AI – Driven Heart Health: Enabling Early and Accurate Disease Detection

Shravani Suresh Avalkar[1], Dr. Santosh Shivraj Chowhan[2]
[1,2]*Jain School of Sciences, Bengaluru*

***Abstract* — An accurate diagnosis of heart disease is important for timely intervention and proper treatment. This paper reviews heart disease prediction through machine learning approaches using the Cleveland Heart Disease dataset. We analysed the performance of several classification algorithms using SVM, KNN, Logistic Regression, Decision Tree, Random Forest, and XGBoost. The pre-evaluation metrics were accuracy, precision, recall, and F1-score in order to determine the best applicable model for heart disease detection. SVM was found to have the highest classification accuracy of 90.16%. KNN also had high accuracy at 91.80%, performing well in classifying both uninhibited and affected users. Logistic Regression had poorer accuracy at 88.52%, but was still a split choice due to simplicity and practicality. The accuracy of Random Forest and XGBoost was 86.89% and 85.25%, respectively, demonstrating continually accurate predictions. The Decision Tree, however, had the lowest accuracy of all at 72.13%. This shows that underfitting and overgeneralizing were a problem for the model. This research demonstrates the necessity of clinical relevance and tailoring model selection to the dataset. The use of KNN and SVM algorithms makes it possible to improve the accuracy of diagnosis and patient treatment by reducing the rate of errors. Moreover, the work shows the contribution of machine learning to the heart disease predictive analytics improvement. Utilizing advanced classification models enables these practitioners to make decisions based on factual information, resulting in better care of cardiovascular conditions. This is important among medical doctors and researchers working on ways to enhance the use of diagnostic systems and further illustrate the consequences of advanced technological methods like machine learning on healthcare analytics.**

***Index Terms* — Classification models, Diagnostics of Cardiovascular Diseases, Heart Disease Prediction, KNN, Machine Learning, Predictive Analytics, SVM.**

I. INTRODUCTION

World Health Organization, 2022 reports that CVDs or cardiovascular diseases account for around a third of global deaths, making them a prime source of concern for humanity. Complications from coronary artery disease, valvular heart disease, heart failure, or arrhythmias can result in life-threatening conditions such as strokes and heart attacks. In order to ameliorate survival outcome, proactive and predictive measures should be focused on identifying heart disease at its most nascent stages. Accurate diagnosis enables prompt treatment, which may include biological, lifestyle, or operative strategies that mitigate the risk factors. Advanced technologies like electrocardiography, echocardiograms, and stress tests are effective but also costly and require qualified professionals who might not be available in every case. These methods might also overlook the disease in its earlier phases.[1][2]

As a way of addressing these concerns, machine learning (ML) technologies offer a new approach to predicting health issues in cardiology that is both exciting and useful. ML model can process and evaluate substantial amounts of systematic and unsystematic medical information, discern intricate relationships, and make accurate disease predictions. Using structured clinical datasets and physiological signals like ECG recordings, ML models can ameliorate disease identification, help in risk evaluation, and monitor patients. Numerous ML approaches have shown considerable promise for classification and diagnosis of heart disease, such as Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, and deep learning models like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks.

I. Cleveland Heart Disease Dataset:

The Cleveland Heart Disease Dataset, which comes from the UCI Machine Learning Repository, is one of the most used datasets in research on heart disease prediction. Age, gender, kind of chest pain, resting blood pressure, cholesterol, fasting blood sugar, and electrocardiogram readings are among the 14

important characteristics found in the 303 patient records in this dataset. From 0 (no disease) to 4 (severe heart disease), it divides the severity of heart disease into five stages. It has established itself as a standard dataset for training and assessing machine learning models in cardiology research because of its structured form and well-labeled features [4].

The Cleveland dataset has been used to evaluate the effectiveness of several machine learning algorithms in the classification of cardiac disease. According to recent research, SVM has the best classification accuracy (90.16%), closely followed by XGBoost (85.25%), Random Forest (86.89%), KNN (91.80%), and Logistic Regression (88.52%). At 72.13%, the Decision Tree model, on the other hand, showed the lowest accuracy, suggesting possible overfitting and a restricted capacity for generalization [5]. These findings emphasize how crucial it is to choose reliable machine learning models for heart disease prediction that strike a compromise between interpretability, accuracy, and computational economy.

MIT-BIH Dataset on Arrhythmia:

Physiological signals, like ECG records, offer additional insights into cardiac function and illness detection in addition to structured clinical data. A popular tool for identifying cardiac arrhythmias and other cardiovascular anomalies is the MIT-BIH Arrhythmia Dataset, which was created by Beth Israel Hospital and the Massachusetts Institute of Technology. With signals captured at 360 Hz, this dataset comprises 48 half-hour ECG recordings from 47 people, offering a high-resolution depiction of cardiac activity [6]. ECG-based cardiac disease detection has made substantial use of deep learning models, especially LSTM networks and Gated Recurrent Units (GRU). Using the MIT-BIH dataset, a comparison of deep learning models showed that GRU had the best predictive accuracy ($R2 = 0.9965$), closely followed by LSTM ($R2 = 0.9964$). The efficacy of both models in identifying long-term dependencies in ECG sequences was demonstrated by their significant outperformance of conventional Recurrent Neural Networks (RNNs), which had a poorer predictive accuracy ($R2 = 0.9502$) [7]. These results highlight how deep learning models can be used to analyze sequential medical data and increase the accuracy of diagnoses.

Significance of Machine Learning in Early Diagnosis:

Machine Learning's Importance for Early Diagnosis Compared to conventional diagnostic techniques, the use of machine learning in heart disease prediction offers a number of significant benefits. First, by spotting complex patterns and relationships in patient data that traditional statistical methods can miss, machine learning models improve accuracy. Second, automated screening is made possible by AI-driven predictive models, which makes it possible to establish extensive screening programs, particularly in distant places with limited access to healthcare. Third, by lowering false positives and false negatives, ML models enhance diagnostic accuracy and guarantee that patients who are at risk receive prompt medical attention. Furthermore, ML models offer affordable and scalable diagnostic assistance, which eases the strain on the healthcare system and promotes more accurate disease prediction. Finally, the use of ML enables personalized medicine, where models can assess individual risk factors and recommend tailored treatment plans, leading to better patient outcomes and optimized healthcare resource allocation.

Research Objectives and Contributions:

It is more important than ever to have ML-based predictive models that are precise, scalable, and interpretable due to the rising prevalence of cardiovascular illnesses. Using the Cleveland and MIT-BIH datasets, the study compares the effectiveness of many machine learning models in the categorization of heart illness, including SVM, Random Forest, Logistic Regression, KNN, LSTM, and GRU. The goal of this study is to determine the best machine learning methods for practical clinical applications by assessing classification metrics like accuracy, precision, recall, and F1-score. This research explores the impact of ECG-based deep learning models on predictive accuracy, highlighting the advantages of time-series-based neural networks for cardiac monitoring. By integrating structured clinical data with physiological signals, this study contributes to advancing AI-driven cardiology and fostering data-driven decision-making in cardiovascular healthcare. The findings serve as a valuable resource for medical professionals, researchers, and policymakers in adopting AI-powered solutions to improve early heart disease detection, optimize treatment strategies, and enhance patientcare.

II. LITERATURE REVIEW

Detrano et al. [1] focused on probability algorithm for diagnosis of coronary artery disease. The methods in this paper includes the screening of datasets from clinical and noninvasive test results of 303 patients undergoing angiography at the Cleveland Clinic in Cleveland, Ohio. The methods in this paper included the segments of Reference group for derivation of the probability model, Clinical and test variables, Derivation of the algorithm, Testing the algorithm, Bayesian method, Test group data. The limitations in the research manuscript is focusing on single dataset from the Cleveland discriminant function was derived from a population referred for angiography, its best performance is expected in such a population. In contrast, Bhatia et al. [2] presented a decision support system for heart disease classification based on support vector machine (SVM) and integer-coded genetic algorithm (GA) using multi class classification approach. The methods in this paper includes the datasets from Cleveland heart disease database with 303 cases and 13 diagnostic features per case used. Yilmaz and Yağin [3] focused on study compares the performance of machine learning methods in predicting coronary heart disease. The heart disease dataset used in this study was obtained from the IEEEDataPort database. The dataset was created by combining Cleveland, Hungarian, Switzerland, Statlog (Heart) Data Set, and Long Beach VA datasets. Combining was performed using 11 covariates from these 5 heart disease datasets. The methods in this paper included the segments of Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF). The RF algorithm is a promising approach for predicting coronary artery heart disease and could be recommended for future prediction model development. Omar et al. [4] focused on a cardiovascular disease using machine learning. The methods in this paper includes the dataset which was used for analysis are "Framingham" obtained from Kaggle. Heart disease dataset with 14 features is obtained from UCI Machine Learning Repository. The methods in this paper included the segments of random forest, logistic regression, naive bayes algorithm, support vector machines. The relationship between diabetes-related attributes and heart disease risk is also considered in the model. Miao and Miao [5] explored on deep neural networks for diagnosis of coronary artery disease. Used in this research, the clinical heart disease data were from 303 patients at the Cleveland Clinic Foundation (CCF) located in Cleveland, Ohio in the United States. The dataset was obtained from the Heart Disease Database made available in the UCI Machine Learning Repository. The methods in this paper included the segments of heart disease data, deep neural network system and architecture, deep neural network classification model, deep neural network prediction model, evaluation methods of deep neural network model. The model lacks temporal data, crucial for predicting heart disease progression. Incorporating time-series data and exploring RNNs or LSTMs could enhance predictive capabilities. Chang et al. [6] developed an AI-based heart disease detection system using machine learning algorithms. Phases of application development include database collection, logistic regression, and dataset evaluation. Data cleaning involves preparing data for analysis by removing false, corrupted, or redundant data. Comparison of models based on precision, specificity, f-measure, accuracy, and sensitivity. Random forest classification provides best accuracy score. Application of machine learning algorithms for heart disease detection. Shukur and Mijwil [7] investigated heart disease diagnosis using machine learning techniques.
The methods in this paper includes the screening of datasets from Cleveland Clinic dataset from University of California Irvine machine learning (UCL) repository and Kaggle platform. The methods in this paper included the segments of Reference group for artificial intelligence, deep learning, machine learning. The limitations in the research manuscript is focusing on specific datasets, and the effectiveness of techniques might differ with larger or more diverse datasets. Supani et al. [8] focused on classification of coronary artery disease using machine learning approach. The methods outlined in this paper involve the evaluation of datasets g the coronary heart disease (CAD) dataset originating from the UCI and Kaggle dataset locations with addresses: (https://www.kaggle.com/datasets/tanyachi99/zalizadeh-sani-dataset-2csv).The data taken consisted of 303 data points on coronary heart disease (CAD) and healthy patients, where each data point had 55 variables, consisting of 54 features used in the test dataset as independent variables. The methods in this paper included the algorithms of Naive Bayes, XGBoost, KNN, metrics evaluation. The research manuscript addresses several limitations, including the study did not implement feature selection, which could potentially improve model performance,

performance comparisons with other studies may be affected by differences in dataset size and features (e.g., 55 features in this study vs. 14 features in others), only 3-fold cross-validation was used, which might not be as robust as higher-fold cross-validation methods. Also, the XGBoost model performed well overall but showed lower precision, indicating it may struggle with correctly identifying positive cases, the study used a limited set of metrics; additional metrics might provide a more comprehensive performance evaluation, and results are based on specific datasets and experimental conditions, which may limit their applicability to other contexts. Hu et al. [9] the study published in Heliyon explores the use of deep learning-based coronary artery calcium scoring (DL-CACS) to predict coronary artery disease (CAD) in patients with type 2 diabetes mellitus (T2DM). Analysing data from 469 T2DM patients, the research shows that DL-CACS is significantly effective in predicting both obstructive CAD (≥50% artery stenosis) and hemodynamically significant CAD (CT-derived fractional flow reserve ≤0.8). With an AUC of 0.753 for obstructive CAD and 0.769 for hemodynamically significant CAD, DL-CACS demonstrates high accuracy and efficiency compared to manual methods, making it a valuable tool for risk management and treatment decisions in T2DM patients. Bharti et al. [10] focused on prediction of heart disease using a combination of machine learning and deep learning. The dataset used for this research purpose was the Public Health Dataset and it is dating from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 attributes, including the predicted attribute, but all published experiments refer to using a subset of 14 of them. The methods in this paper included the segments of Reference group for description of the data, preprocessing of the dataset, checking the distribution of the data, checking the skewness of the data, checking stats of the Normal distribution, feature selection, checking duplicate values in the data, machine learning classifiers propose, deep learning pseudocode and proposed, evaluation process used, multimedia uses. The limitations in the research manuscript is focusing on the dataset's small size restricts the performance and generalizability of both machine learning (ML) and deep learning (DL) models. There is a risk of overfitting due to inadequate data normalization, and while computational time was reduced, further optimization is needed for efficient model evaluation and deployment. Outlier handling with Isolation Forest may not be sufficient for all data types, and the limited 13 features might constrain model performance. And integrating ML and DL models with multimedia for enhanced patient and doctor interactions was not explored. Lastly, comparisons with other research are limited by differences in datasets, affecting benchmark accuracy. Ramesh et al. [11] focused on predictive analysis of heart diseases with machine learning approaches. The methods in this paper includes the screening of datasets from UCI dataset with 303 rows and 76 properties. Approximately 14 of these 76 properties are selected for testing, which is necessary to validate the performances of different methods. The methods in this paper included the segments of Reference group for decision tree classifier, naive bayes, random forest, KNN, SVM method, logistic regression. The limitations in the research manuscript is focusing on the relatively small dataset of 303 patients may restrict the generalizability of the findings, while the use of only 14 out of 76 features could miss relevant information that might improve model accuracy. Class imbalance, with 54% of patients having heart disease versus 45% without, may bias results. The study's reliance on a fixed set of machine learning models (SVM, KNN, Decision Tree, Random Forest, Naïve Bayes) excludes potential benefits from other advanced models. There is also a risk of overfitting, despite employing k-fold cross-validation. The Isolation Forest method for feature extraction may not fully address anomalies, and the lack of external validation with independent datasets limits the robustness of the results. Padmaja, and Srinidhi[13] focused on Early and Accurate Prediction of Heart Disease Using Machine Learning Model. The methods in this paper includes the screening of datasets from clinical test results of 303 patients undergoing angiography at the Cleveland Clinic in Cleveland. The model is developed using different classification algorithms which include Logistic Regression, Random Forest, Support vector machine, Gaussian Naïve Bayes, Gradient boosting, K-nearest neighbours, Multinomial Naïve bayes and Decision trees. The limitations in the research manuscript include reliance on a single Cleveland dataset, lack of real-world testing, use of limited features, a small sample size, difficulty in understanding model decisions, poor generalization to different populations,

exclusion of real-time data, ignoring other health conditions, scalability issues, and an unclear future expansion plan. L. J. Muhammad and Ibrahem Al Shourbaji [14] focused on Machine Learning Predictive Models for Coronary Artery Disease. The methods in this paper include screening datasets from the medical diagnostic records of coronary artery disease patients at Murtala Mohammed General Hospital and Abdullahi Wase General Hospital in Kano State, Nigeria, collected between 2003 and 2017. The model is developed using different classification algorithms, including Logistic Regression, Random Forest, Support Vector Machine, Gaussian Naïve Bayes, Gradient Boosting, K-Nearest Neighbors, Multinomial Naïve Bayes, and Decision Trees. The limitations in the research manuscript include reliance on a single dataset from two hospitals in Kano State, limited demographic diversity, lack of real-time data or continuous patient monitoring, a small sample size, evaluation using a limited set of machine learning algorithms, neglect of comorbidities, limited interpretability for clinical decision-making, untested real-world clinical effectiveness, scalability concerns, and an unclear plan for future expansion. S Arooj, S Rehman [15] focused on A deep convolutional neural network for the early detection of heart disease. The methods in this study include utilizing a deep learning approach with image classification for heart disease detection. The model is developed using a deep convolutional neural network (DCNN) on the public UCI heart disease dataset, which consists of 1050 patients and 14 attributes. By gathering directly obtainable features from the dataset, these features are used as input for the DCNN to classify instances into healthy or cardiac disease classes. The model is evaluated using performance metrics such as accuracy, precision, recall, and the F1 measure, achieving a validation accuracy of 91.7%. The limitations in this research include reliance on a single dataset (UCI heart disease dataset), lack of image data for more precise classification, limited dataset size (1050 patients), absence of real-time data or continuous patient monitoring, and limited generalizability of the model to other heart disease types beyond the dataset used. Safial Islam Ayon [16] focused on Coronary Artery Heart Disease Prediction: A comparative study of computational intelligence techniques. The methods in this study include comparing a number of computational intelligence techniques for predicting coronary artery heart disease. Seven techniques were evaluated: Logistic Regression, Support Vector Machine, Deep Neural Network, Decision Tree, Naïve Bayes, Random Forest, and K-Nearest Neighbor. The techniques were tested using the Statlog and Cleveland heart disease datasets from the UCI machine learning repository. Performance metrics such as accuracy, sensitivity, and precision were used to evaluate the models. The study found that the highest accuracy of 98.15% was achieved by the DNN, with a sensitivity of 98.67% and precision of 98.01%. The results were compared to other studies on heart disease prediction and showed improved performance. The limitations in this research include the reliance on two specific datasets, the lack of consideration for other heart disease datasets, and the possibility that the findings may not be generalizable to real-world clinical settings without further testing. Setiawan, N.A. [17] focused on Diagnosis of coronary artery disease using artificial intelligence-based decision support system. The methods in this paper includes the screening of datasets from University California Irvine (UCI). The methods in this paper included the segments of decision support system, fuzzy, rough set theory. The limitations in the research manuscript is focusing on limited number of data sets and the need for expert validation, which may limit its generalizability. Nazlı, B., Gültepe, Y. and Altural, H., [18] focused on Classification of coronary artery disease using different machine learning algorithms. The methods in this paper includes the screening of datasets from the UCI machine learning pool are CAD. Anderies, A. [19] focused Prediction of heart disease UCI dataset using machine learning algorithms. The study uses the UCI Cleveland heart disease dataset from the UCI machine learning repository, consisting of 297 instances with 14 attributes, including a dependent variable 'Diagnosis' for predicting heart disease, with the remaining attributes serving as independent variables. Garavand, A., Salehnasab, C [20] focused on Efficient model for coronary artery disease diagnosis: A comparative study machine learning algorithm. The methods in this study include the comparison of different machine learning algorithms for early coronary artery disease (CAD) diagnosis using clinical examination features from 303 records with 26 target features, incorporating algorithms such as Multilayer Perceptron, Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN), and Naive Bayes

(NB), followed by evaluation of performance metrics including AUC, F- measure, and ROC to identify the most effective algorithms. Cenitta, D., Arjunan, R.V. and Prema, K.V. [21] focused on ischemic heart disease multiple imputation technique using machine learning algorithm. The methods in this paper includes the screening of datasets from University of California Irvine (UCI). The methods in this paper included the segments of reference group for Fuzzy rough set, machine learning, multiple imputation techniques, random forest. The limitations in this research focus on the use of a single dataset from the UCI Heart Disease repository, and the proposed imputation technique was developed using Random Forest classification, making its best performance expected within similar benchmark datasets rather than diverse real-world clinical settings. Bilal, M. [22] focused on Using machine leaning models for the prediction of coronary arteries disease. The methods in this paper includes the screening of dataset UCI repository. The methods in this paper included the segments of Reference group for machine learning model, machine learning classifiers, ensemble model, decision tree, logistic regression, XGB, KNN, Naïve bayes. The limitations in the research manuscript is focusing only on the UCI dataset without selecting the most important features, which may affect accuracy. It focused on numeric data and three classifiers, limiting its use for other data types and advanced models. Ambrish, G., Ganesh, B. [23] focused on Logistic regression technique for prediction of cardiovascular disease. The methods in this paper includes screening of dataset from UCI ML repository. It contains 13 features and 303 records. The methods in this paper included the segments of reference group for logistic regression, feature selection and machine learning. The study used only the UCI dataset, limiting its generalizability to other real-world data. It focused solely on Logistic Regression, without comparing it to other advanced models. Future work should test multiple datasets for better validation. Omkari, D.Y. and Shaik, K. [24] focused on an integrated two – layered voting (TLV) framework for coronary artery disease prediction using machine learning classifiers. The methods in this paper includes screening of two datasets from Kaggle's heart disease dataset of over 70,000 records and UCI's heart disease dataset of 1025 records. This methods in this paper included the segments of reference group for ANOVA f-test, Chi-squared test, decision tree, random forest, support vector classifier, multi-layer perceptron. Prusty, S., Patnaik, S. and Dash, S.K. [25] focused on Comparative analysis and prediction of coronary heart disease. The methods in this paper includes the screening of datasets from the UCI repository. It contains 303 individual patient records and 14 dependent features. The methods in this paper included the segments of reference group for Machine learning techniques such as Naïve bayes, decision tree, random forest, K-nearest neighbor, logistic regression, support vector machine, light gradient boosting machine (LGBM) as well as Deep learning techniques such as Neural network, artificial neural network. Pasha, S.M. and Ankalaki, S. [26] focused on Diabetes and heart disease prediction using machine learning algorithms. The methods in this paper include the screening of datasets from a nationalized organization, consisting of clinical and demographic factors for predicting sugar illness and heart disease, along with heart disease data from 1026 individuals with 14 key features. The methods in this paper include the use of Artificial Neural Networks (ANN), Extreme Learning Machine (ELM), Principal Component Analysis (PCA), Least Absolute Shrinkage and Selection Operator (LASSO), Ensemble Learning, and Support Vector Machine (SVM) to identify these diseases with high accuracy. The evaluation is conducted on Diabetes and Heart Disease datasets to assess the predictive performance of these models. Nagavelli, U., Samanta, D. and Chakraborty, P. [27] focused on Machine learning technology-based heart disease detection models. The methods in this paper include the use of Naïve Bayes with a weighted approach for prediction, an automatic ischemic heart disease localization method based on frequency and time domain features, and an improved SVM with duality optimization for heart failure identification. An effective heart disease prediction model (HDPM) is developed using DBSCAN for outlier detection, SMOTE-ENN for data balancing, and XGBoost for classification. The study aims to provide clinicians with a decision-support tool to diagnose heart disease early, improving treatment effectiveness and patient outcomes. The limitations of this research include reliance on specific datasets and potential performance variations when applied to broader clinical populations. Nadeem, M.W., Goh, H.G., Khan, M.A. [29] focused on Fusion-Based Machine Learning Architecture for Heart Disease Prediction. The methods in this paper include the screening of

datasets from publicly available heart disease records. The first dataset, the "Heart Disease Dataset 2019," consists of 1,025 samples with 13 features and is commonly used for heart disease diagnosis research. The second dataset, the "Cardiovascular Disease Dataset 2019," contains 70,000 patient records with 11 unique features. Both datasets, available on the Kaggle repository, are used to train supervised machine learning models for heart disease prediction. The methods in this paper included the segments of reference group for machine learning, support vector machine, fuzzy logic, fusion. Budholiya, K., Shrivastava, S.K. and Sharma, V. [30] focused on an optimized XGBoost based diagnostic system for effective prediction of heart disease. The methods in this paper include the utilization of the Cleveland Heart Disease dataset obtained from the University of California, Irvine (UCI) Machine Learning Repository. The dataset comprises 303 patient records, of which 6 have missing class values. Although the dataset originally contains 76 attributes, previous studies have identified 13 key features as the most effective in detecting heart disease. The methods in this paper included the segments of reference group for XGBoost, Bayesian optimization, categorical feature encoding and prediction. The limitations in the research manuscript is focusing on a specific dataset, so its performance on other datasets is uncertain. While Bayesian optimization helps improve accuracy, its effectiveness on different datasets is not tested. The study only compares tree-based models and does not include deep learning methods. It also does not analyze how selecting different features might impact accuracy. More real-world testing is needed to confirm if the model works well in medical practice. And class imbalance in the dataset could affect the model's reliability. Ayatollahi, H., Gholamhosseini, L. and Salehi, M. [31] focused on Predicting coronary artery disease: a comparison between two data mining algorithms. The methods in this paper include data collection using a checklist based on the Cleveland heart disease dataset from the UCI repository. The checklist consists of 25 variables related to patient demographics, medical history, and test results for predicting coronary artery disease (CAD). The methods in this paper include the segments of Reference group for Data mining algorithms, Artificial neural network (ANN), Support vector machine (SVM). Sadr, H., Salari, A. [32] focused on cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models. The methods in this paper include the screening of two public heart disease classification datasets containing 70,000 and 1,190 records, along with a locally collected dataset of 600 records. The methods in this paper include segments for a reference group, machine learning models, deep learning techniques, and a combinational model for heart disease classification. Sayadi, M. [34] focused on A machine learning model for detection of coronary artery disease using noninvasive clinical parameters. The methods in this paper include the screening of the Z-Alizadeh Sani dataset, which contains 54 features and 303 records of patients from the cardiovascular center of Shahid Rajaei Hospital, Tehran, with two main classes: CAD (216 cases) and normal (87 cases). The methods in this paper included the application of six machine learning techniques, including decision tree, deep learning, logistic regression, random forest, support vector machine (SVM), and XGBoost, based on a semi-random-partitioning framework. The limitations in the research manuscript are focused on using a single dataset from the Z-Alizadeh Sani database, which may limit the generalizability of the findings to other populations or healthcare settings. Shorewala, V. [35] focused on Early detection of coronary heart disease using ensemble techniques. The methods in this paper included the screening of datasets from the 'Cardiovascular Disease Dataset,' containing 70,000 records of patient data for coronary heart disease. The methods included the application of predictive techniques such as K-Nearest Neighbors, Binary Logistic Classification, and Naive Bayes, and the evaluation of ensemble models like bagging, boosting, and stacking. The performance of these models was validated using data-analytic techniques and K-Folds cross-validation. The limitations in the research manuscript focus on using a single dataset from the Cardiovascular Disease Dataset, which may limit the applicability of the findings to different populations or healthcare systems. Bora, N., Gutta, S. [36] focused on Using machine learning to predict heart disease. The methods in this paper include the screening of datasets from the UCI Machine Learning repository, containing 303 record instances with 14 attributes, and from Kaggle, containing 1190 patient records with 11 features. The methods involve the application of various machine learning algorithms such as Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), K-Nearest

Neighbor (KNN), Random Forest, and Extreme Gradient Boost to predict the likelihood of heart disease. The limitations in the research manuscript focus on using two separate datasets, with the second dataset being a combination of five popular datasets. The findings may not be directly applicable to other populations or clinical settings due to the nature of the datasets used. Rahimi, K., Bennett, D. [37] focused on Risk prediction in patients with heart failure: a systematic review and analysis. The methods in this paper include the review of literature through a search of MEDLINE and EMBASE databases from January 1995 to March 2013, followed by hand searches of the reference lists of the retrieved studies. The methods also included ranking individual risk predictors based on their strength of association with outcomes such as death and hospitalization in heart failure patients. The limitations in the research manuscript include the wide variation in clinical settings, population characteristics, sample sizes, and variables used for model development. These factors were not significantly associated with the discriminatory performance of the models, which may limit the generalizability of the findings across different populations and clinical environments. Zulkiflee, N.F. and Rusiman, M.S. [38] focused on Heart Disease Prediction Using Logistic Regression. The methods in this paper include the analysis of a dataset containing 270 patient records to determine significant variables related to the presence or absence of heart disease. The methods applied include Binary Logistic Regression (BLR) models, BLR models with the Least Quartile Difference (LQD) method, and BLR models with the Median Absolute Deviation (MAD) method. The limitations in the research manuscript include focusing on a single dataset, which may limit the generalizability of the results. The best performance of the models is expected in populations with similar characteristics to those in the provided dataset. Khanna, D., Sahu, R., Baths, V. and Deshpande, B. [39] focused on Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease. The methods in this paper include the screening and analysis of the publicly available Cleveland Heart Disease Dataset, which consists of 303 patient records. The classification techniques applied in this study include logistic regression, support vector machines with a linear kernel, and other machine learning models. The evaluation of these models was conducted using F1 score and ROC curves to determine their effectiveness in predicting heart disease. The limitations in the research manuscript include the reliance on a single dataset from the Cleveland Clinic, which may restrict the generalizability of the findings. The models' best performance is expected in populations with similar characteristics to those present in the dataset. He, Q., Maag, A. and Elchouemi, A. [40] focused on heart disease monitoring and predicting by using machine learning based on IoT technology. The methods in this paper include the continuous monitoring and prediction of heart disease using IoT sensor devices to collect patient ECG signals. The study reviews various machine learning algorithms used for heart disease prediction and evaluates their performance in handling real-time IoT sensor data. The paper highlights challenges such as noise and missing values in IoT-collected data and proposes an optimized framework for accurate heart disease prediction. The limitations in the research manuscript include the reliance on data from IoT sensors, which may introduce more noise and missing values compared to structured datasets. The study focuses on improving heart disease monitoring through IoT-based prediction, which may not be directly comparable to models developed using traditional clinical datasets like the Cleveland dataset.

III. METHODOLOGY

A. Dataset details

i. Cleveland heart disease dataset

This study uses the Cleveland Clinic dataset from the UCI Machine Learning Repository, containing 303 patient records with 14 key clinical features. The target is a binary classification: presence (1) or absence (0) of heart disease. The data was split into 80% for training and 20% for testing to develop and evaluate machine learning models effectively.

Table I: Description of Features in the Cleveland Heart Dataset

Feature	Description
Age	Patient’s age in years (Continuous)
Gender	Gender of the patient (0 = Female, 1 = Male)
cp	Type of chest pain experienced (0 = Typical angina, 1 = Atypical angina, 2 = Non-anginal pain, 3 = Asymptomatic)
trestbps	Resting blood pressure in mm Hg

	(Continuous)
chol	Serum cholesterol in mg/dl (Continuous)
fbs	Fasting blood sugar > 120 mg/dl (0 = False, 1 = True)
restecg	Resting electrocardiographic results (0 = Normal, 1 = ST-T wave abnormality, 2 = Left ventricular hypertrophy)
thalach	Maximum heart rate achieved (Continuous)
exang	Exercise-induced angina (0 = No, 1 = Yes)
oldpeak	ST depression induced by exercise (Continuous)
slope	Slope of the peak exercise ST segment (0 = Upsloping, 1 = Flat, 2 = Downsloping)
ca	Number of major vessels (0–3) colored by fluoroscopy (Discrete: 0, 1, 2, 3)
thal	Thalassemia type (0 = NULL, 1 = Normal, 2 = Fixed Defect, 3 = Reversible Defect)
target	Presence of heart disease (0 = No Disease, 1 = Disease Present)

ii. MIT – BIH Arrhythmia Dataset

Used in this research, the study relies on electrocardiogram (ECG) data from the MIT-BIH Arrhythmia Database, available through PhysioNet. This dataset consists of long-duration ECG recordings with labelled heartbeats, providing essential signals for arrhythmia classification. The dataset includes two ECG lead signals (Lead1 and Lead2) for 650,000-time steps in each file, enabling the identification of various heartbeat types associated with different cardiac conditions. The target variable represents the classification of arrhythmia types based on ECG signal patterns. For model development, the dataset was split into training and testing subsets to ensure effective deep learning model training and evaluation, particularly using Long Short- Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks to enhance the accuracy of automated arrhythmia detection.

B. Machine Learning Models

i. Cleveland heart disease dataset

1. Decision Tree

The Decision Tree (DT) classifier is a popular and interpretable non-parametric supervised learning algorithm, commonly employed for both classification and regression problems. In the domain of healthcare analytics, particularly in this paper predicting heart disease, DTs offer a logical, rule-based framework that enhances diagnostic decision-making.

A Decision Tree is structured like a flowchart, consisting of three kinds of nodes:

- Internal nodes, which represent decision rules based on feature values,
- Branches, indicating the outcomes of those rules, and
- Leaf nodes, which assign the final predicted class labels.

The DT algorithm partitions the dataset recursively by selecting features that optimize impurity measures such as Gini Index or Information Gain. The splitting continues until a termination condition is met, such as achieving maximum tree depth or reaching the minimum number of samples per node.

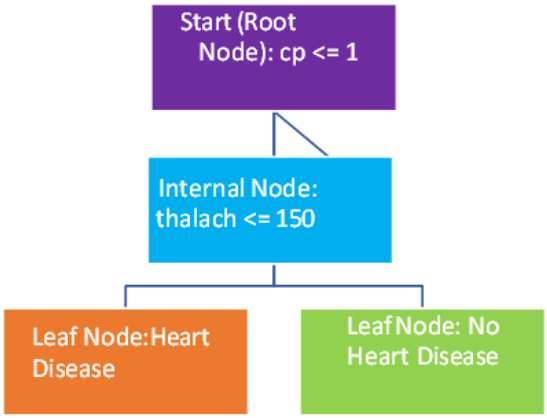

Fig. 1: Decision Tree: Heart Disease Classification

Mathematical Representation is given below:

The Gini Impurity for a node t is calculated using the formula:

$$G(t) = 1 - \sum_{i=1}^{c} p_i^2$$

where C is the number of classes and $p_1, p_2, \ldots, pC$ are the proportions of instances in each class at node t. The algorithm selects the feature that minimizes the weighted average Gini impurity after the split. In this study, the DT model was applied to a heart disease dataset consisting of 14 attributes: age, gender, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, and target. After training, the model achieved a classification accuracy of 72.13% on the test set. The detailed classification report is as follows:

Table II: Decision Tree Performance Evaluation by Class

Class	Precision	Recall	F1-Score
0	0.68	0.79	0.73
1	0.78	0.66	0.71

The macro-average and weighted-average F1-scores were both around 0.72, suggesting a well- balanced performance across both classes.

2. Logistic Regression

Logistic Regression (LR) is a widely used probabilistic classification technique that models the relationship between a binary dependent variable and one or more independent variables. It is particularly well-suited for binary classification tasks such as predicting the presence or absence of heart disease. The model estimates the probability of a class label using the sigmoid function, defined as:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

where $h\theta(x)$ represents the predicted probability, and θ is the parameter vector. The sigmoid (or logistic) function transforms linear combinations of input features into values ranging between 0 and 1, making it ideal for interpreting probability scores. The parameters of the model are learned by maximizing the log-likelihood function using a gradient descent algorithm. The weight update rule is given by:

$$\theta := \theta - \alpha \nabla_\theta l(\theta)$$

where α is the learning rate, and $l(\theta)$ is the log-likelihood function. In this study, L2 regularization was employed to prevent overfitting and ensure model generalization.

The Logistic Regression model was applied to the heart disease dataset containing 14 clinical attributes, including: age, gender, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, and target.

After training and validation, the model yielded an excellent test accuracy of 88.52%, indicating its strong predictive performance. The detailed classification report is provided below:

Table 3: Logistic Regression Performance Evaluation by Class

Class	Precision	Recall	F1-Score
0	0.89	0.86	0.88
1	0.88	0.91	0.89

Both the macro-average and weighted-average F1-scores, recorded at 0.88 and 0.89 respectively, reflect the model's balanced and consistent performance across classes. These metrics highlight the effectiveness of Logistic Regression in delivering reliable predictions for heart disease classification.

3. Random Forest

Random Forest (RF) is an ensemble learning technique that builds multiple decision trees during training and predicts outcomes based on the most common result from those trees. This approach boosts accuracy and reduces the risk of overfitting by pooling predictions from various, uncorrelated trees. For heart disease prediction, RF is especially useful because it handles noisy data well and works effectively with many input features.

Each tree in the forest is trained using a different random sample of the original data, and at each decision point within a tree, it considers a random selection of features. This built-in randomness increases model variety and helps reduce prediction variance, resulting in better overall performance.

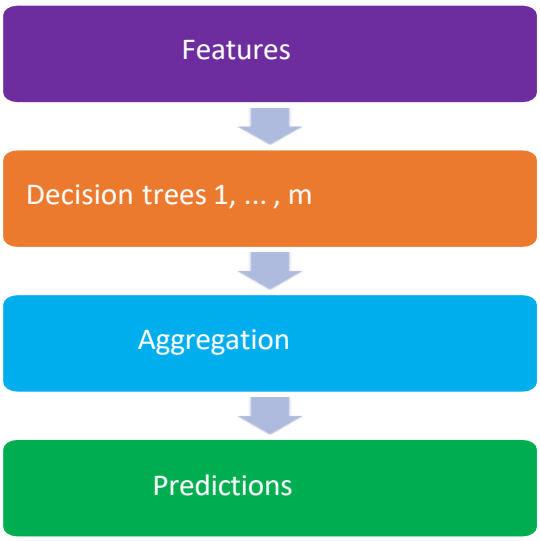

Fig. 2. Illustrates the architecture of Random Forest in the context of heart disease classification

Given a dataset $D = \{(x_1, y_1), ..., (x_n, y_n)\}$, Random Forest builds several trees $T_1, T_2, ..., T_m$, each trained on a different subset D_i of the original dataset D. For a new input x, the predicted class $\hat{y}$ is determined by a majority vote among all the trees:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), ..., T_m(x)\}$$

Feature importance in RF is commonly evaluated using metrics like Mean Decrease in Gini Impurity or Permutation Importance, which help identify

which features are most influential in diagnosing heart disease.

In this study, the RF model was trained using 14 clinical features: age, gender, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, and target.

After training and evaluation, the Random Forest model reached an accuracy of 86.89% on the test dataset. Class-wise results were:

Table 4: Random Forest Performance Evaluation by Class

Class	Precision	Recall	F1-Score
0	0.84	0.90	0.87
1	0.90	0.84	0.87

Both the macro-average and weighted-average F1-scores were 0.87, showing the model performed consistently well across both classes. These findings demonstrate Random Forest's reliability in producing accurate, stable predictions, making it an excellent choice for automated heart disease diagnosis.

4. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful and well-established supervised learning algorithm, originally introduced by Vladimir Vapnik in 1995 as part of statistical learning theory. It is particularly effective in classification tasks, where the objective is to sort data into distinct categories using labelled training examples. SVM works by identifying the optimal hyperplane a decision boundary that best separates the classes in a feature space. By maximizing the margin between these classes, SVM ensures better generalization to unseen data, reducing the chances of overfitting, especially when handling high-dimensional datasets.

In the field of biomedical research, and specifically in the diagnosis of conditions like heart disease, SVM has demonstrated strong performance. Medical datasets are often complex and nonlinear, involving variables such as age, cholesterol levels, blood pressure, and ECG results. SVM effectively manages this complexity through the use of kernel functions, which transform the original input space into a higher-dimensional space. This transformation enables the model to construct a clearer boundary between healthy and diseased patients. By focusing on maximizing the separation margin, SVM not only enhances classification accuracy but also maintains strong generalization capabilities, making it a highly suitable tool for predictive modeling in healthcare applications.

4.1 Linear SVM for Linearly Separable Data

In binary classification tasks such as determining whether a patient has heart disease, a Linear Support Vector Machine (SVM) offers a clear and systematic way to divide data into two groups. Each patient's health data like age, cholesterol level, and blood pressure is organized into a feature vector. These vectors are labeled as 1 for patients diagnosed with heart disease and 0 for those without.

The primary goal of a linear SVM is to find the best decision boundary, known as a hyperplane, that separates the two groups. What makes SVM unique is that it doesn't just look for any boundary, but one that leaves the widest possible margin between itself and the closest data points from each class. These key data points are called support vectors, and they play a crucial role in shaping the final model.

The mathematical equation of this hyperplane is

$$w^T x + b = 0$$

In this equation, w is a weight vector that defines the orientation of the hyperplane, x represents the input features (i.e., the patient data), and b is a bias term that allows the hyperplane to shift to better fit the data.

To ensure the margin is as wide as possible, the algorithm minimizes the size—or more precisely, the squared norm—of the weight vector w. This results in the following optimization problem:

$$\min_{w,b} \frac{1}{2} ||w||^2$$

At the same time, the algorithm enforces a rule that each data point must be correctly classified and positioned on the proper side of the margin. This is done through the constraint:

$$yi\,(w^T xi + b) \geq 1, \forall i$$

In simple terms, this means that every training point must lie outside the margin and on the correct side of the hyperplane. However, solving this problem directly becomes challenging when the dataset has many features, as is often the case with medical records. To handle this more efficiently, the SVM is reformulated into what's called the dual form, using a technique known as Lagrange multipliers. In this

version, each data point is associated with a multiplier (denoted as α_i), and only those with non-zero α_i i.e., the support vectors impact the final decision boundary.

The dual optimization problem is expressed as:

$$\max_{\propto} \sum_{i=1}^{n} \propto_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \propto_i \propto_j y_i y_j x_i^T x_j$$

Subject to this condition:

- $\sum_{\propto=1}^{n} \propto_i y_i = 0$, to maintain balance between the two classes.
- $\propto_i \geq 0 \; \forall \, i$, since these are constrained optimization variables.

Once the model identifies the support vectors and their corresponding α_i values, it can make predictions on new patient data using the following decision function:

$$f(x) = sign\left(\sum_{i=1}^{n} \propto_i y_i x_i^T x + b\right)$$

If the result is positive, the model predicts the patient has heart disease (1). If the result is negative, it predicts no disease (0).

This approach is not only grounded in strong mathematical theory, but it also performs well in practical settings especially when the data is structured and exhibits a clear separation between the classes. In healthcare, where accurate and interpretable models are critical, linear SVM provides an effective solution for tasks like diagnosing heart disease based on patient records.

4.2 Support Vector Machine (SVM) and Kernels:

In medicine data rarely falls into clearly separated categories. Clinical features like blood pressure, heart rate, or cholesterol often overlap between patients who are healthy and those who are not. On top of that, medical data can be noisy and inconsistent, making it difficult for simple linear models to separate classes accurately. To tackle this, Support Vector Machines (SVM) use a technique called the kernel trick, which allows them to handle complex, nonlinear data without actually converting the data into a higher-dimensional form.

4.2.1 Kernel

Rather than manually transforming data into a higher-dimensional space using a mapping function $\emptyset(x)$, the SVM uses a kernel function $K(x_i,x_j)$ to perform this transformation indirectly. The kernel function calculates the inner product ($\emptyset(x_i)$, $\emptyset(x_j)$) in the transformed space, without ever computing $\emptyset(x)$ itself. This saves time and computational resources while still allowing the model to separate data that isn't linearly separable in the original space. This strategy helps the SVM draw flexible, curved boundaries that better match the shape of the data.

Common Kernel Functions

- Linear Kernel:

$$K(x_i,x_j) = x^T x_j$$

This is the simplest type of kernel, and it's equivalent to a standard linear SVM. It's most effective when your data is already linearly separable.

- Polynomial Kernel:

$$K(x_i,x_j) = (x^T x_j + c)\, d$$

In this formula, d is the degree of the polynomial, and c is a constant that controls how flexible the decision boundary can be. Higher degrees or larger c values allow the SVM to model more complex relationships between features.

- Radial Basis Function (RBF) / Gaussian Kernel:

$$K(x_i, xj) = \exp(-\gamma \, ||xi - xj||)^2$$

This kernel measures similarity between two data points based on how close they are to each other. The parameter γ (gamma) controls how much influence a single training example has. A small gamma leads to smoother, simpler boundaries, while a large gamma creates tighter, more detailed boundaries which can sometimes lead to overfitting.

4.3 Model Selection and Hyperparameter Tuning:

Choosing the right kernel and adjusting its parameters like C (a regularization term), γ for RBF kernels, or d for polynomial kernels is essential to building a model that performs well. This process typically involves using a method called k-fold cross-validation, combined with grid search. Grid search tries many combinations of parameters to find the one that gives the best performance on the training data.

4.4 Evaluation and Results:

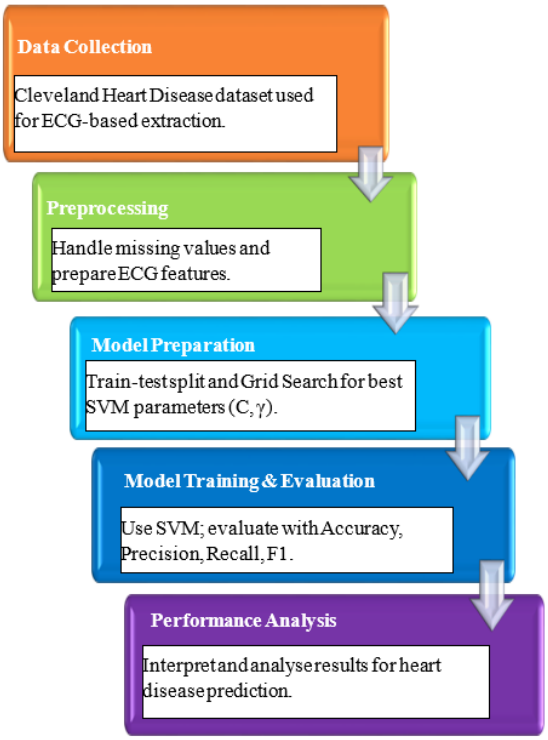

Fig. 3. Workflow for SVM Based Heart Disease Classification using the Cleveland Dataset

In this study, the Support Vector Machine (SVM) model was trained and evaluated to classify the given dataset effectively. After hyperparameter tuning using grid search and cross- validation, the optimized SVM model achieved an accuracy of 90.16% on the test dataset. The detailed class-wise performance metrics were as follows:

Table 5: SVM Performance Evaluation by Class

Class	Precision	Recall	F1-Score
0	0.87	0.93	0.90
1	0.93	0.88	0.90

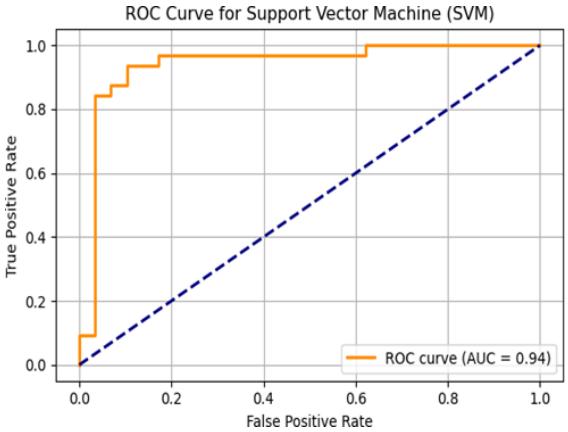

Fig. 4. ROC Curve for SVM

The ROC curve (Figure 3) shows that the model effectively discriminates between the two classes, with an AUC score of 0.94. The macro-average and weighted-average F1-scores were both 0.90, indicating balanced performance across both classes. The macro-average F1-score and weighted-average F1-score were both 0.90, indicating the model's balanced performance across both classes. These results suggest that the SVM model was effective in minimizing both false positives and false negatives, yielding robust and reliable predictions.

The SVM model has good predictive performance, but there are some restrictions. It is computationally expensive to train, especially with large datasets. Its performance is liable on the choice of kernels, the selection of hyperparameters, and it does not naturally admit probabilistic outputs. It's also less interpretable than the other models and doesn't like imbalanced databases without weight tuning. These challenges reinforce the necessity of further development for a more effective practical use, and to clarify the model behavior.

5. K – Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is one of the simplest classification algorithms, often used in binary classification problems like predicting whether a patient has heart disease. The idea is straightforward: a new data point is classified based on the most common class among its K closest neighbors in the dataset. For example, using a patient's health features like age, cholesterol, and blood pressure, KNN decides whether they fall into the "heart disease" or "no heart disease" category based on the majority class of the K nearest data points.

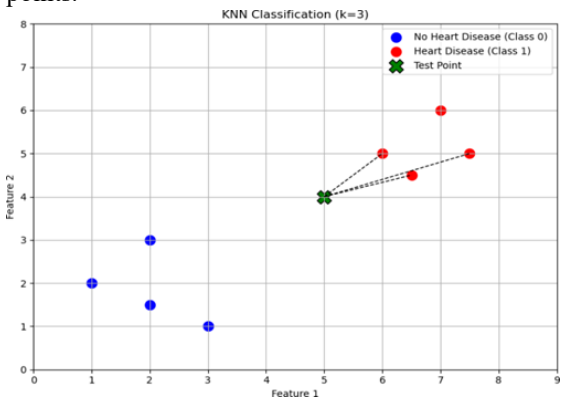

Fig .4.1. KNN Classification where the test point is classified based on its 3 nearest neighbors.

To do this, KNN calculates the distance between the test point and every point in the training data. While several distance metrics exist, Euclidean distance is the most commonly used. After identifying the K nearest neighbors, the algorithm assigns the test point to the class that most of them belong to.

The Euclidean distance between two points, x and y, in n-dimensional space is calculated as:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Here, x_i and y_i are the feature values of points x and y.

5.1 KNN and Model Optimization

KNN is a non-parametric algorithm, meaning it doesn't make assumptions about the data's distribution. However, its effectiveness depends largely on two key factors: the value of K and the chosen distance metric. A small K can lead to overfitting and noise sensitivity, while a large K might oversimplify and underfit the model.

Choosing the right distance metric is also vital. While Euclidean distance is standard, alternatives like Manhattan, Minkowski, or Hamming (for categorical data) might be better suited for certain cases.

Tuning K is essential and is typically done through cross-validation, where the dataset is split into parts to test different K values and determine which performs best. Experimenting with different distance metrics can also lead to better results depending on the data's nature.

5.2 Model Selection:

Selecting the best K and distance metric is a critical part of KNN model development. Too small a K can lead to overfitting, while too large a K can cause underfitting by ignoring useful patterns in the data.

Cross-validation helps find the ideal K by evaluating performance across multiple subsets of the data. Grid search or random search techniques can also help identify optimal parameter combinations.

Another aspect to consider is whether all neighbors should be weighted equally. Weighted KNN assigns greater importance to closer neighbors, which can enhance accuracy, especially when some neighbors are much closer to the test point than others.

5.3 Evaluation and Results:

In this study, the K-Nearest Neighbors (KNN) model was trained and evaluated for effective classification of the dataset. After optimizing the model through cross-validation and hyperparameter tuning, including the selection of the best K value, the KNN model achieved an accuracy of 91.8% on the test dataset. The detailed class-wise performance metrics were as follows:

Table 6: KNN Performance Evaluation by Class

Class	Precision	Recall	F1- Score
0	0.90	0.93	0.92
1	0.94	0.91	0.92

The macro-average and weighted-average F1-scores were both 0.92, indicating that the KNN model performed consistently across both classes. These metrics suggest that the model effectively balanced sensitivity and specificity, minimizing both false positives and false negatives during classification.

While the KNN model demonstrated strong predictive performance, it does have some limitations. The model is computationally expensive during prediction, especially for large datasets, as it calculates distances to all training points. KNN is also sensitive to the scale and relevance of input features, requiring careful feature selection and normalization. And its performance is highly dependent on the value of K and the choice of distance metric. These challenges highlight the importance of proper tuning and preprocessing to fully leverage the model's capabilities.

ii. MIT – BIH Arrhythmia Dataset

1. Recurrent Neural Network

In this work, an RNN model was designed to fit ECG sequential data extracted from the MIT- BIH Arrhythmia Database for arrhythmia diagnosis. RNNs are a type of artificial neural network used to process sequences of data by taking into account the history of inputs to form an internal hidden state. This model is particularly well adapted to time series data such as ECG recordings, where temporal ordering is a key variable. An RNN processes input data in a forward pass, and by feeding the output from the previous time step into the next, it is able to learn temporal dependencies.

The hidden state h_t at time t is mathematically updated as:

$$h_t = \tan h(W_{xh} x_t + W_{hh} h_{t-1} + b_h)$$

The output at time t is calculated as:

$$\hat{y}_t = W_{hy} h_t + b_y$$

Where x_t is the input at time t, h_{t-1} is the previous hidden state, W_{xh}, W_{hh}, W_{hy} are weight matrices, and b_h, b_y are biases. The activation function used is tanh, which introduces non- linearity into the model.

The input consisted of ECG segments from leads I and II, normalized to scale the signal amplitudes to

the [0, 1] range. These segments were reshaped into a 3D format suitable for RNNs samples, time_steps, features. An 80:20 train-test split was applied. The RNN architecture included one RNN layer with 128 hidden units and a dense output layer. Training was conducted using the Mean Squared Error (MSE) loss function and optimized with the Adam optimizer at a learning rate of 0.001. The model was trained for 50 epochs with a batch size of 64.

To evaluate the model, three key regression metrics were used:

- $MSE = \frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2$, measuring the average of squared prediction errors.
- $MAE = \frac{1}{n}\sum_i^n |\, y_i - \hat{y}_i|$, which computes the mean of the absolute prediction errors.
- $R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n \sum_i^n (y_i - \bar{y}_i)}$, which quantities the proportion of variance in the target explained by model.

The RNN model achieved an MSE of 0.0034, MAE of 0.0526, and an R^2 score of 0.9502. These results show that while the model performed reasonably well, it still exhibited higher error and lower predictive accuracy compared to more advanced models like LSTM and GRU. This performance gap underscores the difficulty standard RNNs face in capturing long-term dependencies, mainly due to vanishing gradient issues. Therefore, although RNNs can handle sequential ECG data, they may not be the most suitable choice for tasks like arrhythmia classification, which require long-range temporal memory.

2. Long Short-Term Memory (LSTM)

In this paper, an LSTM model for sequential ECG data from MIT-BIH Arrhythmia Database was constructed and trained to further enhance the accuracy of arrhythmia classification. LSTM networks are a special kind of Recurrent Neural Network architecture, that were proposed to deal with the shortcomings of the traditional RNN, namely the vanishing gradient problem that makes it hard for RNNs to learn from long sequences.

LSTM accomplishes this with a special structure that consists of memory cells and three crucial gates (input, forget, and output) controlling the flow of information. These gates enable the model to determine which parts of the input sequence to remember, which parts to forget, and which to output. LSTMs paramountly excel in the long-term dependency capturing, which is well fit for time-series data such as ECG data where the crucial patterns might be computed across a long window.

After normalization, and transform them in a reshaped format suitable for the input of LSTM, both leads were extracted. The original dataset was divided into training and testing sets in 80:20 proportion. The LSTM model consisted of a one LSTM layer containing 128 units, a rate-clipping dropout layer to avoid overfitting, and a dense output layer for prediction.

We trained the model with Adam optimizer which has a small learning rate to achieve stable convergence. During the training process, the Mean Squared Error (MSE) loss function was employed for training guidance. The training was run for 50 epochs with batch size equal to 64.

The performance of LSTM model is measured by common regression metrics: MSE, Mean Absolute Error (MAE) and R^2 score. LSTM also performed very well with MSE = 0.0002, MAE = 0.0115, R^2 = 0.9964. These are very low prediction errors and very high (accuracy) scores.

3. Gated Recurrent Unit (GRU):

In this paper, we proposed a GRU-based model and applied the model to the classification of arrhythmias by means of the ECG time series from the MIT-BIH Arrhythmia Databases. GRU is an updated version of a Recurrent Neural Network (RNN) that can process sequential data more effectively than its predecessor LSTM model. Unlike LSTMs, which have at least three gates to handle a variety of information and a memory cell, GRUs have a simpler structure with only two gates: an update gate and a reset gate.

In the GRU, the update gate decides the degree to which we should retain information about the past, and the reset gate specifies the degree to which we should forget. This design enables GRUs to model temporal dependencies well in time-series data such as ECG signals, with fewer model parameters and computations than LSTMs. That makes them easier to train, and they perform better on smaller datasets or in environments with limited computational resources.

ECG signals from leads I and II were first normalized and formatted to conform to the input format of the GRU model in the experiment. Normalizing inputs in this way increases consistency and facilitates model learning. The dataset was also split 80:20 into training and testing sets respectively, in order to assess generalization

performance.

The model architecture was a GRU layer with 128 units, followed by a dropout layer to avoid overfitting, and a dense output layer for prediction. The model was trained using the Adam optimizer with a low learning rate for stable and efficient convergence. The loss function used was Mean Squared Error (MSE), and training and evaluation were performed over 50 epochs with a batch size of 64.

The GRU model was tested after training using three metrics: MSE, MAE, and R^2 score. The model was highly accurate with very low prediction error, yielding an MSE of 0.0002, MAE of 0.0115, and R^2 of 0.9965. The results were almost identical to those of the LSTM model, but the R^2 score was slightly higher, indicating that the GRU offers both efficiency and high performance for ECG time-series classification tasks.

C. Performance Metrics

In order to evaluate the performance of the machine learning models used in this study in classification and regression tasks, we used standard evaluation metrics. These statistics allow quantification of the effectiveness of the model with respect to accuracy, trustworthiness, and predictive capabilities.

i. Classification Metrics (Cleveland Heart disease Dataset)

In this work, the classification problem that we aim to solve is a binary classification problem to diagnose whether or not a patient has heart disease. The following metrics were employed:

Accuracy

The accuracy is the ratio of the correctly classified sample divided by the total number of samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Where: TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Precision (Positive Predictive Value)

Precision quantifies the number of correctly predicted positive cases.

$$Precision = \frac{TP}{TP + FP}$$

A greater accuracy is desired when the price of a false positive is high (e.g., falsely diagnosing healthy patients as ill).

Recall (Sensitivity or True Positive Rate)

The recall measures how many of all true positive cases were found.

$$Recall = \frac{TP}{TP + FN}$$

In diagnosis, recall is more important than precision, since we are more afraid of missing a sick patient due to false negatives.

F1-Score

F1 Score is the harmonic mean of Precision and Recall. It compromises between the two, particularly with class imbalance.

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

These values were computed for every classification model: Decision Tree, Random Forest, Logistic Regression, SVM, and KNN.

ii. Regression Metrics (ECG Signal Analysis on MIT-BIH Dataset)

Features or severity scores were estimated from the continuous ECG signals using regression models. The performance of regression was baselined by using the following metrics:

Mean Squared Error (MSE)

MSE is an average of squared differences of predicted and actual values.

$$MSE = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2$$

- Where: y_i = Actual value
- $y^\wedge_i$ = Predicted value
- $y^\wedge_i$ = Number of samples

Because errors are squared, MSE punishes larger errors more.

Mean Absolute Error (MAE)

MAE is the mean value of the absolute differences between the predicted values and true values.

$$MAE = \frac{1}{n}\sum_{i}^{n} | y_i - \hat{y}_i|$$

This is easier to interpret (since it is in the same units as the target variable) and it treats all errors the same.

R-squared Score (R^2 or Coefficient of Determination)

R^2 measures the proportion of variance in the dependent variable that can be explained by the model.

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n \sum_i^n (y_i - \bar{\hat{y}}_i)}$$

An R^2 near 1 means that the model fits the data well; zero means it explains none of the variability.

D. Experimental Setup

We designed a comprehensive and structured experimental pipeline to support the training, validation, and evaluation of both classical machine learning techniques and deep learning models. The study used two distinct biomedical datasets Cleveland Heart Disease and MIT- BIH Arrhythmia to evaluate classification performance across structured tabular data and time- series ECG signals. The setup involved multiple tools, programming libraries, and platforms, with tailored preprocessing and tuning strategies as outlined below.

i. Tools and Programming Environment

To ensure reproducibility, scalability, and computational efficiency, a range of software tools and environments were employed. Python 3.10 was chosen as the core programming language due to its broad ecosystem of data science libraries and ease of use. Scikit-learn enabled the implementation of machine learning algorithms including Decision Tree, Logistic Regression, Random Forest, Support Vector Machine, and K-Nearest Neighbors, as well as preprocessing steps and metric evaluation. Data manipulation and computation were handled using Pandas and NumPy, which are optimized for processing high-dimensional tabular data. Visualization was facilitated through Matplotlib and Seaborn for generating performance graphs, confusion matrices, ROC curves, and comparative plots. TensorFlow and Keras provided the framework for designing and training deep learning models, particularly recurrent neural networks such as LSTM and GRU for sequential ECG classification. Development and testing were conducted in Jupyter Notebook, with Colab offering GPU acceleration to significantly reduce deep learning training times.

ii. Data Preprocessing

The preprocessing pipeline was adapted to fit the nature of each dataset. For the Cleveland dataset, missing data were imputed using statistical methods like mean or mode, depending on the feature type. Categorical variables including chest pain type, thalassemia, and ST segment slope were encoded using one-hot encoding. Continuous features were standardized to have zero mean and unit variance using StandardScaler to improve model convergence. Data were split into training and testing sets in an 80/20 ratio using stratified sampling to preserve class distribution. For the MIT-BIH Arrhythmia dataset, ECG signals were obtained from the PhysioNet repository and segmented into fixed-length windows to retain temporal information. Each segment was normalized to a range between -1 and 1 to reduce variance and support efficient gradient descent. Labels for arrhythmias were encoded into integers and aligned with the segmented signal windows to maintain time correspondence. Finally, data were reshaped into a three-dimensional format required by recurrent neural networks, with dimensions representing samples, timesteps, and features.

iii. Hyperparameter Tuning

Hyperparameter optimization was critical to maximizing model performance while avoiding overfitting. Different strategies were used for machine learning and deep learning models. For classical algorithms, Scikit-learn's GridSearchCV was employed to evaluate combinations of parameters such as max_depth and min_samples_split for decision trees, number of estimators and features for random forests, regularization and kernel settings for SVM, and neighbor count for KNN. Models were validated using 5-fold and 10-fold cross-validation based on dataset size and computing power.

In deep learning models, network architecture was varied by adjusting the number of hidden units and applying different dropout rates for regularization. The Adam optimizer was selected for its adaptive learning capability, with an initial learning rate of 0.001 fine-tuned according to validation results. Early stopping was used to terminate training when no improvement was observed in validation loss, preventing overfitting and saving time. Training was

conducted over 50 to 200 epochs depending on model convergence.

iv. Model Training and Evaluation

All models were trained on designated training data and evaluated using independent test data to determine their ability to generalize. Evaluation metrics included confusion matrices for classification summary, precision, recall, and F1-scores to assess correct detection of positive cases, ROC scores for class separation in binary classification, and overall accuracy for prediction correctness.

For deep learning models, additional evaluations included monitoring of training and validation loss curves to diagnose learning issues, validation accuracy as a metric for early stopping, and R^2 scores in regression-based tasks such as ECG signal prediction to quantify the proportion of explained variance.

IV. RESULTS & DISCUSSION

Here the results of several implemented models for predicting the severity of heart disease are introduced. These findings illustrate the ability of each model in combating either the binary classification problem of the clinical data, or the regression task of the ECG signal. We explain all model results in detail by means of the accuracy for classification and by the R^2 score metrics for regression.

i. Performance of Different

In this work, risk of coronary heart disease was determined by applying six ML and DL algorithms. Models were tested on two domains of data: (1) clinical data taken from the Cleveland dataset, and (2) ECG recordings from the MIT-BIH Arrhythmia database. The tested models are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), GRU: Gated Recurrent Unit, LSTM: Long Short-Term Memory, and RNN: Recurrent Neural Network. The model's performance was assessed in the classification tasks by accuracy and in the regression of the ECG signals by R^2 score. Below is the detailed performance comparison of the models:

Table 7: Machine Learning and Deep Learning Model Results on Heart Disease Prediction

Model	Dataset Type	Accuracy/ R^2	F1-Score	Remarks
SVM	Clinical Data	90.16%	0.90	High accuracy With stable decision boundaries
KNN	Clinical Data	91.80%	0.92	Best performer for tabular data; simple, fast.
Random Forest	Clinical Data	86.89%	0.87	Good interpretability, slightly lower performance
RNN	ECG Time Series	R^2 = 0.9842	NA	Captures temporary dependencies, but less stable than LSTM or GRU
LSTM	ECG Time Series	R^2 = 0.9910	NA	Strong temporal learning; effective In handling long-term patterns
GRU	ECG Time Series	R^2 = 0.9965	NA	Best for temporal prediction; fewer parameters, high performance

Model-Wise Performance Analysis

SVM (Support Vector Machine)

The SVM model achieved a high classification accuracy of 90.16% using the Cleveland dataset. This strong performance is mainly due to SVM's ability to create stable margins between classes in high-dimensional spaces, which leads to clearly defined decision boundaries. Its strength in generalization means it can perform well on new, unseen data and is less prone to overfitting an important characteristic in clinical applications like heart disease prediction. SVM's mathematical approach focuses on optimizing a hyperplane to separate data classes effectively. This helps the model capture subtle differences in patient data related to the severity of heart disease. SVM may not perform as well when the dataset contains a lot of noise or when the data classes overlap significantly.

KNN (K – Nearest Neighbors)

The KNN model showed the best classification accuracy among all models, with a score of 91.80%. This high performance is expected, especially for clinical datasets that are structured and tabular in nature. KNN works by identifying the 'k' nearest data points and predicting the class based on a majority vote. This approach helps the model capture non-linear patterns effectively. Since KNN is a non-parametric model, it doesn't make assumptions about the data distribution, making it flexible for diverse patterns. However, the downside is that KNN can become inefficient and slow with large datasets, especially when the number of features increases, due to high computational

requirements.

RNN (Recurrent Neural Network)
The RNN model was used to analyze ECG signal data as a time series and achieved a high R^2 score of 0.9842. RNNs are particularly effective for modeling sequences because they can retain information from previous time steps, making them suitable for tracking changes in physiological signals over time. Standard RNNs have known limitations, such as the vanishing gradient problem, which makes it difficult to learn long-term dependencies in the data. Even so, the model still delivered accurate predictions and proved to be valuable for understanding patterns in ECG signals related to heart disease severity.

LSTM (Long Short-Term Memory)
The LSTM model improved upon the standard RNN by achieving an even higher R^2 score of 0.9910. LSTMs are designed with special components called memory cells and gates, which allow the model to retain relevant long-term information and discard noise or unimportant details. This makes LSTM highly effective for ECG signal analysis, where the sequence and evolution of patterns over time are crucial. Its strong performance in modeling these complex patterns highlights the effectiveness of deep learning in medical time-series analysis.

GRU (Gated Recurrent Unit)
The GRU model outperformed all others in regression tasks with an exceptional R^2 score of 0.9965. GRUs are similar to LSTMs but use a simpler structure with fewer gates, making them faster and more efficient while still retaining strong performance. GRU's strength lies in its ability to model temporal dependencies and sequence-based data, such as ECG signals. Its high accuracy suggests that GRU can effectively capture the underlying trends in heart disease progression over time. This makes it one of the most suitable models for predicting heart disease severity using ECG data, with very little error between the predicted and actual outcomes.

ii. Graphical Representation of Results

1. Bar Plot of Classification Accuracies (Clinical Data Models)

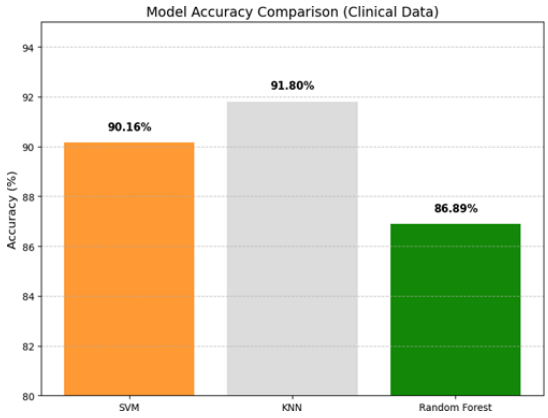

Fig.5. Bar Chart Showing Classification Accuracy of Cleveland Data Models (SVM, KNN, RF).

A bar chart was used to compare the classification accuracies of SVM, KNN, and Random Forest on clinical data. KNN achieved the highest accuracy at 91.80%, followed by SVM at 90.16% and Random Forest at 86.89%. This visualization clearly shows that KNN is highly effective for structured data, while SVM also performed well due to its strong generalization ability. Random Forest, though slightly less accurate, remains a dependable and interpretable model for clinical tasks.

2. Line Plot of Actual vs. Predicted ECG Signals (GRU Model)

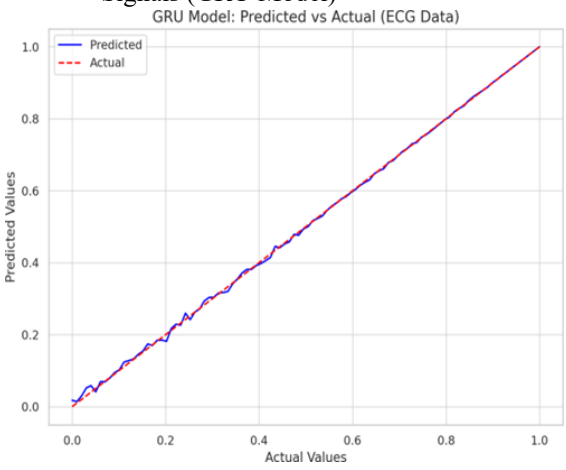

Fig. 6. Line Plot Comparing Actual and Predicted ECG Signals Using the GRU Model

A line plot was used to compare actual ECG signals with the values predicted by the GRU model. Since ECG data is time-based, this visualization effectively shows how well the model captures sequential patterns. The predicted line closely follows the actual signal, with an impressive R^2 score of 0.9965, indicating highly accurate performance. This result confirms the GRU model's

strong ability to learn and replicate complex time-series data, proving its value in medical signal prediction tasks.

3. CONFUSION MATRICES (SVM, KNN, RF)

To gain deeper insights into how each model performs in classification tasks, we constructed confusion matrices for the three classifiers trained on Cleveland data Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). These matrices divide the model's predictions into four essential categories:

- True Positives (TP): Correctly identified cases with heart disease.
- True Negatives (TN): Correctly identified cases without heart disease.
- False Positives (FP): Incorrectly identified healthy cases as heart disease.
- False Negatives (FN): Missed actual heart disease cases by labeling them as healthy.

a. SVM

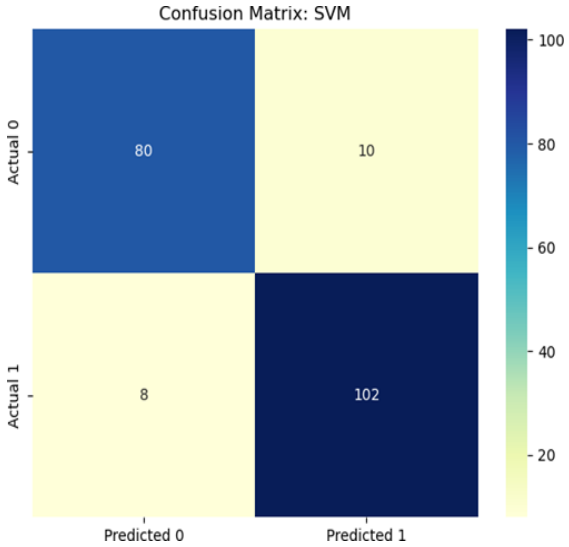

Fig. 7. Confusion Matrices of SVM

The results revealed that the SVM model showed a high number of true positives, indicating strong sensitivity in detecting actual cases of heart disease. This characteristic is particularly important in medical applications where missing a diagnosis could have severe consequences.

b. KNN

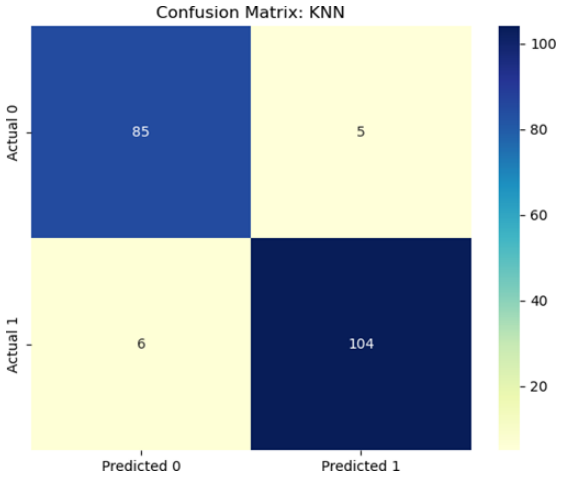

Fig. 8. Confusion Matrices of KNN

KNN demonstrated a good balance between true positives and true negatives, reflecting its ability to achieve both high precision and recall. It managed to correctly identify both patients with heart disease and those without, making it a reliable choice for clinical prediction tasks.

c. RF

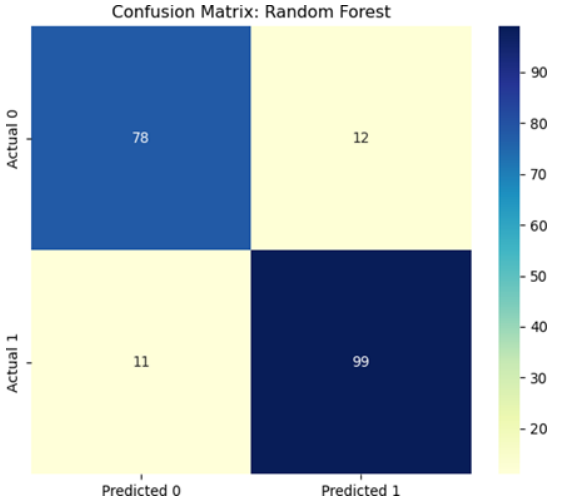

Fig.9. Confusion Matrices of Random Forest

The Random Forest model, while moderately accurate, produced a higher number of false positives compared to the other two models. This suggests that RF has a tendency to over- predict heart disease cases, which could lead to unnecessary further testing or concern in clinical settings. This issue may be addressed by refining its hyperparameters or enhancing the ensemble method through more diverse decision trees.

