EPH – International Journal of Mathematics and Statistics

ISSN (Online): 2208-2190
Volume 10 Issue 01 March 2024

DOI:https://doi.org/10.53555/eijms.v10i1.74

A COMPARATIVE ANALYSIS OF MACHINE LEARNING FOR THE EARLY DIAGNOSIS AND IDENTIFICATION OF CARDIOVASCULAR DISEASES

Purvash Dangi[1*]

[1]Jamnabai Narsee International School (JNIS), purvashdangi@gmail.com*

**Corresponding Author: Purvash Dangi, Jamnabai Narsee International School*
Email: purvashdangi@gmail.com

Abstract
With the advent of advanced digital technologies revolutionizing the healthcare sector through artificial intelligence and machine learning, there is a significant focus on accurate machine learning models to predictchronic heart diseases. The late detection and misdiagnosis of chronic diseases like cardiovascular disease can significantly increase the mortality rate. Therefore, a dependable, accurate, and workable system must identify these illnesses in time for appropriate treatment. Large-scale and sophisticated data processing has been automated using machine learning techniques and algorithms for various medical datasets. Several researchers have recently employed machine-learning techniques to assist the medical community and experts detect heart-related illnesses. The approach assists in predicting chronic diseases like heart failure and effective rehabilitation and timely management. This paper surveys various machine-learning models and compares their accuracy and other parameters in diagnosing heart-related disease. In this paper, we aim to compare various machine-learning models, and a predictive model is proposed for heart disease prediction based on the stacking of various classifiers. This model suggests fostering accurate decision-making. This proposed model will enhance prediction accuracy and eliminate anomalies, thus justifying the selection of the stackingclassifier as the most accurate machine-learning model to predict heart failure about 98%.

Key Words: *Machine learning (ML), Cardio Vascular Diseases (CVDs), Stacking classifier*

I. Introduction

There are multiple predictors like diabetes, blood pressure, age, obesity, and high cholesterol, which increasethe chances of developing heart disease; this highlights the need for strategic techniques fostered through predictive digital machine learning to identify the chances of developing heart disease and severity of the heart disease [1]. With the advent of increasing incidences of stress, graveyard work shift schedules, excessive smoking, and drinking, there is a multitude of CVDs globally [2]. CVDs are increasing at an alarming rate, taking an estimated 17.9 million lives each year as per WHO, raising impetus to the need for advanced predictive ML models to predict their incidence and curb the implications of late diagnosis and prognosis[3].In Asian countries, there is an increasing incidence of CVDs, with half of the patients diagnosed dying within 1-2 years of diagnosis. Late diagnosis and delayed prognosis could result into fatalities, which gives rise to the need for developing accurate and predictive Machine learning models to predict CVDs [4]. Coronary disease and cardiovascular diseases are types of heart disease. In Asian countries, there is a severe lack of awareness regarding CVDs, and the diagnosis is often during the time of death of the patient. To develop a model, we have compared various machine learning to highlight the need for early detection and timely intervention [5]. Traditionally, medical practitioners adopted a curative versus a less preventive approach byscreening patients for heart diseases through blood work, ECG, and angiography. However, with the rising incidence of CVDs, there was this glaring impetus to more predictive recourses [9]. There are multiple invasive techniques to detect coronary heart diseases, such as coronary arteriography (CAG). However, we propose a predictive model using stacking in which the predictions of a collection of classifiers are given as inputs to the next-level learning algorithm [6]. Multiple machine learning models like ANN, Decision tree classifier, ensemble techniques, logistic regression, and support vector machines exist. However, the stackingmodel has a better prediction strategy, with better nonlinear fitting ability and practical utility.

Stacking Ensemble Machine Learning Algorithm (SEMLA) is an effective technique that is not just predictivebut helps accurately manage the worst-case scenarios. This study aims to explore multiple base models like stacking compared to individual models; in Asian countries, stacking models can be highly beneficial, as theywill also help reduce patient screening time. This signifies the quest for a more predictive and robust model like stacking to validate the existing limitations of the existing models, which lack predictive capabilities forearly detection of CVDs[8]—leveraging the predictive accuracy of the stacking algorithm for the betterment of humanity by assisting cardiologist and medical - practitioners to gain better insights in the healthcare sector. The UCI Heart Disease Dataset from the UCI Machine Learning Repository is open to the public and is one of the most used datasets in this research area. The Statlog dataset is also widely used. In the clinical detectionof diseases, such ML models aim to improve accuracy and reduce the total cost of the computation [9]. The proposed classifier outweighs other machine learning models with the highest accuracy of 98% [10].

