehensive strategy, which guarantees accurate and trustworthy diagnostic forecasts.

VI.ALGORITHM

A. XGBoost

An enhanced version of the gradient booster framework is called XGBoost (Extreme Gradient Boosting). By maximizing the predictive model's speed and accuracy, it improves model performance. XGBoost employs an ensemble tree-based technique in which a number of decision tree models are trained one after the other, correcting each other's mistakes. Regularization techniques are integrated to avoid overfitting, and parallel processing is supported to expedite computation. With its built-in techniques to accommodate missing values and its great efficiency in processing huge datasets, XGBoost is a reliable option for predictive modeling.

B. Support Vector Machine

Support Vector Machine (SVM) is a type of supervised learning applied to regression and classification issues. SVM uses the best-fitting hyperplane to classify data into groups. It uses a kernel approach to transform the given data into an area with additional dimensions wherein linear division is possible. Optimizing the margin, also known as support vectors, between the closest points of different classes is the primary goal of support vector machines (SVMs). SVM is versatile and efficient in high-dimensional domains because it may use a range of kernel functions, such as highly directional basis functions (RBF), linear, and polynomial.

C. Forest Random

Forest Random (RF)With the use of many decision trees built during training, a random forest is a method of ensemble learning that produces the mean prediction (regression) or mode of the categories (classification) for each individual tree. All the tree is trained by using a distinct collection of features and data that is chosen at random via bootstrapping. This unpredictability lessens overfitting and enhances model generalization. Robust against noise and anomalies, Random Forest can manage extensive datasets with elevated dimensionality and yield feature importance estimations, rendering it a potent and comprehensible model.

D. Linear Regression

A fundamental statistical technique for predicting a dependent variable that is continuous based on a number of independent variables is called linear regression. It is predicated on the input features, with the target variable having a linear relationship. By reducing the sum of squared discrepancies in observed and predicted values, the model seeks to choose the line that best fits the data. Ordinary least-squares (OLS) is one method used to do this. Implementing linear regression is easy, as it provides a direct interpretation of coefficients and acts as a foundation for more intricate models. It can be dependent on outliers and is only capable of capturing linear correlations.

E. Convolutional neural networks (CNNs)

Long short-term memory (LSTM) networks are unique in that they can manage dependencies that arise in sequential data by employing memory cells with user input, ignore, and output gates. By controlling information flow, these gates help classic RNNs overcome the vanishing gradient issue. Key factors in maximizing model performance are dropout rates, learning rate, and LSTM unit count. By adjusting these settings, you may make sure that the LSTM network responds to the features of datasets related to cardiovascular illness. To improve predicted accuracy, backpropagation is used to modify weights, sequence data is processed, and cell memories are updated during training. Several validation strategies are used to reduce overfitting and enhance generalization skills, such as early termination based on performance indicators from an independent validation set.

F. ResNet

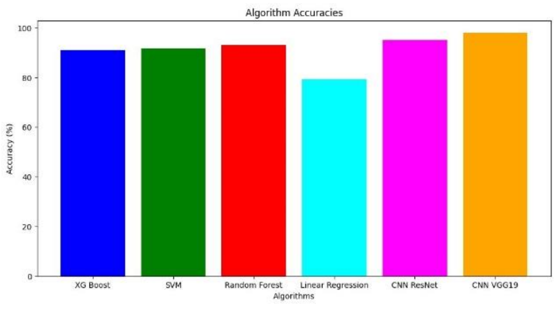
The vanishing gradient problem in very deep networks was addressed by developing a ResNetA deep neural network model known as ResNet (Residual Networks). Gradients can now traverse the network directly thanks to the addition of shortcut connections that avoid one or more tiers. The training of incredibly deep networks containing hundreds or thousands of layers is made possible by this architecture. Because of its architecture, ResNet can learn complicated features efficiently and keep up its performance as network depth grows. In numerous computer vision applications, it has produced state-of-the-art results, and it serves as the basis for numerous sophisticated neural network structures.

G. VGG19

Convolutional neural networks (CNNs) with the architecture VGG19 are renowned for their deep structure and excellent accuracy in recognizing images. There are three completely linked layers and sixteen convolutional layers, totaling 19 layers in this structure. Every convolution block in VGG19 consists of several convolutional layers, which are then succeeded by a max-pooling layer. This helps to gradually decrease the input's spatial dimensions while increasing its depth. Smaller 3x3 convolutional filters are used throughout the network to enable VGG19 to effectively learn complex characteristics from images. In spite of its complexity, VGG19 is easy to comprehend and use, which makes it a popular option for a range of computer vision applications, such as tasks involving object recognition and medical picture analysis.

