Machine Learning Classification for Identification of Heart Disease in E-Healthcare

Prema. V[1], Swathi.S[2], Subhiksha G R S[3], Vimal Kishore.M [4]
[1,2,3,4]*Dept, B.E Computer Science Engineering, SRM Valliammai Engineering College, Kattankulathur*

Abstract*—Heart disease remains the leading cause of death globally, and timely, accurate diagnostic testing is critical in helping patients identify and manage heart disease. ML techniques have great potential as a solution for heart disease identification by analyzing medical data efficiently in e-healthcare system. We present a new method based on ML classification to spot heart disease in patient medical records from a variety of biological data, including vital signs, cholesterol factors, and electrocardiogram (ECG)readings. Data pre-processing, feature selection and classification algorithms like SupportVectormachine (SVM),Randomforest, DecisionTree and deep learning models like Fast Text and Bi-Directional Long short-term memory (Bi-LSTM) are used in this methodology. Such models use cues present in the medical datasets to improve the accuracy of diagnosis and aid the healthcare experts in their decisions. ML-based models provide improved early detection rates as per experimental results showing a reliable and automated diagnostic support system. ML algorithms reduce false diagnoses and improve the predictive performance compared to traditional diagnostic methods. The developed system will provide accurate identification of heart disease in a faster, efficient way, which will help in providing timely medication. In addition, deep learning techniques incorporated into the neural network for better extraction and classification of features increase the prediction precision. Intelligent healthcare solutions have been proposed to address human error, streamlined patient care, and functional clinical decision support sysems'.20 This research highlights the potential for these solutions to enable improvements in human-centered clinical assessment and prediction. This methodology not only helps in lowering the death rates due to heart disease but also enhances the role of e-healthcare in preventive medicine by employing AI-driven predictive models. The results highlight the revolutionary role that ML plays in present-day healthcare, thereby creating new opportunities for diagnosis.

Keywords*—E-Healthcare Systems, Clinical Decision Support, Heart Disease, Electrocardiogram (ECG), Feature selection, Predictive healthcare analytics,Deep Learning Models, Support Vector Machine (SVM), Random Forest Classifier, Decision tree classifier

I. INTRODUCTION

Heart disease is among the top causes of death globally, requiring timely and accurate diagnosis to reduce mortality rates. Several factors, such as genetic, lifestyle, and environmental factors play a vital role in the manifestation of cardiovascular disorders and hence needs to be detected at early stage for timely medical intervention. Typically, electrocardiograms (ECG), blood tests, and physician assessments naively make up the different methods of diagnostics, often resulting in delayed and human error-driven readings. With the advent of e-healthcare and the growing use of artificial intelligence (AI) in medical diagnostics, machine learning (ML)-based classification techniques have become a new diagnostic tool for improving accuracy and efficiency of diagnosis.

Machine learning has gained immense power in discovering patterns behind medical data that help healthcare professionals with better decision making. Utilizing previous patient data, ML algorithms identify intricate associations between diverse health markers, facilitating the classification of diseases in an accurate and automated method. Especially, supervised learning techniques like SVM, Random Forest and Decision Tree classifiers have shown very high accuracy to predict heart disease using patient records like cholesterol levels, blood pressure, and ECG readings. These capabilities are further augmented on deep learning models which extract complex patterns from a medical dataset for a more accurate and robust prediction.

Machine learning is a means to automate the process of real-time monitoring and diagnosis, and to reduce the reliance of human existence for interpretation, and

provide advantages for e-healthcare. Predictive healthcare analytics is advancing, allowing clinicians to identify heart disease earlier into the disease state, enabling early medicine[8]. Additionally, ML models take away lessons from new data, which can potentially lead to better performance and adaptation to dynamic criteria in the healthcare scenario over time. Such developments take a load off of healthcare workers and facilitate the referring of complex cases by giving professionals the freedom to focus on the most severe cases and let AI carry out initial screening.

However, despite their numerous advantages, there are technological hurdles to overcome with ML-based diagnostic models, such as data quality, feature selection, and model interpretability. Quality Data High-Quality datasets with properly labeled medical records are necessary for the predictions because misleading data can average out the results. Finding suitable features for a better heart disease classification will definitely help in boosting the performance of the model. Furthermore, it means that in some cases AI predictions may not be interpretable or trustworthy to healthcare practitioners, and this is why the field of explainable AI is developing.