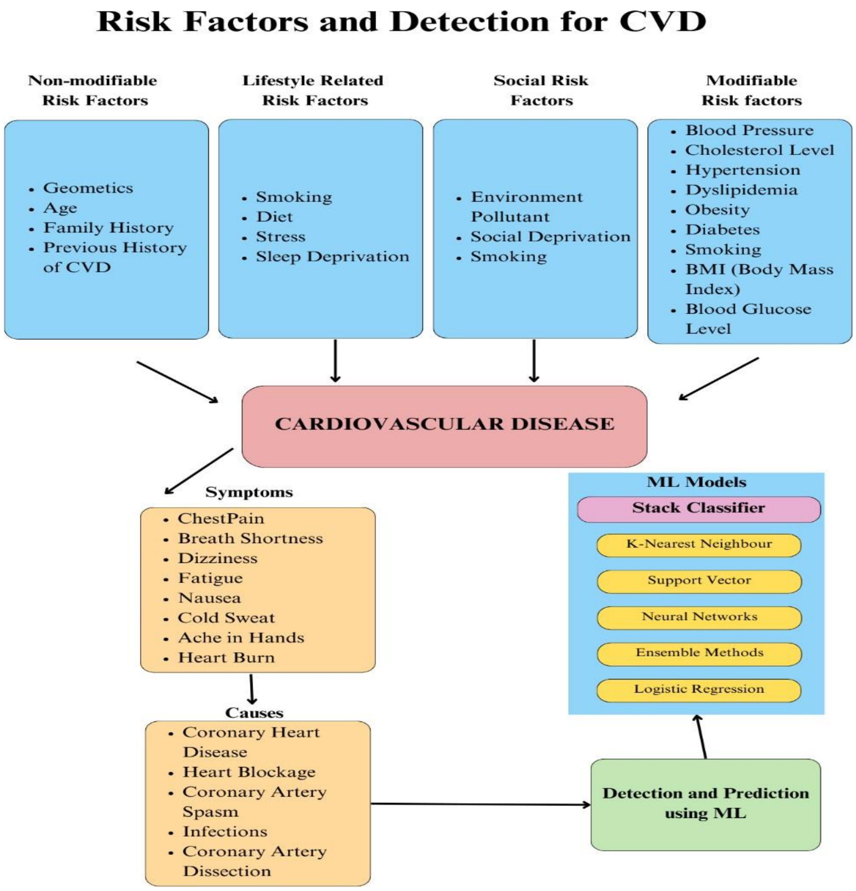

Fig 1. Overview of Cardiovascular Risk

A thorough summary of the machine learning algorithms, detection techniques, and risk factors for cardiovascular disease (CVD) is shown in Figure 1. It lists a number of risk factors, such as genetics and age that cannot be changed, lifestyle choices like diet and smoking, social risk factors like pollution, and changeable risk factors like cholesterol and blood pressure. Furthermore, Figure 1 gives the overview of the typical CVD symptoms such as dizziness, fatigue, breadth shortness, and chest pain etc. Numerous machine learning methods, including Stack Classifier, K-Nearest Neighbour, Support Vector Machine, Neural Networks, Ensemble Methods, and Logistic Regression, are used to predict the heart disease. These ML algorithms can predict the CVD by analysing a variety of risk factors and symptoms.

II. Literature Review

As we read the literature addressing this important health, issue related to heart disease, the combination of machine learning (ML) into the early detection and diagnosis of cardiovascular disease (CVD) has received importance in recent years. Several studies has suggested the innovative frameworks and applications using various machine learning models and approaches, all of which have help us to understand how to predict theCVD.

The creation of several ML models for CVD categorization was highlighted in presentation of "iCardo: A Machine Learning Based Smart Healthcare Framework for Cardiovascular Disease Prediction" in [11]. The authors highlighted the role that feature selection methods like Ridge and LASSO helps in increasing the prediction accuracy. When comparing the effectiveness of numerous ML algorithms like Random Forest, K-nearest neighbours, SVM etc. their results shown that systolic blood pressure and body mass index plays an important variables influencing CVD prognosis. A Systematic Review of Machine Learning and IoT [12] evaluated the usefulness of combining (ML) with the Internet of Things (IoT) for predicting cardiovascular diseases (3D) by using the data present on these devices. IoT devices can track physical indicators like bloodpressure, blood sugar, and heart rate. Additionally, they can keep an eye on environmental cues like the location of a patient.

In the recent paper [13] it compares seven different machine-learning models, such as gradient boosting and convolutional neural networks etc. using two datasets and a variety of pre-processing methods to improve thequality of the data for analysis. In their thorough evaluation of the literature on early cardiac disease forecasting, the authors emphasized the use of machine learning (ML) techniques to find hidden patterns in huge datasets, which can lead to improved diagnostic skills. This study demonstrated how machine learning could automate the creation of models and adapt them to new data, further revolutionizing the healthcare industry. These papers show how machine-learning applications for diagnosing and detecting cardiovascular disease are developing. While addressing the issues of data dependability and integrating technology into clinical practice, they demonstrate an agreement regarding the efficacy of several machine-learning algorithms. It is crucial to assess these approaches closely as the field develops to make sure they are applicable and effective in actual healthcare settings.

A thorough summary of numerous papers on machine learning algorithms for heart disease prediction is provided in Table I. To improve prediction accuracy, each study uses different techniques, like stacking classifiers, random survival forests, and neural networks. The study presents noteworthy results, such as the high accuracy rates of models such as artificial neural networks (97.5%) and hybrid techniques (88.7%) attained. Common research gaps, however, are the requirement for hybrid models, the investigation of dimensionality reduction methods, and the applicability of results to a variety of demographics. There are also lots of limitations; for example, a lot of studies only use small datasets or particular cohorts, which can limit their generalizability.

Table I. Literature Review for CVD using Machine Learning Methods

Paper Reference	Year	Research Methodology	Research Gaps	Findings	Limitations
[25]	2017	MLP neural network+ Backpropagation using Cleveland dataset	Improvement potential exists through hybrid models and dimensionality reduction techniques	accuracy=95% for heart disease prediction.	Small dataset; potential for enhanced performance through dimensionality reduction.
[27]	2017	Random Forest used for analyzing clinical and imaging data	Traditional risk scores may not capture all cardiovascular risk factors.	Identified novel predictors improving prediction accuracy beyond traditional scores.	Limited to the MESA cohort; findings may not generalize to other populations.
[26]	2018	Comparative analysis of various ML techniques using Cleveland Heart Disease dataset	Lack of comprehensive comparison of techniques considering dimensionality reduction and output class variations.	ANNs achieved the highest accuracy (97.5%) among techniques studied.	Relied on a single dataset; potential for overfitting due to excessive dimensionality reduction.
[24]	2019	10-fold cross-validation evaluating three ML algorithms (DT, NB, SVM) on South African Heart Disease dataset	Limitations of conventional techniques in predicting CHD due to data complexity.	NB outperformed SVM and DT in accuracy; deemed most effective algorithm for predicting CHD.	Small number of instances; imbalanced nature of CHD class may hinder ML performance.
[28]	2020	Coefficient and Pearson correlation coefficient utilized followed by enumeration algorithm	Model parameters not optimal; training multiple models is time-consuming.	Achieved accuracy of 95.43%, sensitivity 95.84%, specificity 94.44%.	Time-consuming model training; results may vary based on parameter optimization.
[15]	2020	Utilized Z-Alizadeh dataset for stacking and voting techniques	Identifies gaps in heart disease identification; introduces novel combination of techniques.	Stacking-based model achieved highest accuracy in prediction compared to other techniques.	Limited to Z-Alizadeh Sani dataset; may not generalize across different datasets or populations.
[14]	2020	Machine learning classifiers predicting heart failure patient survival focusing on serum creatinine and ejection fraction as key predictors.	Previous studies lacked focus on critical features from medical records for accurate predictions.	Serum creatinine and ejection fraction are most predictive features for survival outcomes in heart failure patients.	Based on a small dataset from 2015; may not generalize to larger or more diverse populations.
[16]	2021	Experimental study using PIMA Indian diabetes dataset with correlation technique and stacking technique	Limited exploration of advanced stacking techniques compared to traditional methods.	Stacking technique outperformed AdaBoost with an accuracy of 78.2%.	Focused primarily on a single dataset; results may not be applicable across different populations.
[17]	2021	Utilized sklearn, pandas, matplotlib libraries analyzing binary heart disease data from UCI repository	Current works primarily use conventional ML techniques without hybrid approaches for better results.	Hybrid model demonstrated an accuracy level of 88.7% in experimental results.	Limited by reliance on specific libraries; results may vary based on data quality and preprocessing methods.
[18]	2021	Integration of multiple classification methods through ensemble learning	Existing classification methods rely heavily on underlying assumptions that may introduce bias.	Enhanced CHD classification accuracy from 70% to 87.7%; sensitivity: 0.903, specificity: 0.843.	Subjective echocardiogram interpretation could affect prediction accuracy; relies on low-quality image processing.
[19]	2023	Proposed stacking model with performance metrics including accuracy, precision, recall	Limited data instances and attributes hinder ML approaches' effectiveness in real-world scenarios.	Achieved prediction accuracy of 92% with high precision (92.6%) and sensitivity (92.6%).	Limited amount of data may affect model reliability; further validation needed across diverse datasets.
[20]	2024	Comparative analysis of ten ML classifiers with refined feature set (SF-2)	Need for further exploration in feature engineering and explainable AI for better model interpretability.	XGBoost algorithm achieved high accuracy (97.75%) emphasizing critical role of feature selection.	Quality and variety of data directly impact model performance; class imbalance may introduce bias.
[21]	2024	Development of a sensor system for human odor detection using neural networks	Current diagnostic systems lack precision leading to misdiagnosis; insufficient use of neural networks.	System can accurately identify human odor patterns with up to 86% accuracy using multiple sensors.	Requires specialized datasets; practical use in diverse medical environments not fully validated.
[22]	2024	Collection and preprocessing heart disease datasets applying feature selection techniques	Limited exploration of advanced technologies like ML in early detection and management CVDs.	FCBF feature selection improved CVD prediction models achieving up to 78% accuracy with Extra Tree algorithms.	Focuses on traditional diagnostic methods which may not adequately address CVD burden globally.
[23]	2024	SMOTE algorithm used alongside stacked classifier for feature selection	Smaller dataset due to lack of publicly available datasets poses challenges for validation.	G-mean: 87.53%, sensitivity: 93.02%, ROC Score: 87.69%; outperforms existing research methodologies.	Dataset limitations could affect generalizability; reliance on SMOTE raises questions about overfitting risks.