VII.RESULTS

In this study, we used combined machine learning as well as deep learning models to construct a prediction system for heart illness, which we then integrated into a Flask web-based application to provide predictions in real time. A strong basis for prediction was given by the machine learning models, which included linear regression, support vector machine (SVM), random forest (RF), and XGBoost. Out of all of them, XGBoost's accuracy of 90% showed how reliable and effective it is at processing structured ECG data. With accuracy rates of 89%, support vector machine and Random Forest both fared well. Even though it was less complex, linear regression produced an accuracy rate of 85%, which was used as a standard.



In order to extract intricate patterns from the ECG data for deep learning, we utilized the ResNet and VGG19 convolutional neural network designs. With the greatest accuracy of 96%, VGG19 surpassed other models because of its deep feature extraction capabilities. With a 94% accuracy rate, ResNet likewise displayed impressive results. Users' easy interaction allowed for the development of an extensive instrument for heart ailment identification, and the incorporation of these mathematical frameworks into the Flask interface allowed for real-time prediction. The attached bar chart shows the visual contrast between model accuracies and shows that VGG19 is the best-performing model.

VIII.FUTURE DISCUSSION

Future presentations may explore various approaches to the mission's expansion. To further improve prediction accuracy, these include investigating the integration of new deep learning architectures or ensemble techniques. Furthermore, examining the integration of actual patient information streams and ongoing model optimization via feedback mechanisms may enhance flexibility and efficacy in dynamic healthcare settings. Additionally, evaluating the prediction model's scalability and implementation across various healthcare environments and demographics may shed light on its wider application and influence on public health campaigns. These approaches seek to develop deep learning-based cardiovascular disease prediction with the ultimate goal of providing more efficient, individualized healthcare.

IX. CONCLUSION

In summary, deep learning-based improvements in heart disease prediction, especially with respect to ECG picture interpretation, indicate a revolutionary path forward for the medical field. More precise and individualized risk evaluations may be possible with the combination of complex neural network architectures with multi-modal data sources like genetic data and thorough medical records. The implementation of explainable AI approaches in clinical practice is contingent upon the continuous development of strategies that further augment the openness and trust of predictive models. The ultimate goal of these developments is to provide medical professionals with strong instruments to better identify, treat, and prevent cardiovascular illnesses, which will enhance patient outcomes and promote precision healthcare across the board.

X. REFERENCES

[1] Roy, T. S., Roy, J. K., & Mandal, N. (2022). Classifier identification using deep learning and machine learning algorithms

[2] Yaganteeswarudu, A. (2020, June). Multi disease prediction by using machine learning and Flask. In 2020 5th International conference on (ICCES) (pp. 1242-1246). IEEE.

[3] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction by using machine learning. SN Computer Science, 1(6), 345.

[4] Katarya, R., & Meena, S. K. (2021). Machine learning techniques for heart disease prediction: a comparative study and analysis. Health and Technology, 11(1), 87-97.

[5] Kavitha, M., Gnaneswar, (2021, January). Heart disease prediction using hybrid machine learning model. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1329-1333). IEEE.

[6] Boukhatem C. Heart disease prediction using machine learning. IEEE Advances in Science and Engineering Technology International Conferences 2022 Feb 21–24, Dubai, United Arab Emirates

[7] Golawar, Dhiwar, A., Wakchoure, S., Bhoite, S., & Khadke, D. (2022, May). Heart disease prediction using machine and deep learning. In 2022 international conference on computational intelligence and sustainable engineering solutions (CISES) IEEE.

[8] Barhoom, A. M., Abu-Nasser, B. S., & Abu-Naser, S. S. (2022). Prediction of heart disease using a collection of machine algorithms.

[9] Riyaz, L., Butt.Zaman, M., & Ayob, O. (2022). Heart disease prediction using machine learning techniques: a quantitative review. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3 (pp. 81-94). Springer Singapore.

[10] Nancy, A. A., D., Raj Vincent, P. D., Srinivasan, K., & Gutierrez Reina, D. (2022). Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. Electronics, 11(15), 2292.

[11] Arumugam, K., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings, 80, 3682-3685.

[12] Valarmathi, R., Sheela, T. (2021). Heart disease prediction using hyper parameter optimization tuning. Biomedical Signal Processing and Control, 70, 103033.

[13] Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M. (2020). Exploring feature selection and classification methods for heart disease. Digital health, 6, 2055207620914777.

[14] Wang, K., Tian, J., Zheng, C., Yang, H., Ren, J & Zhang, Y. (2021). 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, 137, 104813.

[15] Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. In Machine learning and the internet of medical things in healthcare (pp. 89-111). Academic Press.

[16] Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning. Healthcare Analytics, 2, 100032.

[17] Desai, R. J., Wang, S. V., Vaduganathan, & Schneeweiss, S. (2020). Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. JAMA network open, 3(1), e1918962-e1918962.

[18] Singh, A. (2020, February). Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE.