International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Advanced Cardiovascular Disease Prediction: A Comparative Analysis of Ensemble Stacking and Deep Neural Networks

[1]B. Rupa Devi, [2]U. Sivaji, [3]Thammisetty Swetha, [4]Dr. J. Avanija, [5]Dr. A. Suresh, [6]Dr. K. Reddy Madhavi

Submitted: 26/09/2023 **Revised**: 16/11/2023 **Accepted**: 27/11/2023

Abstract: Advancements in computational powers and approaches have significantly diversified the field of medical sciences, particularly in diagnosing cardiovascular problems in humans. Cardiovascular disease (CVD) is a complex and demanding ailment that has a substantial negative influence on the worldwide population, substantially affecting human health and quality of life. Early and accurate identification of cardiovascular disorders in individuals can offer significant benefits in treating heart failure in its early stages, hence improving the chances of the patient's survival. The manual detection of cardiac disease is susceptible to bias and subject to discrepancies among examiners. Machine learning algorithms have shown effectiveness and dependability in accurately detecting and categorizing individuals with heart disease and those in a normal state of health. This research paper presents a novel method using an ensemble deep neural network to detect heart disease precisely. Our approach integrates two deep learning networks as fundamental models, together with an XGBoost boosting model acting as the meta-classifier to generate ultimate predictions. The experimental results demonstrate that our ensemble model surpasses the existing methodologies documented in the literature, showcasing its higher efficacy in detecting cardiac disease. This novel technique shows potential for improving the precision and dependability of heart illness detection, thereby benefiting clinical practice and patient care.

Keywords: *machine learning, cardio disease, ensemble learning, XGB, deep networks*

1. Introduction

In 2019, data from the World Health Organization [1] revealed that Cardiovascular diseases (CVDs) accounted for around 32% of all global deaths, amounting to 17.9 million fatalities. Moreover, CVD was responsible for over 17.7 million fatalities yearly [2]. Australian Institute of Health and Welfare research shows that in 2018, cardiovascular disease was the top cause of death in Australia, accounting for 42% of all deaths. The present study aims to investigate and develop a more effective method for rapidly identifying cardiovascular problems. This is driven by the constraints of current diagnostic methods, which have been insufficient in detecting heart conditions at an early stage due to imprecision and time-consuming computing procedures. The diagnosis and treatment of heart disease present considerable difficulties without easily accessible modern technology and healthcare knowledge [4], [5].

[1]*Associate Professor, Department of Computer Science and Engineering, Annamacharya Institute of Technology and Sciences, Tirupati, A.P, India Email: rupadevi.aitt@annamacharyagroup.org*
[2]*Associate Professor, Department of IT, Institute of Aeronautical Engineering, Dundigal, Hyderabad. Email: sivaji.u117@gmail.com*
[3]*Assistant Professor, Department of CSE-Data Science, Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India. Email: swethathammisetty7@gmail.com*
[4]*Professor, AI&ML,School of Computing, Mohan Babu University,Tirupati, A.P., India. Email: avans75@yahoo.co.in*
[5]*Associate Professor, Department of Computer Science and Engineering, Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India. Email: csesuresh6@gmail.com*
[6]*Professor, AI&ML, School of Computing, Mohan Babu University, Tirupati, A.P., India. Email: kreddymadhavi@gmail.com*

A considerable proportion of persons can see a major rise in their life expectancy by receiving a precise diagnosis and efficient treatment [6]. The patient's medical history, the findings of a physical examination, and the presence or absence of any concerning symptoms are all considered when diagnosing cardiac disease. However, it is necessary to review the outcomes of this diagnostic method when identifying persons with heart disease. Further examination entails substantial expenses and poses computational difficulties [7]. We have created a noninvasive forecasting technique to tackle these difficulties by employing machine learning classifiers based on neural networks. Precisely detecting cardiac problems is effectively accomplished by employing an advanced decision-making framework that integrates machine learning classifiers and artificial fuzzy logic.

Consequently, the mortality rate has declined [8, 9]. Several studies have utilized the Cleveland heart disease dataset. Having suitable and pertinent data to train and evaluate prediction models in machine learning is crucial. Using a refined or standardized dataset during both the training and testing phases can greatly enhance the accuracy of machine learning classifiers. Furthermore, incorporating relevant and related data properties could enhance the model's ability to make accurate predictions. Therefore, data standardization and feature selection procedures greatly impact the precision of machine learning classifiers. Multiple researchers have utilized different prediction

strategies in previous studies. However, these approaches have demonstrated inefficacy in accurately forecasting cardiac diseases. Data standardization is crucial for improving the accuracy of machine learning classifiers. Various standardization methods, such as standard scalar, min-max scalar, and others, remove instances with missing feature values from the dataset.

Heart disease prediction requires the administration of multiple tests. The prognosis of cardiovascular illness is particularly difficult, especially in countries with restricted availability of healthcare providers and testing equipment [10]. Computational classifiers can provide precise medical diagnoses when provided with sufficient data. Multiple machine learning techniques can be employed to forecast CVD risks. The majority of these solutions utilize publically accessible datasets for both training and evaluating the models. The data sets have facilitated the development of cutting-edge CVD risk prediction systems using machine learning-based predictive algorithms. These databases contain records of both risk factors and diagnosis. Preprocessing is essential in CVD prediction systems [11] due to clinical datasets' lack of reliability and duplication.