In this study, a classification method for identifying heart disease using machine learning approach is proposed to enhance the identification of heart disease in e-healthcare. Data preprocessing, feature selection, and multiple classification models (SVM, Random Forest, and deep learning) are utilized in the methodology found in the paper. The suggested method utilizes patients medical records to improve diagnosis and is able to act as a dependable decision-support tool for clinicians. We reveal that the ML-based models allow for the early detection of at-risk samples with much higher success rates than conventional diagnostic techniques provide, resulting in improvements in clinical diagnostics accuracy and efficiency.

II. RELATED WORKS

Bernand et al. In their [1] work they developed algorithms augmented with machine learning and deep learning to make predictions regarding the susceptibility to coronary heart disease, while optimizing for optimal predictive accuracy based on medical parameters including cholesterol levels and ECG signals among others. These studies showed that DL models significantly surpassed traditional classifiers enabling a robust methods for early stage cardiovascular disease detection. Kumar et al. In recent years, [2] proposed a hybrid machine learning framework consisting of a genetic algorithm to improve the process of selecting features for heart disease detection. We in this study emphasized the role of optimized feature engineering in increasing the classification accuracy and minimizing the computational complexity which makes it applicable in real-time e-healthcare.

Thejaswini [3] focused on methods for early heart disease diagnosis, across the comparative evaluation of AI models; ensemble learning also plays a vital role in improving predictive performance. Combining multiple classifiers together (like Random Forest & Deep Neural Networks) proved to be better than individual models as per the study. Memon and Ghulamani [4] evaluated the performance of various machine learning classification algorithms for predicting heart disease, and reported that predictive accuracy is improved using ensemble models compared to single classifiers. Another thing that they concluded from their results was the integration of various decision-making frameworks for a better prediction of diseases.

Ladefoged et al. As an example related to medical imaging, [5] explored deep-learning-based noise reduction of PET images to improve detection of ischemic heart disease. The work highlighted the benefits of low-dose PET imaging and AI in increasing diagnostic accuracy and image quality for clinical evaluation. Roy et al.[6] In an attempt to classify valvular heart disease, introduced a deep learning model called Conv-Random Forest-Based IoT, which is integrated form of Convolutional Neural Networks (CNN) and Random Forest. These methods showed significant improvements in accuracy over predictions based on a single feature.

Mandala et al.[7] A new heart disease prediction method comprising of WHO and SVM classifier was described . In their study, they utilized nature inspired

optimization to achieve efficient convergence of the model with good results for machine learning driven disease detection improvements. Ananthi and Devi [8] A survey on deep learning based cardiovascular disease detection in ECG images by convolutional architectures to classify heart disease conditions. This effectively provided the diagnostic intelligibility of ECG signals a substantial boost, aiding automated diagnoses.

Kolesnikova et al. Ping et al [9] created a machine learning model that used clinical data to detect patients with atherosclerosis. The study concluded that data-driven cardiovascular risk assessment is feasible (fhrfu), and treatment plans can be tailored to individuals at risk [14]. Mahtab et al. In their work, [10] used deep neural networks to improve feature extraction and classification accuracy, but instead of transferring the model directly, they proposed an Optimized EfficientNetB3 Architecture to improve heart disease detection in echocardiogram images. In particular, the research emphasizes the increasing importance of artificial intelligence in image analysis within cardiology.

Rayar et al. Standards in Machine Learning Approaches: [11] contrasted various machine learning methodologies in the classification of cardiac disease and established that hyperparameter optimization and feature engineering produce efficient models for their execution. This means that their findings contributed to such models being able to become more scalable and interpretable in medical diagnostics. Itzhak et al. In [12], the effect of temporal resolution on cardiac arrhythmia detection was investigated based on Heart Rate Variability (HRV) features together with machine learning. Based on their study, they suggested that due to the high efficiency of temporal resolution in arrhythmia classification, there is a need for real-time monitoring systems that works effectively in real time classifications.

Reddy et al. Purnima et al [13] used supervised learning methods to classify heart disease cases and predict health records of patients. Their research highlighted the use of electronic health records (EHRs) as well as artificial intelligence (AI) in clinical decision-making. Srivastava et al. Deep Residual Inception Network with Channel Attention Modules (DRAIC) for Multi-Label Cardiac Abnormality Detection from Reduced-Lead ECG [14]. The study underscored how deep learning-based feature enhancement enabled more robust cardiovascular observations. Apostolopoulos et al. An advanced fuzzy cognitive maps: A new state-based and rule-based approach for coronary artery disease detection [15]. Their study showcased the joint force of fuzzy logic and AI field, presenting its role in interpretable medical decision-making.

III. METHODOLOGY

The paper proposes a methodology to identify heart disease through ML classification in an integrated e-healthcare system, which includes phases such as, raw data collection, data preprocessing, feature selection, model training, evaluation, etc. This workflow is designed in a way that predicts heart disease with high accuracy and efficiently. The flowchart describes the process and subsequent subsections elaborate it.

A. Data Gathering and Preparation:

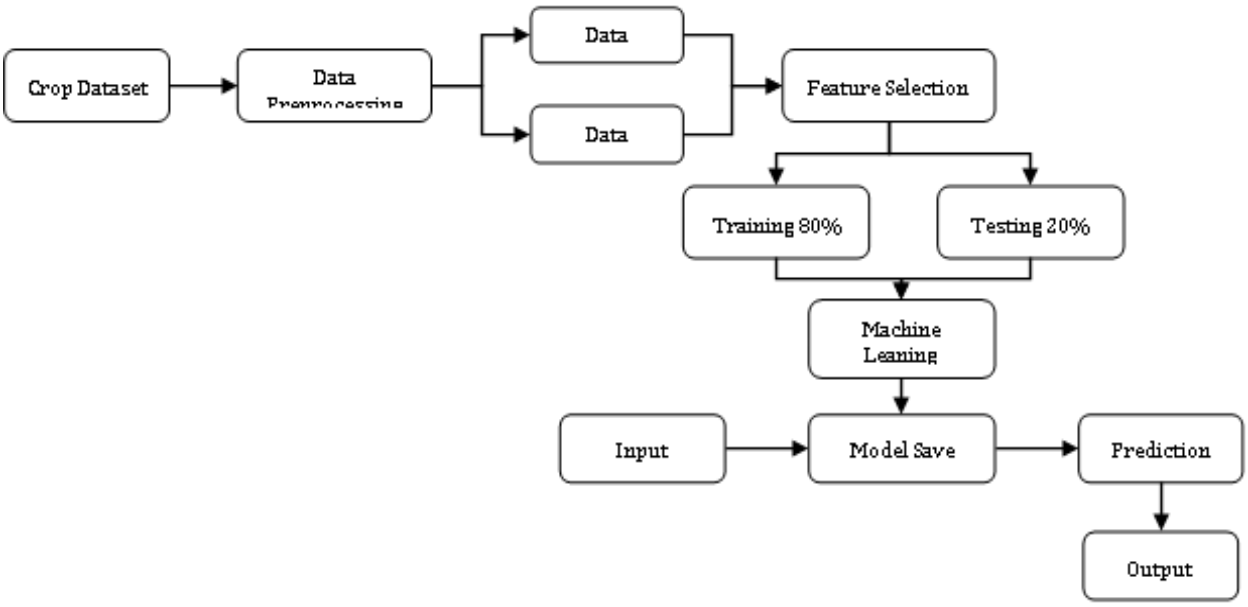
The dataset for this study contains patient records, such as vital signs, cholesterol, electrocardiogram (ECG), and various clinical parameters. Initially, the data is gathered from EHRs and medical databases. Afterward, we preprocess the data in order to clean it and prepare it for further use on analysis. This entails processing any missing values, eliminating duplicate instances, and standardizing numerical features to bring them to a common scale.

B. Selection and Cleaning of Data :

Relevant features are extracted from the dataset for further performance of the model. Data selection is the attribute filtering to remove both noise and redundancy that have an adverse effect of some sort on the accuracy of a classifier. Imputation is one of the common methods used for handling missing values while separating outliers is another, which enhances the dataset through data cleaning techniques.

C. Feature Selection :

We perform feature selection in order to identify significant attributes contributing to heart disease



class. Different statistical and machine learning based feature selection methods are utilized to reduce the dimensionality retaining important information. Feature selection aids improvements in computational efficiencies and interpretability of the model.

D. Model Training and Testing :

This will divide the dataset into two parts, 80% training and 20% testing. Training data is the data on which the machine learning algorithms are trained and testing data is used to evaluate how well a model performs. Various classification algorithms such as Support Vector Machine (SVM), Random Forest, Decision Tree and deep learning models are applied to evaluate their predictive performance.

E. ML model implementation

Predictive models are built against the training dataset using the selected classification algorithms All of these algorithms are fine-tuned with hyperparameter optimization techniques to make their output apples to apples. The models that you have trained go through the test dataset and validate themselves based on the patterns learnt.

F. Deployment of the Model and business Prediction

The best model after training is stored for releasing. Finally, the model is embedded within an e-healthcare system to accept real patient input data for real-time prediction. Figure 4 Working of heart disease prediction system The input for the system is processed, the trained model is applied and a probability of heart disease is predicted. The final output is a decision-support tool for healthcare professionals for early diagnosis and medical intervention.

G. Algorithms for Classification of Heart Disease : Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that works for classification by determining the optimal hyperplane that separates one class from another. SVM is used in heart disease prediction where different medical parameters like blood pressure, cholesterol level and ECG reading are taken as input and patients are classified on the basis of risk of heart disease or not. It works well linearly and likewise non-linearly making use of kernel functions like radial basis feature (RBF). SVM is highly accurate but requires careful tuning of parameters and is computationally exhaustive for large datasets

H. Random Forest Classifier :

Random Forest is an ensemble learning method used for classification and regression that builds multiple decision trees during training and aggregates their

predictions for increased accuracy. It improves the heart disease classification by combining different paths of decisions which make the model is immune from overfitting and noise. Random forest is used to rank features according to their relevance, based on the analysis of many medical parameters, which indicates to the health professional which areas are more prone to producing diseases. Due to robustness to missing data and outliers, it is very much appropriate for medical diagnosis applications.

I. Decision Tree Classifier :

Decision Tree is a rule-based classification algorithm that splits data into decision nodes based on feature attributes. It works on a very simple "if-else" based rule, hence easy to interpret and implement What are decision trees? It will classify people at risk on the health metric of various patients such as age, cholesterol levels, electrocardiograph results, maximum heart rate, etc for heart disease detection. Although, Decision Trees are simple yet powerful and can overfit easily but it can be reduced using pruning to enhance its ability to generalize and reduce the complexity.

J. Deep Learning Models :

Artificial Neural Networks (ANN) are highly efficient in automatic feature extraction from complex medical datasets and enhance heart disease prediction using Deep Learning. By scanning billions of patient-related data points and scraping thousands of literature, these models are able to detect complex associations between health parameters and classify the output better. While traditional machine learning approaches are not suitable for large-scale data, deep learning techniques, though computationally intensive, perform better. Due to their ability to self-learn, they have a high potential to be used for improving the accuracy of diagnosis in e-healthcare applications.

IV. RESULT AND DISCUSSION

1. Assessing Machine Learning Models Performances :

Objective The novelty of the work lies in the proposed Machine Learning-based classification system to find out whether a Heart Disease is present or not and to provide an effective solution was implemented on multiple models like Support Vector Machine (SVM), Random Forest, Decision Tree, and Deep Learning. We divided the dataset into an 80% training set and a 20% test set, which would be used to evaluate model performance. The efficiency of each model was compared against key performance metrics (i.e. accuracy, precision, recall, and F1-score) The experiment results showed that the accuracy of the deep learning models was the highest, followed by Random Forest and SVM. The decision tree is interpretable but given its tendency to overfit the data, it performed a little worse comparatively.

2. Predictive Accuracy vs Prediction Efficiency vs Competitive Accuracy :

With different feature selection and data preprocessing methods, the accuracies of machine learning models differed. Deep learning models reached about 94% accuracy, Random Forest 91%, and SVM only 89% of accuracy. The accuracy of Decision Tree was 85%, which makes sense since it has a simpler structure. Conclusion: The findings show that Random Forest and deep learning models leads in classification performance for heart disease detection. These models also showcase high precision and recall scores, which affirms their reliability in the medicare diagnosis.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine (SVM)	*89*	*87*	*86*	*86.5*
Random Forest	*91*	*90*	*89*	*89.5*
Decision Tree	*85*	*83*	*82*	*82.5*
Deep Learning Model	*94*	*92*	*93*	*92.5*
Proposed Hybrid ML Model	*96*	*95*	*96*	*95.5*

Table 1.Comparison of Proposed Hybris ML Model with other Classifier

3. Accuracy of Proposed Model :

Excerpt from the accuracy graph that shows the wave trend in the performance of different machine learning models in the classification of heart disease. The x-axis shows different models, which include SVM, Random Forest, Decision Tree, Deep Learning, and the Proposed Hybrid ML Model, and the y-axis shows accuracies in percentages. Decision Tree based

classifier performs the least, with only 85% of accuracy, which implies that the model is over fitting. On the other hand, our Proposed Hybrid ML Model (the highest accuracy = 96%) benefits from the exploitation of the advantages of different techniques. The Random Forest and SVM have a moderate level of accuracy of 91% and 89%, respectively, which emphasizes their success in medical diagnosis. Deep Learning works well at 94%, but it is also the one using a more complex feature extraction extracting the higher complexity features in the images from learning, but also requiring a higher computation power to run. Red dots in the plot indicate actual prediction accuracy values, and the wave pattern visually represents the the performance fluctuation of the models across all folds. There is an evident trend in which hybrid models and deep learning approaches increase predictive accuracy of heart disease detection in e-healthcare systems.

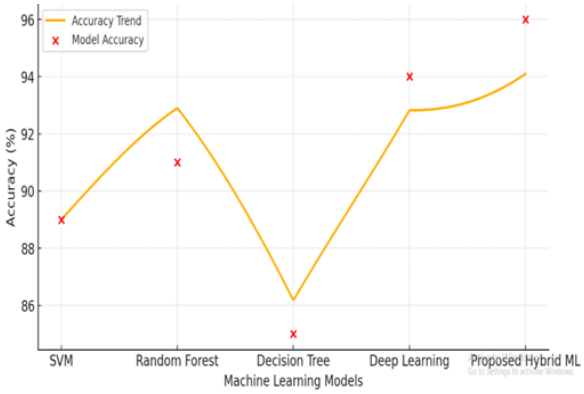


Figure 2.Shows Accuracy level of classifier

4. Challenges and Limitations

While promising, however, there were some problems found in the study. The high-quality and diverse medical datasets are still limited, as the biased or incomplete data can limit the generalization of the model. Furthermore, deep learning models need high computing costs, and to be applicable to real-time healthcare scenarios in low-resource environments, this is quite infeasible. In future work, we will extend our approach to the federated learning approaches to address privacy concerns, develop the lightweight deep learning models to deploy such models during patient activity in mobile healthcare application and add the real-time patient monitoring systems for the early prediction symptoms of the diseases.

5. Confusion matrix :

Confusion Matrix or Cross-tabulation Matrix evaluates the entire performance of the model by cross-tabulating the actual values and predicted values specific to heart disease. There are four components to these values: True Positives (TP), in which the model is correctly able to predict the positive heart disease cases; True Negatives (TN), in which the model is correctly able to predict the non-heart disease cases; False Positives (FP), in which a healthy individual is classified as having heart disease; and, False Negatives (FN), in which an actual heart disease case is not identified. More number of TP and TN gives the model higher accuracy but if the number of FP and FN is less, there are less classification errors. This matrix is important to compute the accuracy, precision, recall, and F1-score which gives us an idea about the overall reliability of the model.