For example, despite employing a stacking-based model to attain a 95% accuracy rate, one study encountereddifficulties with choosing the best parameters and laborious training procedures. Furthermore, some research suggested that the complexity of heart disease risk variables might not be well captured by the predictive models now in use. Overall, this evaluation of the literature highlights the promise of machine learning in enhancing heart disease prognoses while highlighting the need for additional research to fill in the gaps and overcome the limitations that have been found. The results support improved techniques for feature selectionand the incorporation of various datasets to support prediction performance in clinical situations.

Research Objectives

1) To compare traditional ML algorithms vs. Stacking algorithm to predict CVD accurately
2) To highlight the significant features that enhance the stacking algorithm's prediction accuracy.
3) To develop a comprehensive framework of multiple classifiers which enhance the robustness andreliability in predicting CVDs

Hypothesis

1. Stacking algorithms will demonstrate superior predictive accuracy compared to individual classifiersin diagnosing cardiovascular disorders.
2. Incorporating a diverse set of features will significantly enhance the performance of stacking modelsin predicting cardiovascular diseases.

Research Questions

1. How effective are stacking algorithms compared to traditional machine learning methods in diagnosing cardiovascular disorders?
2. What features contribute most significantly to the predictive accuracy of stacking models incardiovascular disease detection?

III. Research Methodology:

Models

In this work, we will be comparing numerous models against each other to identify which one works the best. Each model is designed to identify patterns in the dataset and make a prediction about the newly entered data based on the pattern it has identified. The models we will work with are Logistic regression, Decision Tree classifier, Ensemble techniques, Support Vector Machine, K Nearest Neighbours, and Artificial Neural Network. The performance of each of these models is affected by a few critical parameters known as hyper parameters. We will experimentally select the optimal value for each model for these hyper parameters. This evaluation is presented in the results and discussion section. All the models are implemented using the Tensor Flow framework [29]. As such, only the possible hyper parameter values implemented by this framework are considered.

Logistic regression

Logistic regression is an extension of linear regression [30]· Linear regression is a machine learning technique that predicts a quantitative response Y based on a single predictor X [31]. Logistic regression wasdeveloped by statistician David Cox in 1959 and is based on the principle of probability. Logistic regressionhelps us estimate the probability of Y based on one or more predictor variables. It determines how much theprobability of an event changes based on numerous other events. Logistic regression can only output binaryvalues and is thus suitable for use here as our final output is one of two options (autistic or non-autistic). The main parameters that affect how the model works are the regularization parameters, types of functionsused, and the penalty term. The different penalty terms are l1, l2, and elastic net.

Decision Tree classifier

The algorithm used by the decision tree classifier is comparable to that of a biological dichotomous key. A dichotomous key operates by posing a series of yes/no questions until you finally arrive at a single output after responding to all of the questions [32]. The output of a decision tree classifier is represented by the leaves, which are follow-up questions based on the path taken, the primary node, or main question, is the root. The stem connects each question to the next. In this case, a decision tree makes sense. The questions mentioned in the dataset above can be made into a tree-like structure split into yes or no branches, with eachyes or no branch leading to a different path and a different output. The main parameter, which affects how the model works, is the criterion, used to split. The different criteria are gini, entropy, and log loss.

Ensemble techniques

Ensemble techniques combine several individual techniques to arrive at a final output. This algorithm is inspired by how humans work together to solve a problem. We combine the opinions of several experts to make crucial decisions [33]. Ensemble techniques similarly induce several individual functions into a moreextensive algorithm and produce an output. Specifically, we consider an ensemble of Decision Tree classifiers. This type of ensemble is also known as a random forest. The most critical parameter for this ensemble technique is the number of estimators. There can be any number of estimators.

Support Vector Machine

Support vector machine (SVM) algorithms work by dividing lines that partition the data samples into different regions [34]. These dividing lines are defined using additional artificial data points called supportvectors, hence the name. The model's prediction is consecutively determined by which side of dividing thenew input falls on. In its basic form, SVMs are only usable for datasets with few variables but can be extended by processing the data using a predefined kernel. The type of kernel used is, therefore, an important hyper parameter. The kernels are linear, poly, radial basis function, sigmoid, and precomputed.

K-Nearest Neighbours

The k-nearest neighbours (KNN) algorithm works based on distance [35]. Important parameters affect this model's performance: the function to calculate the distance of the new input from its counterparts and the number of neighbours (data points) already existing. There can be any number of neighbours.

Artificial neural network

The artificial neural network algorithm is inspired by the human brain, in the sense that our brain has severalneurons that finally connect to the brain and a body part [36]· The artificial network similarly has several nodes, which connect layer by layer. There is an input layer where the data is inputted, then the hidden layers where a function is carried out on the values output from the previous layers to modify these values, and finally, the output layer. Nonlinearity is introduced into the model through activation functions. We consider the neural network's architecture the main hyper parameter for our evaluation. As such, we considerthe number of layers in the model and the number of nodes used in each layer.

Stack Classifier

A stacking classifier is an ensemble method in which the output from multiple classifiers is passed as inputto a meta-classifier for the final classification task.

Metrics

In this subsection, we will discuss how we compare the models. The metrics we have chosen to compare the different models each other are accuracy, precision, recall, and F1 score.

Accuracy

Accuracy is a measure of the correctness of the model. The formula to calculate accuracy is

$$Accuracy = \frac{correct\ predictions}{total\ number\ of\ predictions}$$

Correct predictions are the number of instances in which the model predicted the correct class and the total number of predictions is the number of instances in which the model made predictions.

Precision

Precision is a measure that checks the quality of the model's optimistic predictions. The formula tocalculate precision is

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

True positives refer to the number of people the model correctly predicts as having autism. False positivesrefer to the number of people that the model predicts as having autism, but in reality, they do not.

Recall

Recall is similar to precision in that it works. This metric also checks how many of the true cases themodel finds. The formula to calculate recall is

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

False negatives are the frequency of instances where the model incorrectly predicted an individual as not having autism when, in reality, they do.

F1 score

There is usually a trade-off between precision and recall: If we wanted a highly precise model, the model would be designed to predict instances it is highly confident about. This would lead the model to miss positive instances, leading to a lower recall score. However, we want the model to have a high recall score.In that case, it will predict instances with lower confidence as positive, resulting in false positives and a lower precision score. Which of the two approaches is preferred depends on the context of the problem. Wewill touch on this aspect in the discussion. To bypass the trade-off between recall and precision, we also usethe F1 score. The F1 score is a combination of the recall plus the precision. The formula to calculate the F1score is

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

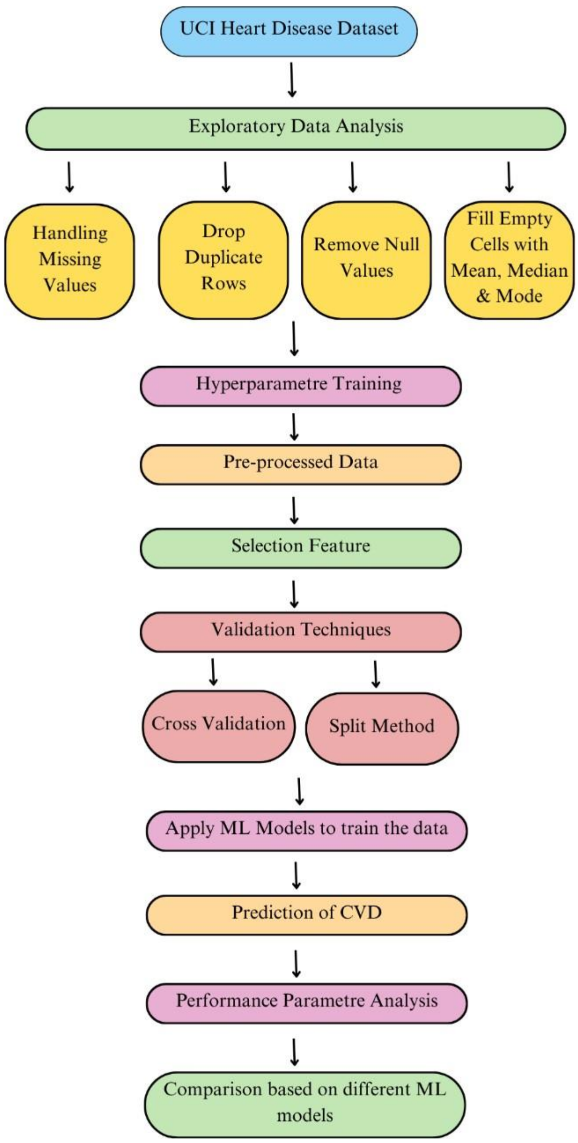

Figure 2: An Overview of the Framework for Explainable Machine Learning to predict the risk of CVD.

The procedure for utilizing machine learning to forecast cardiovascular disease (CVD) is depicted in Figure 2. First, the UCI Heart Disease Dataset is used, and then preliminary data analysis is performed on it. Managing missing values, deleting null values, removing duplicate rows, and populating empty cells with the proper statistical measurements are all part of this process. The data then goes through pre-processing, whichincludes feature selection and hyperparameter tweaking. We then use validation techniques like split method or cross-validation to assess the model's performance. The data is trained using a variety of machine-learningalgorithms, and the generated models are then utilized to predict CVD. In the end, performance parameters are used to analyze the various ML models' performances, facilitating comparison and the identification of thetop-performing model settings.

Key Steps: Analysis of the Diagram: Predicting Cardiovascular Disease Using Machine Learning

1. **Data Acquisition:** The process begins with the UCI Heart Disease Dataset, a patient data collection relevant to cardiovascular health.

2. **Exploratory Data Analysis (EDA):**

o **Data Cleaning:** Handles missing values and duplicates to ensure data quality.

o **Feature Engineering:** Prepares the data for analysis by removing irrelevant features orcreating new ones.

3. **Model Training:**

o **Hyper parameter Tuning:** Optimizes model parameters for better performance.

o **Model Selection:** Chooses appropriate machine learning algorithms (e.g., Stack Classifier, K-Nearest Neighbor, Support Vector Machine, Neural Networks, Ensemble Methods, and Logistic Regression).

o **Training:** Trains the selected models on the pre-processed data to learn patterns andrelationships between features and the target variable (heart disease).

4. **Model Evaluation:**

o **Validation Techniques:** Use techniques like cross-validation to assess model performanceand prevent overfitting.

o **Performance Metrics:** Evaluates model accuracy, precision, recall, F1 score, and otherrelevant metrics.

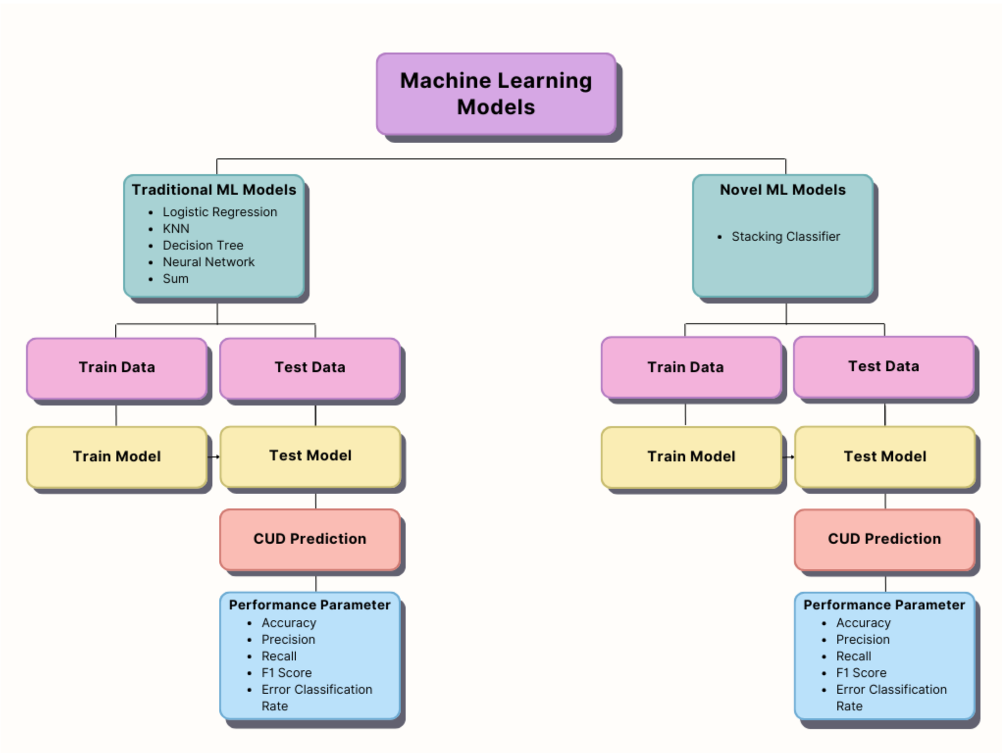

Figure 3: Comparison of Different Machine Learning Models for Predicting CVD

Figure 3 illustrates the process of using machine-learning models to predict chronic obstructive pulmonary disease (CUD). It outlines two categories of models: traditional and novel. Traditional models include Logistic Regression, K-Nearest Neighbours (KNN), Decision Tree, Neural Network, and Sum. Novel models focus onthe Stacking Classifier. The process involves splitting the data into training and testing sets, training the models on the training data, and using the trained models to predict CUD on the test data. Performance parameters such as accuracy, precision, recall, F1-score, and error classification rate are used to evaluate the effectiveness of different models in predicting CUD.

IV. Results and Discussions

In this section, we present our study's empirical evaluation. First, we discuss the impact of the different hyper-parameters for each model. Then, using the best scoring model for each approach, we compare all approaches.

Many previous studies focussed on the need for advanced machine learning algorithm is that are predictive to detect CVDs early, as early identification is a prerequisite requirement to curb fatalities and manage worse- case scenarios [14. It is vital to note that traditional ML models had the issue of class imbalance and only contained a comprehensive dataset containing pertinent indicators related to CVDs and data set containing attributes of patients [15]. Previous Research focussed on multiple traditional ML methods like logistics regression, KNN, decision tree, neural network, and sum, and some research focussed on modern ML algorithms. However, the current Research compares traditional vs. Modern ML algorithms to predict CVDs. Our Research focuses on developing a theoretical framework based on advanced machine learning stacking models to enhance accurate predictions and enable medical practitioners to strategize the line of treatment. Future studies need to focus on creating effective intelligent systems as heart ailments are leading cause of fatalities across the globe and more so in Asian countries where there is a lack of resources and infrastructure.Future studies must incorporate multiple datasets applicable to all population groups. The dataset with the patient's clinical records should consider their disease and lifestyle, as they are vital predecessors for CVDs. Furthermore, more comprehensive studies analysing the feature selection in the context of stacking models will be synergistically better regarding diagnostic outcomes than individual classifiers.

V. Conclusion:

This work presents a comparative study of six machine-learning methods for CVD detection. We compared all six methods based on accuracy, recall, precision, and f1-score. We found that different approaches had thebest performance depending on the metric.

References

1] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learningtechniques. *IEEE access*, *7*, 81542-81554.

2] Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE Access*, *9*, 106575-106588.

3] https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1

4] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, *9*, 19304-19326.

5] Nikam, A., Bhandari, S., Mhaske, A., & Mantri, S. (2020, December). Cardiovascular disease prediction using machine learning models. In *2020 IEEE Pune section international conference (PuneCon)* (pp. 22-27). IEEE.

6] Wang, J., Liu, C., Li, L., Li, W., Yao, L., Li, H., & Zhang, H. (2020). A stacking-based model for non-invasive detectionof coronary heart disease. *IEEE Access*, *8*, 37124-37133.

7] Fatima, R., Kazi, S., Tassaddiq, A., Farhat, N., Naaz, H., & Jabeen, S. (2023). Stacking Ensemble Machine Learning Algorithm with an Application to Heart Disease Prediction. *Contemporary Mathematics*, 905-925.

8] Fatima, R., Kazi, S., Tassaddiq, A., Farhat, N., Naaz, H., & Jabeen, S. (2023). Stacking Ensemble Machine Learning Algorithm with an Application to Heart Disease Prediction. *Contemporary Mathematics*, 905-925.

9] Mohapatra, S., Maneesha, S., Mohanty, S., Patra, P. K., Bhoi, S. K., Sahoo, K. S., & Gandomi, A. H. (2023). A stacking classifiers model for detecting heart irregularities and predicting Cardiovascular Disease. *Healthcare Analytics*, *3*, 100133.

10] A. Kathar, S. Iyer, H. Taneja and Y. Jadhav, "Early Prediction of Heart Failure by Using Stacking Classifier," *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*, Pune, India, 2024,pp. 1-8, doi: 10.1109/MITADTSoCiCon60330.2024.10575846.

[11] N. Sinha, T. Jangid, A. M. Joshi, and S. P. Mohanty, "iCardo: A Machine Learning Based Smart HealthcareFramework for Cardiovascular Disease Prediction," 2022. [PDF]

[12] A. Cuevas-Chávez, Y. Hernández, J. Ortiz-Hernandez, E. Sánchez-Jiménez, et al., "A Systematic Review of Machine Learning and IoT Applied to the Prediction and Monitoring of Cardiovascular Diseases," 2023. ncbi.nlm.nih.gov

[13] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, et al., "Machine Learning-Based Predictive Models for Detectionof Cardiovascular Diseases," 2024. ncbi.nlm.nih.gov

14] Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC medical informatics and decision making, 20, 1-16.