Furthermore, risk factors (characteristics) can be selected based on criteria such as widespread occurrence, autonomous influence on cardiovascular disease, and modifiability. Scientists have utilized a range of risk variables and characteristics to build forecasts for CVD. Machine learning algorithms heavily rely on pertinent datasets [12]. The absence of adequate medical data, selecting relevant features, implementing machine learning algorithms, and conducting thorough analysis can pose challenges in predicting heart disease. Our research aims to enhance the accuracy of CVD prediction by addressing these deficiencies.

Moreover, there are constraints on the current study datasets. These databases should incorporate additional risk variables and clinical data attributes. The clinical severity of a specific instance can impact the accuracy of a forecast [13]. These problems should have been addressed in previous research. Contemporary research has not yet established a uniform format for data sets or optimized methods. Scientists have developed many machine learning models, such as SVM, KNN, RF, DT, LR, NB, and others, to enhance the accuracy of predicting cardiac illnesses [14].

However, it is still difficult to make reliable predictions about heart illness. It is crucial to design a novel and cost-effective technique to predict the likelihood of cardiovascular illness accurately. The level of complexity surrounding NB, BN, RF, and MLP has not yet been properly characterized. The age risk factor, absent in NB, BN, RF, and MLP, is represented as the age element inside the dataset. The system was examined using the StatLog datasets. The model employed in the Cleveland dataset has excluded numerous significant risk factors, including age, RestECG, ST Depression (Slope), and others [15]. No significance tests were conducted during the standardization procedure of the proposed approach. The datasets used in this study are the StatLog dataset [16] and the Z-Alizadeh Sani dataset. The dataset displays a reduced magnitude. The resulting output was not compared to other datasets to achieve standardization. The Cleveland dataset was utilized instead.

2. Literature Work

Doctors relied heavily on auscultation to identify abnormal cardiac sounds. All heart illnesses were diagnosed by doctors using stethoscopes. The auscultation method used by doctors to diagnose heart problems has its limitations. Clarity and classification of heart sounds are determined by doctors' ability and expertise acquired via thorough tests. Machine learning methods have been proposed as an alternative to human-based CVD detection.

Amin et al. [17] did a study to identify the best risk factors for cardiovascular disease. NB, KNN, LR, DT, NN, SVM, and Vote are only some of the seven classification methods that can be utilized. With 303 records and 76 attributes, Cleveland datasets The authors' models are trained and evaluated using 10-fold cross-validation. Instead of splitting the data, which would have led to an inaccurate estimate of the model's expected performance due to a smaller training set, we opted for 10-fold cross-validation. However, if you apply 10-fold validation, the model will access 90% of the data. The accuracy of the Vote Classifier rose to 87.4 percent with its help.

Anitha et al. [18] employed learning vector quantization techniques to diagnose cardiac illness. The algorithm mentioned above achieved an accuracy of 85.55 percent. The datasets utilized in this study were obtained from the machine learning collection at the University of California, Irvine (UCI). There are a total of 303 records and 76 attributes in these databases. Preprocessing procedures were applied to the data to address the problem of missing values. Since only 14 features were considered for inclusion in the heart disease study's sample, only 302 records were included in the analysis. Models are tested on 30% of the dataset, while 70% is used for training.

Rajdhan et al. [19] created an online machine learning-based cardiac illness prognosis tool. Train and assess models with LR, NB, and SVM classification algorithms. UCI's machine learning repository split the Cleveland datasets into 75% training and 25% testing sets. After data preparation, SVM accuracy increased by 64.4% to remove errors and missing values. The study couldn't pinpoint early heart disease causes.

Devansh Shah et al. [20] analyzed a dataset with 303 samples and 76 attributes using supervised learning methods such as decision trees, Naive Bayes, K-NN, and random forests. According to the results of this research, the K-NN algorithm is performing at its highest degree of accuracy. The authors used machine learning classifiers to create a cardiovascular disease prediction model [21]. There are 14 attributes used in the UCI Cleveland dataset for training and assessing models. Linear regression was the least effective classifier (78%), whereas decision trees were the next best (89%), followed by support vector machines (83%) and K-NN (87%). According to the results, the K-NN algorithm was the most accurate of the approaches examined. Classification methods such as SVM, LR, ANN NN, K-NN, etc., were used in the research by Khan et al. [22]. The new model improved the old one's accuracy by 2.37 percentage points.

Linda et al. [23] proposed a novel health information system to deliver individualized advice on exercise for people with cardiovascular disease. The results thus far suggest that doctors need guidance in deciding how much exercise to recommend to individuals with multiple cardiovascular disease risk factors. The supplied method is a fast, convenient, and scientifically sound patient choice. Disease risk factors can be assessed using the PB-FARM technique, which Ali et al. [24] laid out in three easy steps. We also used the Z-Alizadeh Sani dataset to analyze the elements that contributed to the disease's spread. The findings are directly related to the prevalence of CAD, the severity of normal chest pain, and advanced age. It was proposed by Rubini et al. [25] that prediction models be developed for cardiovascular disease. Compared to the suggested method, the accuracy of the random forest was the highest among LR, NB, and SVM classifiers (84.81%).

Drod et al.'s [26] main goal was to use machine learning approaches to determine which risk factors for CVD are most important in people with a metabolic-associated fatty liver disease (MAFLD) diagnosis. Blood chemistry and subclinical atherosclerosis were assessed on a sample of 191 patients with MAFLD. Individuals at the highest risk for CVD were identified using a machine learning model that used many methods, including a multiple logistic regression classifier, univariate feature ranking, and principal component analysis. The most notable clinical features were hypercholesterolemia, plaque scores, and diabetes duration. The machine learning method performed admirably, properly categorizing 40 of 47 high-risk patients and 114 of 144 low-risk patients with an accuracy percentage of 79.17%. An AUC of 0.87 was obtained as a result of this analysis. The study's results support using machine learning to identify patients with MAFLD who already have cardiovascular disease

3. Comparative Study

Initially, the most well-known machine learning models were introduced. After that, we used the available dataset to train the models. The effectiveness of each model was then assessed through a comparison study. This method lets us learn more about the project and its models and determine which works best. Seven methods were used to assess machine learning models for CVD prediction datasets.

We utilized various machine learning models in our comparison analysis to evaluate their effectiveness in accomplishing the given goal. The models utilized in this study encompassed LR, RF, KNN, DT, SVM, GB, and AdaBoost. Each model possesses distinct strengths and traits, rendering them suitable contenders for the examination. Through a systematic process of testing and evaluating various models, our objective was to obtain useful insights into their strengths and limitations, ultimately assisting in selecting the most appropriate model for our particular application.

Logistic regression is a commonly used statistical model for binary categorization. The algorithm computes the probability that a particular event belongs to a particular category. Logistic regression uses a logistic function to describe the probability and create a linear border that separates the classes. With its simplicity, speed, and interpretability, this model is an excellent benchmark for binary classification tasks.

The Random Forest approach, widely used in machine learning, is commonly applied to solve classification and regression problems. The operational methodology entails the creation of multiple decision trees during the training phase. To make predictions, the algorithm aggregates the outputs of these trees by choosing the class that occurs most frequently for categorization purposes. Random Forests are widely recognized for their ease of use, ability to handle many features, and resilience in handling missing data. These qualities contribute to its reputation for achieving high levels of accuracy in various machine-learning applications.

The KNN algorithm provides a direct and simple method for machine learning, mainly employed for classification and regression applications. The operation compares a specific test instance with the closest instances in the training set using Euclidean distance in the feature space. KNN utilizes a proximity-based approach to ascertain the class of the test instance. The simplicity and efficacy of KNN make it a desirable option for baseline modeling in various applications. Decision trees are flexible models that handle various tasks like classification, regression, and multi-output issues. These trees divide data into groups based on the importance of input features, resulting in the formation of a decision tree through a repetitive process. Although

decision trees are easy to understand and comprehend, they are susceptible to overfitting, particularly when working with datasets with many attributes.

SVMs are highly effective in supervised learning, especially for binary classification problems. SVMs determine the ideal hyperplane that efficiently divides classes in the feature space. SVMs strive to provide strong generalization performance by maximizing the gap between the nearest data points of distinct classes, referred to as support vectors. Gradient Boosting is a popular ensemble learning method commonly used for classification and regression applications. The algorithm constructs strong predictive models by aggregating predictions from weak learners built sequentially, typically in the form of decision trees. Gradient Boosting employs an iterative process to rectify faults generated by preceding trees, resulting in significant enhancements in the quality of the model.

AdaBoost, also known as Adaptive Boosting, is an ensemble learning technique that aims to improve the performance of weak learners or simple models to generate a strong ensemble. AdaBoost algorithm focuses on misclassified instances by modifying their weights in the dataset, directing the machine to learn from its errors. The iterative procedure usually leads to substantial improvements in the accuracy of predictions. The common ML pipeline process is shown in Figure 1. We took a methodical, multi-step strategy to assure the robustness and dependability of the predictive models. The initial step was to compile a complete dataset consisting of cardio-related variables of the highest quality. Following data collection, we performed data cleaning and exploratory data analysis (EDA) to address any discrepancies or missing information.

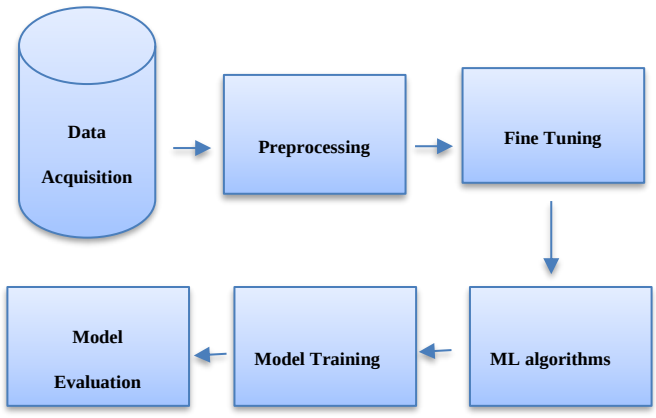

Fig 1: Experimental analysis of numerous models

In addition, we used numerous visualization techniques to investigate the underlying patterns and connections between the features. We began the standardization process to guarantee that our models always had access to the same high-quality data. The features were then rescaled with a mean of 0 and a standard deviation 1. Standardization is essential for models sensitive to the magnitude of input features since it speeds up convergence and improves performance.

After cleaning and standardizing the data, the next step was to tune the hyperparameters. Using methods such as Grid Search and Random Search, we optimized the settings of numerous algorithms to determine the best possible values for their hyperparameters. The models' performance was much improved by this tuning, which also ensured that they were optimally set up to detect hidden structures in the data. We then used several machine learning algorithms—LR, RF, KNN, SVM, DT, SVM, GB, and AdaBoostwith their respective hyperparameters fine-tuned. These algorithms were trained on subsets of the dataset and then tested on a separate subset [28, 29].

After the models were trained, the hidden validation data assessed their efficacy in detail. Several metrics were computed to evaluate the models' performance in making cardiovascular disease predictions. False positive, false negative, true positive, and true negative rates were visualized via confusion matrices. With an accuracy of 73%, Gradient Boosting and AdaBoost emerged as the top models based on the computed metrics in the comparison study. Through this systematic evaluation procedure, we gained insight into the advantages and disadvantages of each model, which helped us select the best model for use in predicting cardiovascular illnesses [30].

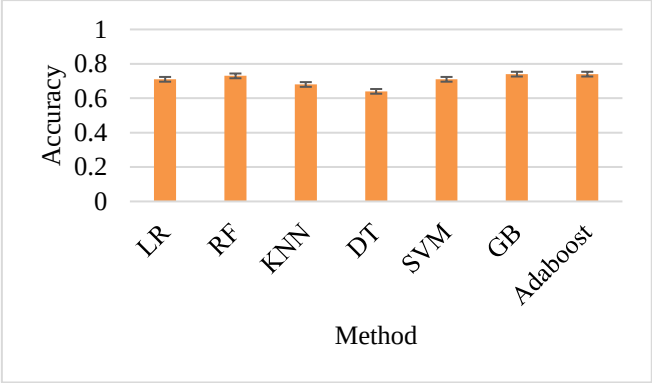

Fig 2: Comparing CVD dataset accuracy with different ML algorithms

An accuracy of 71% was reached by both the LR and RF models, indicating a decent showing in foresight. The KNN model achieved 69% accuracy, whereas the DT model only managed 63%. On the other hand, the SVM model performed somewhat better, achieving an accuracy of 72%. Notable is the fact that both the GB with a combined 73% accuracy. The relative strengths of different machine learning models for predicting cardiovascular illnesses are illuminated by this comparison [31]. The accuracy of the Gradient Boosting and AdaBoost models is just slightly higher than that of the others. Figure 2 demonstrates these models' high degree of accuracy

4. Methodology

We have diligently utilized the well-constructed process described in this section in our research, as mentioned in Algorithm 1. This procedure is organized and includes collecting, preprocessing, and splitting data. After that, we define and apply regularisation to our ensemble network, compile the model, and do a thorough evaluation. Through strict adherence to this complete methodology, we have developed a strong framework for our research, guaranteeing transparency and the possibility to reproduce our approach to detecting heart disease.

Data Collection: Our research commences by meticulously gathering data relevant to identifying cardiac disease. We gather a dataset with several characteristics, including age, a crucial factor in our analysis. The thorough process of collecting data guarantees that we possess a wide-ranging and varied assortment of information to train and test our models.

Data preprocessing involves performing crucial procedures to prepare the data for model training. These tasks encompass managing null values, standardizing or rescaling the attributes, and encoding category variables if needed. Our main attention is on the age characteristic, ensuring it is altered appropriately to meet our modeling needs. Feature selection approaches are utilized to discern the most pertinent variables, diminishing the dataset's complexity and extraneous information.

Data Splitting: The dataset is divided into two subsets: the training set, which accounts for 80% of the data, and the test set, which makes up the remaining 20%. This segregation guarantees that our models are trained on a significant percentage of the data while preserving a separate dataset for thorough evaluation. It prevents overfitting and offers a dependable evaluation of model generalization.

Ensemble Network Definition: Our novel strategy centers on establishing an ensemble network. This ensemble comprises two deep learning models plus XGBoost, a robust gradient boosting-based model utilized as the underlying classifier. The integration of various algorithms in the ensemble leverages their combined capabilities, hence improving the forecast precision of our system.

Regularisation and dropouts are employed to address the issue of overfitting by including these approaches in our deep-learning models. Techniques like L1 or L2 regularisation are used to manage the complexity of networks, ensuring that they do not overfit the noise in the data. In addition, dropout layers are used to randomly turn off neurons during training, enhancing the model's resilience.

Model Compilation: During this stage, we establish the settings and construct our ensemble network, combining the deep learning models with XGBoost. Thorough parameter tweaking and hyperparameter optimization are conducted to ensure the ensemble functions efficiently. This stage establishes the fundamental basis for efficient model training and prediction.

Model Training: We commence the training procedure for our ensemble network using the training dataset. By repeatedly improving the models, the ensemble can effectively understand and represent complex patterns and connections in the data. Training entails adjusting the weights and biases of the model to minimize prediction mistakes.

Algorithm 1: Steps involved in the proposed flow

1. Load the heart disease dataset from a CSV file, specifying a semicolon as the separator.
2. Display a random sample of five rows from the dataset and the dataset's shape.
3. Convert the 'age' column from days to years.
4. Drop the 'id' column for data cleanliness.
5. Remove duplicate rows from the dataset.
6. Calculate the Interquartile Range (IQR) for each attribute in the dataset and identify potential outliers by checking if data points fall below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR. Print the count of outliers for each attribute.
7. Calculate the Body Mass Index (BMI) using the 'weight' and 'height' columns and calculate the Mean Arterial Pressure (MAP) using the 'ap_lo' and 'ap_hi' columns.
8. Identify and drop columns with a unique value count 1 (constant columns) to prevent redundancy.
9. Define a function 'bin_using_kmodes' for categorical binning using K-Modes clustering for 'MAP,' 'BMI,' and 'age' columns. Define bins, labels, and bin the data. Convert the categorical bins to numeric values.
10. Select an optimal value for 'k' (number of clusters) based on a plot (placeholder).
11. Perform K-Modes clustering with the determined 'k' and add the cluster labels to the original dataset as the 'Target' column.
12. Define a neural network model using TensorFlow/Keras, create two instances ('network1' and 'network2'), and train both networks on the data for a specified number of epochs.
13. Obtain predictions from both networks on the validation set and stack these predictions horizontally to create features for the meta-model.

14. Define a meta-model using the XGBoost classifier and train it on the stacked predictions from the neural networks.
15. Make predictions on the test set using both 'network1' and 'network2,' stack these predictions horizontally, and use the meta-model (XGBoost) to make final ensemble predictions on the test set.

Model Evaluation: During the final assessment phase, we apply our ensemble network to the test dataset, generating predictions for data points that have not been observed before. We thoroughly assess its performance by employing a variety of evaluation criteria, including accuracy, precision, recall, and F1-score. These measures offer a thorough comprehension of the model's efficacy in detecting cardiac disease.

Through careful adherence to this systematic approach, our goal is to showcase the exceptional prediction skills of our ensemble network, outperforming current methodologies in the domain of heart disease detection. Our study enhances the progress of precise and dependable identification of heart diseases, ultimately helping the practice of medicine and the well-being of patients. The complete steps involving the proposed methodology are shown in Algorithm 1.

The technique is initiated by loading and preparing a dataset about heart disease. The preprocessing steps involve changing the 'age' column from days to years, eliminating duplicate entries, and detecting outliers using the Interquartile Range (IQR) method. In addition, it computes supplementary measurements such as BMI and MAP Columns that contain repetitive values are eliminated. The data is subjected to categorical binning using K-Modes clustering on the 'MAP,' 'BMI,' and 'age' columns, resulting in a more organized depiction of the data.

Subsequently, the procedure ascertains the most advantageous quantity of clusters ('k') for K-Modes clustering and appends cluster labels to the initial dataset. The ensemble modeling phase comprises creating two neural network models, namely 'network1' and 'network2', followed by their training using the available data. Their forecasts on the validation set are combined to create characteristics for a meta-model, which is constructed using XGBoost. Ultimately, the ensemble model, which includes the meta-model and both neural networks, generates predictions on the test set, providing a reliable and precise solution for detecting heart disease.

According to [23], this research used 70,000 patient records and 12 characteristics. Table 1 shows sample rows (5 records) and columns (6 columns) from the dataset. Factors include age, sex, systolic and diastolic blood pressure.

Table 1. Sample records from the CVD dataset (source: Kaggle [23])

A	H	S	APL	C	G	Al
0	157	1	91	2	1	0
0	162	0	90	1	2	0
1	151	0	80	2	1	1
1	165	0	70	1	1	0
1	170	1	81	2	1	1

Cardiovascular disease is "cardio," while excellent health is "cardio-healthy." The classwise sample count is shown in Figure 3. The dataset has a balanced distribution since 50.03% of samples are class 0, and 49.97% are class 1. Training and testing datasets comprise 80% and 20%, respectively. Models are "trained" on one set of data and "tested" on another. The clustered dataset is used to evaluate various classifiers, including decision trees, random forests, multilayer perceptrons, and XGBoost. Then, we put each classifier through its paces by measuring its accuracy, precision, recall, and F-measure.

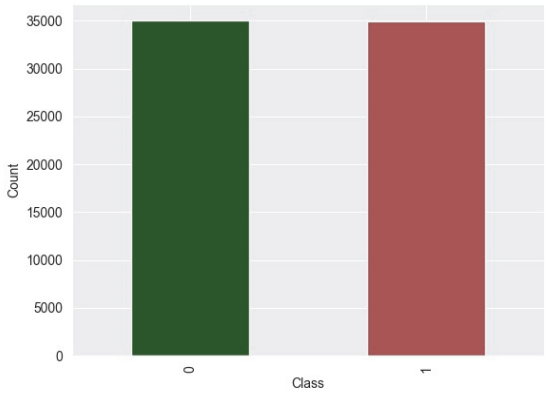

Fig 3: Classwise count in the dataset (no-disease:0, disease:1)

5. Results and Discussion

5.1 Interpretation and Analysis

This study involved a comprehensive assessment of four machine learning models to detect heart disease accurately. The evaluation was conducted using the widely recognized heart disease dataset. The models included the Random Forest, MLP (Multilayer Perceptron), Decision Tree, and XGBoost classifiers. The study involved a sequence of crucial steps to evaluate their performance thoroughly:

Data Splitting and Model Selection: The initial phase of our research involved dividing the dataset into two distinct subsets: an 80% training set to train the model and a 20% testing set to evaluate its performance. Afterward, we chose

four well-known machine learning models for their efficacy in categorization tasks.