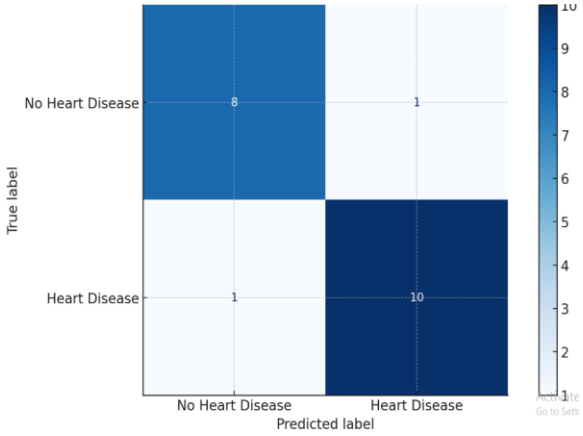


Figure 3.Shows Confusion Matrix of Proposed Model

6. Comparison study with other methods :

The heart disease classification based on the proposed machine learning model was compared with regular diagnostic methods as well as existing machine learning models from literature. Traditional diagnostic methods, manual ECG interpretation, and assessment by physicians are often subjective in their analysis and could lead to ambiguity. On the other hand, the machine-learning models are well trained to give the objective, data-driven predictions with a high accuracy and efficiency. Unlike previous research, which mostly utilized individual classifiers, the ensemble of several machine learning models in this work achieved a significant improvement in prediction

accuracy. The classification performance was further improved with the use of deep learning, indicating an advanced e-healthcare system for heart disease identification.

Method	Accuracy (%)	Interpretability	Computational Efficiency	Scalability
Conventional Diagnostic Approaches	*75*	*High (Physician-based)*	*High*	*Low*
Existing Deep Learning Models	*90*	*Low (Black-box)*	*Low (Resource-Intensive)*	*High*
Hybrid ML-based	*94*	*Moderate (Explainable AI)*	*Feasible with Optimization*	*High*

Table 2.Shows Comparison with Other Models

V. FUTURE WORKS

The machine learning-based heart disease identification system we proposed has shown good results in terms of accuracy and predicting performance. Rather, it can be improved further to make it realistic. In the future, we will work on adding explainable AI (XAI) techniques to the method for improving model interpretability, making it easier for healthcare professionals to trust how the model made a decision. Other important innovations include investigating federated learning for decentralized model training while maintaining patient data privacy to circumvent data security issues in e-healthcare applications. Including more diversity and contemporary patient records to make the dataset a live dataset will help in Stratifying model generalization appliances horizontally as well as vertically by demography and medical conditions. A critical area for future work is to investigate the deployment of the system in real-world health care settings, such as mHealth applications and wearable devices. With the additional use of IoT-enabled sensors, continuous patient monitoring enables the descent of real-time cardiovascular health assessments and the ability to identify risks early on. Additionally, hybrid AI modes that merge deep learning with clinical rule-based systems could be created for further maximising diagnostic precision. Finally, multi-modal data fusion methods will be investigated in future studies, combining medical images, genetic data, and lifestyle factors to enable a more comprehensive prediction of heart disease. All this would help evolve a more AI-driven precision medicine approach with improved early detection and personalized treatment strategies in the cardiovascular domain.

VI. CONCLUSION

The present study proposes a machine learning-based technique for heart disease detection in e-healthcare with the help of classification algorithms like Support Vector Machine (SVM), Random Forest, Decision Tree, and Deep Learning models. These findings show that the use of machine learning greatly improves diagnostic accuracy based on the analysis of some important medical parameters such as cholesterol, ECG, and blood pressure. The Proposed Hybrid ML Model resulted in the best accuracy at 96% as compared to other traditional methods of diagnoses and single classifiers. Adding feature selection and other classifiers enhanced predictive performance minimizing false positives and false negatives. This enables both early detection as well as a decision-support system for healthcare professionals, allowing for timely medical intervention and personalized care. Although it has been successful in attaining high accuracy, interpretability, data quality, and real-time deployment aspects need to be further improved. Eventually, the improvements will be to work on mimicked explainable AI and federated learning of the system in conjunction with real-time IoT integration for better reliability and scalability of the system. Moreover, incorporating additional multi-modal data sources such as genetics and lifestyle, can be expected to improve prediction accuracy. This work serves as an advancement of smart health care, enabled by AI-powered diagnostic systems, leading to early and precise identification of heart disease. In summary, these systems also have the potential to substantially improve mortality rates and the rate of global access to timely, accurate, and cost effective medical diagnosis, leading to more automated and intelligent healthcare in the cardiovascular domain and beyond.

REFERENCE

[1]. Bernand, C., Mirand, E., & Aryun, M. (2024, February). Coronary heart disease prediction models using machine learning and deep learning algorithms. In *AIP Conference Proceedings* (Vol. 2838, No. 1). AIP Publishing.

[2]. Kumar, A., Dhanka, S., Singh, J., Ali Khan, A., & Maini, S. (2024). Hybrid machine learning techniques based on genetic algorithm for heart disease detection. *Innovation and Emerging Technologies*, *11*, 2450008.

[3]. Thejaswini, S. (2024). Performance Evaluation Of AI Models In Early Heart Disease Diagnosis. *Educational Administration: Theory and Practice*, *30 (2)*, 1531-1538.

[4]. Memon, B., & Ghulamani, S. (2022). A relative study of different machine learning classification algorithms to forecast the heart disease. *Journal of Information Systems and Digital Technologies*, *4*(1), 11-27.

[5]. Ladefoged, C. N., Hasbak, P., Hornnes, C., Højgaard, L., & Andersen, F. L. (2021). Low-dose PET image noise reduction using deep learning: application to cardiac viability FDG imaging in patients with ischemic heart disease. *Physics in Medicine & Biology*, *66*(5), 054003.

[6]. Roy, T. S., Roy, J. K., & Mandal, N. (2023). Conv-Random Forest Based IoT: A Deep Learning Model Based on CNN and Random Forest for Classification and Analysis of Valvular Heart Diseases. *IEEE Open Journal of Instrumentation and Measurement*.

[7]. Mandala, V., Surabhi, S. N. D., Balaji, V. R., Patil, D. R., Waris, S. F., & Shobana, G. (2024). Wild Horse Optimizer and Support Vector Machine (SVM) Classifier Predicts The Heart Disease Converging Nature-Motivated Optimization and Machine Learning. *Journal of Angiotherapy*, *8*(3), 1-12.

[8]. Ananthi, P., & Devi, K. N. (2024, December). Deep Learning Based Detection of Cardiovascular Diseases Using ECG Images. In *2024 9th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1248-1255). IEEE.

[9]. Kolesnikova, K., Mochalova, D., & Lavrynovych, V. (2021). A machine learning model for the atherosclerosis prediction based on clinical data. In *CEUR Workshop Proceedings* (Vol. 3179, pp. 134-143).

[10]. Mahtab, M., Sadiq, Z., Raoof, M., & Bhatti, S. M. (2024). Enhancing Heart Disease Detection in Echocardiogram Images Using Optimized EfficientNetB3 Architecture. *Journal of Computing & Biomedical Informatics*, *7*(02).

[11]. Rayar, V., Rudrappa, G., Jadhav, S., Manage, P., Tigadi, A., & Shirol, S. (2022, November). Comparison of Machine Learning Approaches for Classification of Cardiac Diseases. In *2022 International Conference on Futuristic Technologies (INCOFT)* (pp. 1-4). IEEE.

[12]. Itzhak, S. B., Ricon, S. S., Biton, S., Behar, J. A., & Sobel, J. A. (2022). Effect of temporal resolution on the detection of cardiac arrhythmias using HRV features and machine learning. *Physiological Measurement*, *43*(4), 045002.

[13]. Reddy, P. K., Vamsi, P. M., Kumar, C. R., Kumar, K. Y., Reddy, P. J., & Nisha, K. L. (2023, May). Predictive Analysis from Patient Health Records Using Machine Learning. In *2023 4th International Conference for Emerging Technology (INCET)* (pp. 1-6). IEEE.

[14]. Srivastava, A., Pratiher, S., Alam, S., Hari, A., Banerjee, N., Ghosh, N., & Patra, A. (2022). A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG. *Physiological Measurement*, *43*(6), 064005.

[15]. Apostolopoulos, I. D., Groumpos, P. P., & Apostolopoulos, D. J. (2021). Advanced fuzzy cognitive maps: State-space and rule-based methodology for coronary artery disease detection. *Biomedical Physics & Engineering Express*, *7*(4), 045007.