15] Natarajan, K., Vinoth Kumar, V., Mahesh, T. R., Abbas, M., Kathamuthu, N., Mohan, E., & Annand, J. R. (2024). Efficient Heart Disease Classification Through Stacked Ensemble with Optimized Firefly Feature Selection. International Journal of Computational Intelligence Systems, 17(1), 174.

16] Kalagotla, S. K., Gangashetty, S. V., & Giridhar, K. (2021). A novel stacking technique for prediction of diabetes. Computers in Biology and Medicine, 135, 104554.

17] Zhang, J., Zhu, H., Chen, Y., Yang, C., Cheng, H., Li, Y., ... & Wang, F. (2021). Ensemble machine learning approach for screening of coronary heart disease based on echocardiography and risk factors. BMC Medical Informatics and Decision Making, 21(1), 187.

18] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021, January). Heart disease prediction using hybrid machine learning model. In 2021, 6th International Conference on Inventive Computation Technologies (ICICT)(pp. 1329-1333). IEEE.

19] Mohapatra, S., Maneesha, S., Mohanty, S., Patra, P. K., Bhoi, S. K., Sahoo, K. S., & Gandomi, A. H. (2023). A stacking classifiers model for detecting heart irregularities and predicting Cardiovascular Disease. Healthcare Analytics,3, 100133.

20] Ullah, T., Ullah, S. I., Ullah, K., Ishaq, M., Khan, A., Ghadi, Y. Y., & Algarni, A. (2024). Machine Learning-Based Cardiovascular Disease Detection Using Optimal Feature Selection. IEEE Access.

21] Kathar, A., Iyer, S., Taneja, H., & Jadhav, Y. (2024, April). Early Prediction of Heart Failure by Using Stacking Classifier. In 2024 MIT Art, Design and Technology School of Computing International Conference

(MITADTSoCiCon)(pp. 1-8). IEEE.
22] Naeem, A. B., Senapati, B., Bhuva, D., Zaidi, A., Bhuva, A., Sudman, M. S. I., & Ahmed, A. E. (2024). Heart disease detection using feature extraction and artificial neural networks: A sensor-based approach. IEEE Access.

23] El-Sofany, H. F. (2024). Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques. IEEE Access.
24] Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July). Prediction of coronary heart disease using machine learning: an experimental analysis. In Proceedings of the 2019 3rd international conference on deep learning technologies (pp. 51-56).
25] Karayılan, T., & Kılıç, Ö. (2017, October). Prediction of heart disease using neural network. In 2017 international conference on computer science and engineering (UBMK) (pp. 719-723). IEEE.
26] Swain, D., Pani, S. K., & Swain, D. (2018, December). A metaphoric investigation on prediction of heart disease using machine learning. In 2018 International Conference on Advanced Computation and Telecommunication (ICACAT)(pp. 1-6). IEEE.
27] Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., ... & Lima, J. A. (2017). Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circulation Research, 121(9), 1092-1101.
28] Li, W., Yao, L., Li, H., & Zhang, H. (2020). A stacking-based model for non-invasive detection of coronary heart disease. IEEE Access, 8, 37124-37133.
29] Goldsborough, P. A Tour of TensorFlow. *arXiv:1610.01178 [cs]* 2016.
30] James, G.; Witten, D.;Hastie, T.; Tibshirani, R.; Taylor, J. Linear regression.In *An Introduction toStatistical Learning: With Applications in Python*; Springer: New York, 2023; pp. 69–134
31] LaValley, M. P. Logistic Regression. *Circulation* 2008, *117* (18), 2395–2399. https://doi.org/10.1161/circulationaha.106.682658.
32] Patel, H. H.; Prajapati, P. Study and Analysis of Decision Tree Based Classification Algorithms.*InternationalJournal of Computer Sciences and Engineering* 2018, *6* (10), 74–78. https://doi.org/10.26438/ijcse/v6i10.7478.
33] Przemysław, K.; Lughofer, E.; Bogdan, T. Hybrid and Ensemble Methods in Machine Learning. *Journal ofUniversal Computer Science* 2013, *19*, 457–461.
34] Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends. *Neurocomputing* 2020, *408* (1), 189–215. https://doi.org/10.1016/j.neucom.2019.10.118.
35] Zhang, Z. Introduction to Machine Learning: K-Nearest Neighbors. *Annals of Translational Medicine* 2016, *4* (11), 218–218. https://doi.org/10.21037/atm.2016.03.37.
36] Madhiarasan, M.; Louzazni, M. Analysis of Artificial Neural Network: Architecture, Types, and Forecasting Applications. *Journal of Electrical and Computer Engineering*2022, *2022*, 1–23. https://doi.org/10.1155/2022/5416722.