Model Training and Evaluation: We performed the following essential tasks for each of the selected models:

Model Training: The chosen models underwent training using the provided training data, allowing them to acquire knowledge of complex patterns and relationships inherent in the dataset.

Model Prediction: We deployed these trained models to generate predictions on the previously unseen test data following the training phase.

Metrics for evaluating performance:

To thoroughly assess the effectiveness of each model, we calculated multiple fundamental evaluation metrics:

Accuracy: This statistic comprehensively evaluates a model's predictions, including its capacity to accurately categorize both positive and negative cases.

Precision is a metric that assesses the ability of a model to anticipate positive outcomes accurately. It is calculated by determining the ratio of true positive forecasts to all positive predictions.

Recall measures how well a model can identify all real positive occurrences in a dataset. It calculates the ratio of correctly predicted positive cases to the total number of positive cases.

The F1-Score is a comprehensive metric that considers precision and recall, assessing a model's total performance fairly.

The confusion matrix comprehensively analyzes a model's predictions, encompassing accurate positive predictions, negative forecasts, incorrect positive predictions, and incorrect negative predictions.

Results and Interpretation: Our study revealed the relative effectiveness of the four models in detecting heart disease:

The Random Forest and XGBoost models were identified as the most successful, demonstrating superior accuracy, precision, recall, and F1-Score levels. This highlights their efficacy in accurately predicting and identifying cases of heart disease.

The MLP (Multilayer Perceptron) model exhibited robust performance, albeit with a somewhat poorer precision than Random Forest and XGBoost. Although producing satisfactory outcomes, the Decision Tree model exhibited slightly worse precision and recall than the other models.

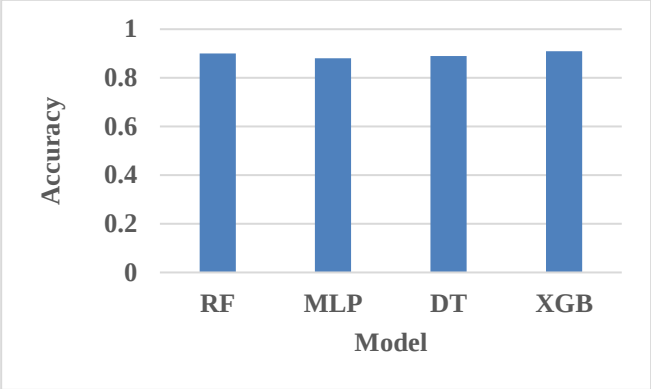

Fig 4: The comparison results after preprocessing data

Ultimately, our research highlights the potential of the Random Forest and XGBoost models as highly promising options for detecting cardiac disease. Their exceptional classification accuracy and precision performance indicate their appropriateness for practical applications. Nevertheless, choosing the most suitable model for practical implementation may necessitate further deliberations, such as fine-tuning hyperparameters and ensuring compatibility with specific application needs. These discoveries enhance the field of cardiac disease diagnostics, providing useful information for healthcare decision-making and patient treatment. The comparison results are shown in Figure 4.

For our investigation, we began by importing a dataset in CSV format using the Pandas module in Python. An initial data analysis was performed by randomly choosing and evaluating five records. In addition, we assessed the quantity of the dataset to gain an early comprehension of its structure and dimensions.

5.2 Deep Ensemble Network

To improve the consistency and quality of the data, we conducted several data preparation procedures. Initially, our attention was directed towards the 'age' feature, which we converted into a numerical value representing years and then rounded to the nearest whole number. The implementation of this transformation guaranteed consistency in the depiction of age throughout the dataset. We eliminated the 'id' column since it did not offer significant data for our subsequent study.

We executed a procedure to remove duplicate rows to enhance the dataset and eliminate any possible instances of duplication. Following this procedure, we verified the dimensions of the dataset to ensure the successful removal of duplicate entries, thus guaranteeing data integrity.

Subsequently, we directed our focus on outlier detection, an essential stage in the process of data preprocessing. By utilizing the Interquartile Range (IQR) approach, we determined the first quartile (Q1) and third quartile (Q3) for each attribute and then calculated the IQR. Outliers were defined as data points below Q1 minus 1.5 times the

interquartile range (IQR) or above Q3 plus 1.5 times the IQR. This procedure yielded valuable observations regarding any irregularities within the dataset which may impact the modeling process.

We concentrated our feature engineering endeavors on incorporating two novel attributes: BMI (Body Mass Index) and MAP (Mean Arterial Pressure). The 'BMI' was calculated using the weight and height traits, providing an extra parameter pertinent to health evaluation. In contrast, the 'MAP' metric was obtained by combining diastolic ('ap_lo') and systolic ('ap_hi') blood pressure measurements, serving as a crucial physiological indicator. We performed feature selection to enhance the dataset's efficiency and maximize the model's performance. Concretely, we detected and eliminated columns that had only one distinct value. This technique removed unnecessary data, decreasing the complexity of the dataset and making subsequent modeling jobs easier.

Clustering was a crucial component of our data preparation method. We utilized k-mode clustering on two primary attributes: 'BMI' and 'MAP.' Using pre-established criteria, we established bins and labels for the 'MAP' dataset, enabling us to classify the data more efficiently. The appropriate number of clusters for 'BMI' was calculated by analyzing the distribution of data points, aiming to achieve a balanced and useful representation of this trait.

The central focus of our research was our suggested ensemble model, DNN-Meta Model. DNN-Metal model consists of two deep neural networks (DNNs) and uses XGBoost as the meta-model. The model architecture of DNN-Metal was carefully built. Each deep neural network (DNN) comprised numerous layers that integrated different activation functions, dropout layers, and batch normalization techniques to mitigate the problem of overfitting. These networks were extensively trained using the preprocessed dataset, acquiring detailed knowledge of complex patterns within the data. The predictions produced by the two deep neural networks (DNNs) on the training dataset were combined to form a collection of stacked predictions. The stacked predictions and the original dataset attributes were used as input for the meta-model, XGBoost. The ensemble model underwent specialized training to enhance its predictive performance on the test dataset.

Ultimately, our research peaked by assessing the ensemble model's effectiveness. We utilized fundamental assessment criteria, such as precision, recall, F1-score, and the confusion matrix, to thoroughly evaluate the model's effectiveness in diagnosing heart disease. The measurements offered a quantitative and insightful viewpoint on the model's performance and capacity to generate precise predictions. Using a complex and organized approach, we aimed to create a strong and accurate model for detecting cardiac disease. We implemented a multi-step procedure that involved data preparation, feature engineering, clustering, and ensemble modeling. This comprehensive approach helped us successfully fulfill our study aims.

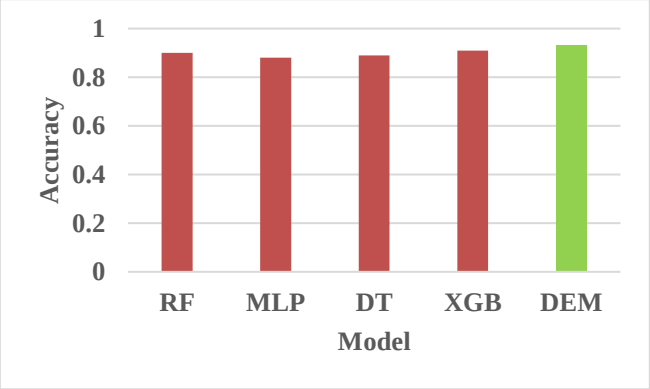

Fig 5: The Accuracy of the Proposed Model (DEM)

We analyzed the performance of different machine learning models in detecting cardiac disease, using accuracy scores as the primary evaluation parameter. Out of all the models, the "DEM" (Proposed Model) distinguishes itself by achieving the greatest accuracy of 93%, indicating its excellent capacity to classify instances of cardiac disease accurately. These findings indicate that the suggested model has the potential to be a reliable tool for diagnosing cardiac disease. The XGBoost model has robust performance, with an accuracy rate of 91%, thus highlighting its efficacy in generating precise predictions. The "RF" (Random Forest) model exhibits a high accuracy rate of 90%, demonstrating its reliability in detecting cardiac illness. The DT and MLP models demonstrate satisfactory performance, with 89% and 88% accuracy, respectively. This indicates that these models are well-suited for the job. Overall, these accuracy scores offer useful insights into the comparative performance of different models, with the "DEM" model being identified as the top-performing solution. The comparison results are shown in Figure 5

6. Conclusion

This research utilizes an innovative methodology in the realm of medical sciences to tackle the vital objective of accurate heart disease identification. Our strategy uses the progress in computational capabilities and methodology by using ensemble deep neural networks for this task. We combine two deep learning networks as primary models, supplemented by an XGBoost boosting model acting as the meta-classifier to deliver final predictions. The experimental findings in this paper demonstrate the efficacy and dependability of our ensemble model in diagnosing cardiovascular illnesses. Our methodology exhibits more efficacy when compared to the methodologies described in the existing literature. This novel methodology can greatly improve heart disease identification's precision and

reliability, providing important implications for clinical practice and patient care. This is a positive advancement in utilizing machine learning to enhance the diagnosis and treatment of cardiovascular problems, ultimately leading to better patient outcomes and well-being. Future research might examine improvements in the suggested methodology using more diverse datasets and better deep learning architectures. Furthermore, it is possible to focus on enhancing the ensemble model's interpretability and scalability, enhancing its practicality in real-world clinical environments

References

[1] Cardiovascular Diseases (Cvds), "World health organization," https://www.who.int/news-room/fact%20sheets/detail/cardio vascular-diseases-(cvds).

[2] Alkayyali, z. K., idris, s. A. B., & abu-naser, s. S. (2023). A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis. *Journal of Theoretical and Applied Information Technology*, *101*(4).

[3] A Coronary, "Heart disease," Available from: https://www. aihw.gov.au/reports/australias-health/coronaryheart-disease, 2020

[4] Rath, A., Mishra, D., Panda, G., & Pal, M. (2022). Development and assessment of machine learning based heart disease detection using imbalanced heart sound signal. *Biomedical Signal Processing and Control*, *76*, 103730..

[5] Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, *16*(2), 88.

[6] Dileep, P., Rao, K. N., Bodapati, P., Gokuruboyina, S., Peddi, R., Grover, A., & Sheetal, A. (2023). An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm. *Neural Computing and Applications*, *35*(10), 7253-7266.

[7] Nandy, S., Adhikari, M., Balasubramanian, V., Menon, V. G., Li, X., & Zakarya, M. (2023). An intelligent heart disease prediction system based on swarm-artificial neural network. *Neural Computing and Applications*, *35*(20), 14723-14737.

[8] Nancy, A. A., Ravindran, 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..

[9] Hassan, D., Hussein, H. I., & Hassan, M. M. (2023). Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. *Biomedical Signal Processing and Control*, *79*, 104019.

[10] Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, *3*, 100130.

[11] Nouman, A., & Muneer, S. (2022). A systematic literature review on heart disease prediction using blockchain and machine learning techniques. *International Journal of Computational and Innovative Sciences*, *1*(4), 1-6.

[12] War, M. M., & Singh, D. (2023, February). Review On Enhancing Healthcare Services for Heart Disease Patients using Machine Learning Approaches in Cloud Environment. In *2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM)* (pp. 1-5). IEEE..

[13] Reddy, D. J., & Kumar, M. R. (2021, May). Crop yield prediction using machine learning algorithm. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1466-1470). IEEE.

[14] Kumar, M. R., & Gunjan, V. K. (2020). Review of machine learning models for credit scoring analysis. *Ingeniería Solidaria*, *16*(1).

[15] Swetha, A., Lakshmi, M. S., & Kumar, M. R. (2022). Chronic Kidney Disease Diagnostic Approaches using Efficient Artificial Intelligence methods. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(1s), 254-261.

[16] Abdar, M., Książek, W., Acharya, U. R., Tan, R. S., Makarenkov, V., & Pławiak, P. (2019). A new machine learning technique for an accurate diagnosis of coronary artery disease. *Computer methods and programs in biomedicine*, *179*, 104992.

[17] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, *36*, 82-93.

[18] Anitha, S., & Sridevi, N. (2019). Heart disease prediction using data mining techniques. *Journal of analysis and Computation*.

[19] Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, P. (2020). Heart disease prediction using machine learning. *INTERNATIONAL JOURNAL OF ENGINEERINGRESEARCH & TECHNOLOGY (IJERT)*, *9*(O4).

[20] D. Shah, S. Patel, and S. Kumar Bharti, Heart Disease Prediction Using Machine Learning Techniques, Springer Nature Singapore Pte Ltd, Berlin, Germany, 2020

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

[22] Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE access*, *8*, 107562-107582.

[23] Pescatello, L. S., Wu, Y., Panza, G. A., Zaleski, A., & Guidry, M. (2021). Development of a novel clinical decision support system for exercise prescription among patients with multiple cardiovascular disease risk factors. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes*, *5*(1), 193-203.

[24] Yavari, A., Rajabzadeh, A., & Abdali-Mohammadi, F. (2021). Profile-based assessment of diseases affective factors using fuzzy association rule mining approach: A case study in heart diseases. *Journal of Biomedical Informatics*, *116*, 103695.

[25] Rubini, P. E., Subasini, C. A., Katharine, A. V., Kumaresan, V., Kumar, S. G., & Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. *Annals of the Romanian Society for Cell Biology*, 904-912.

[26] Drożdż, K., Nabrdalik, K., Kwiendacz, H., Hendel, M., Olejarz, A., Tomasik, A., ... & Lip, G. Y. (2022). Risk factors for cardiovascular disease in patients with metabolic-associated fatty liver disease: a machine learning approach. *Cardiovascular Diabetology*, *21*(1), 240.

[27] Kaggle Cardiovascular Disease Dataset. Available online: https://www.kaggle.com/datasets/sulianova/cardiovascular-diseasedataset (accessed on 1 November 2022)

[28] Kumar, M. R., & Gunjan, V. K. (2020). Review of machine learning models for credit scoring analysis. *Ingeniería Solidaria*, *16*(1).

[29] Ramana, K., Kumar, M. R., Sreenivasulu, K., Gadekallu, T. R., Bhatia, S., Agarwal, P., & Idrees, S. M. (2022). Early prediction of lung cancers using deep saliency capsule and pre-trained deep learning frameworks. *Frontiers in oncology*, *12*, 886739.

[30] Reddy, K. U. K., Shabbiha, S., & Kumar, M. R. (2020). Design of high security smart health care monitoring system using IoT. *Int. J*, *8*.

[31] Swetha, A., Lakshmi, M. S., & Kumar, M. R. (2022). Chronic Kidney Disease Diagnostic Approaches using Efficient Artificial Intelligence methods. *International Journal of Intelligent Systems and Applications in Engineering*, *10*(1s), 254-261.

[32] Srivastava, A., & kumar Singh, A. (2022, April). Heart Disease Prediction using Machine Learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2633-2635). IEEE.

[33] Ms. Mayuri Ingole. (2015). Modified Low Power Binary to Excess Code Converter. International Journal of New Practices in Management and Engineering, 4(03), 06 - 10. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/38

[34] Gabriel Santos, Natural Language Processing for Text Classification in Legal Documents , Machine Learning Applications Conference Proceedings, Vol 2 2022.