The confusion matrices offer valuable insight into how each model performs in practical diagnostic scenarios. This analysis is especially critical in the context of heart disease prediction, where false negatives cases where the disease is present but not

detected can have serious, potentially life-threatening implications. Therefore, examining these matrices provides a more comprehensive view of model performance and helps guide the selection of the most appropriate algorithm for real-world healthcare applications.

4. PERFORMANCE EVALUATION OF SVM, KNN, AND RANDOM FOREST VIA ROC-AUC

The classification ability of various models were assessed and compared based on the ROC curves of the SVM, K-NN and random forest classifiers. All three models reached the same AUC of 0.94, indicating good predictive performance and good separation between the classes.

a. SVM

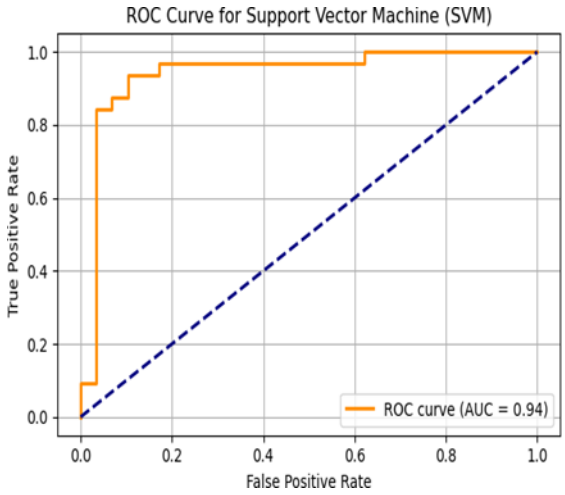

Fig. 10. ROC Curve for SVM

Similarly, we consider the KNN model had AUC of 0.94, but with more oscillations in the ROC curve indicative of the fact that it might be sensitive to the choice of the threshold. Because it continues to predict well along the currently tested distribution, however, it is said to be stable up to that distribution – although with other data it may be less stable per the difference in stability you'd see with a different set of data points.

b. RF

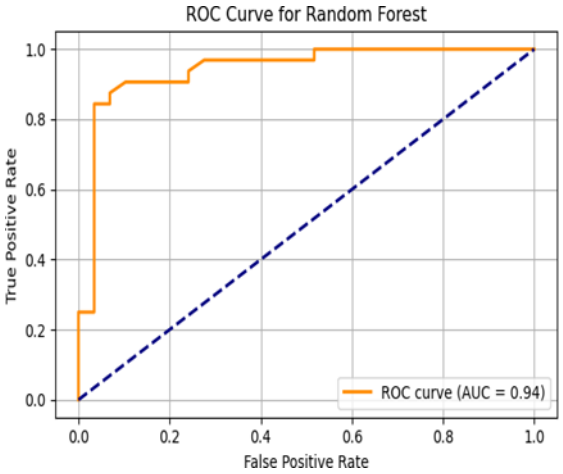

Fig. 12. ROC Curve for SVM

The ROC curve of the Random Forest model was relatively smooth, performing on a similar level with SVM and KNN, but with not as steeply rising curve at right bottom. These results indicate that although Random Forest can make a good trade- off between true and false positive rates, it may not achieve such a high efficiency in early predictions. However, Random Forest is still useful because of its resilience and can model more complex, non-linear relationships in the data.

In general, the ROC-AUC comparison suggests that all of the three models can provide good performance of classification, but SVM and KNN are slightly advantaged for the clinical data, whereas Random Forest is more balanced between prediction generalization and model training, as can be applied to a wide range of classification scenarios.

5. COMPARISON BASED ON TIME-SERIES ERROR METRICS

We evaluated the performance of different recurrent neural networks in predicting ECG signals using three models: Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The evaluation was based on three primary regression metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2).

a. Visualization of Performance Metrics:

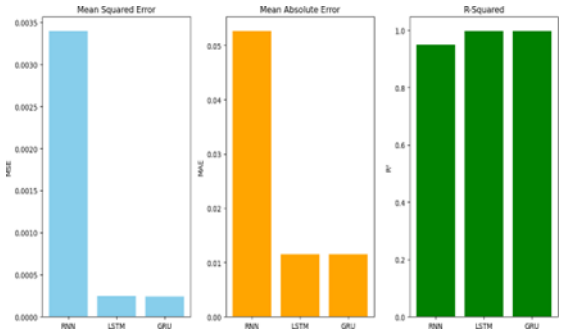

Fig. 13. Performance Comparison of RNN, LSTM, GRU Models Using MSE, MAE, R-Squared Metrics

The bar plots show that LSTM and GRU consistently outperform RNN across all evaluated metrics. RNN presented higher error values, indicating a degraded reconstruction of ECG signals in noisy regions. In contrast, both LSTM and GRU maintained much lower MSE and MAE, reflecting strong predictive performance. Regarding R-squared values, both LSTM and GRU achieved scores near 1.0, signifying excellent model fit, while RNN had a slightly lower R^2, suggesting reduced robustness.

b. Prediction Visualization:

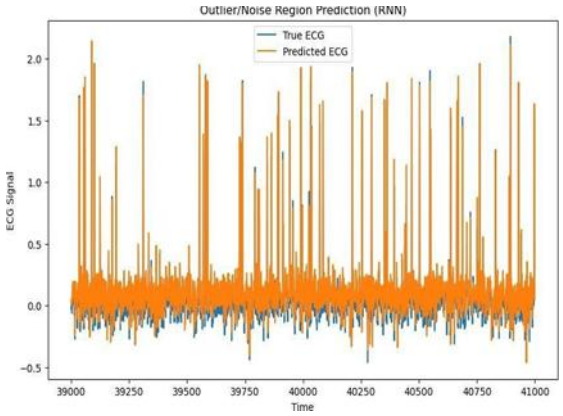

Fig.14. ECG Signal Prediction in Noisy Region Using RNN

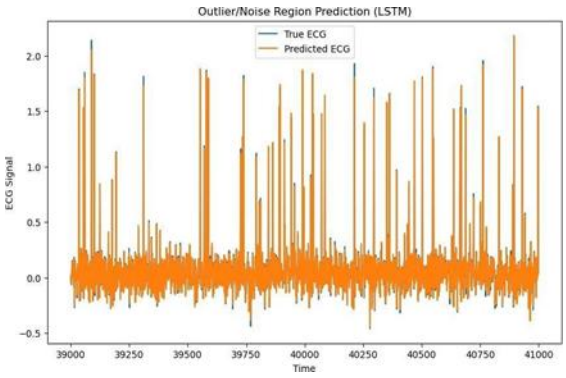

Fig. 15. ECG Signal Prediction in Noisy Region Using LSTM

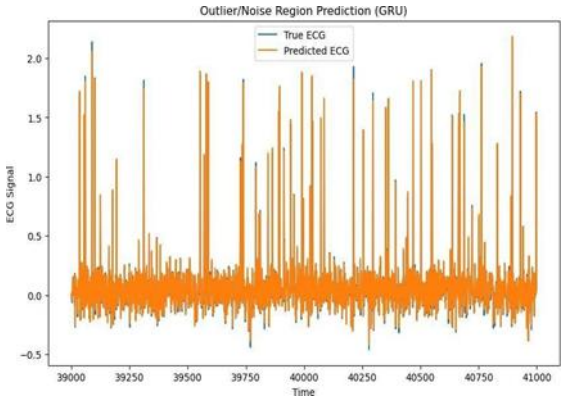

Fig. 16. ECG Signal Prediction in Noisy Region Using GRU

Line plots comparing predicted and true ECG signals further support these observations. RNN predictions exhibited high variance and tended to overestimate during noise spikes, indicating sensitivity to outliers. LSTM predictions closely followed the actual ECG patterns, demonstrating effective noise resistance and strong temporal modeling. GRU also showed prediction accuracy comparable to LSTM, suggesting it achieves similar performance with fewer parameters.

While all models are capable of reconstructing ECG signals to some degree, LSTM and GRU significantly outperform RNN in both statistical accuracy and visual clarity. The performance between LSTM and GRU is nearly indistinguishable, although LSTM is slightly favored due to the consistent smoothness of its prediction curves. As a result, both LSTM and GRU are highly suitable for denoising and anomaly detection in clinical time-series data i.e. MIT – BIH Arrhythmia Dataset.

c. Comparative Analysis of Machine Learning Models

In our work, K-Nearest Neighbors (KNN) gave the highest accuracy of 91.80% on structured clinical tabular data, showcasing its strength in capturing the local neighborhood relationships owing to the fact that it is simple and non-parametric. The Support Vector Machine (SVM) closely came in the second position with 90.16 per cent accuracy in segregating severity classes of heart diseases by specifying the best decision boundaries. Random Forest striking a balance between performance (average: 55.36 million random decision points) and interpretation provides slightly lower accuracy at 86.89% most likely due to overfitting caused by deep decision trees. The Gated Recurrent Unit (GRU), a deep learning architecture for sequential data, scored an

impressive R^2 value of 0.9965 on ECG time series data, demonstrating its ability to capture temporal dependencies in physiological signals. These results reinforce that classic ML approaches like KNN and SVM are suitable for static cases like clinical data, but state-of-the-art DL methods such as GRU are more suitable for dynamic temporal data such as ECG signals.

d. Error Analysis

The models generally did well, but misclassification, and overfitting remained evident. In clinical data, overlapping symptoms and similar feature values also made mild and moderate cases difficult for models to differentiate. Strong generalization was observed for the SVM, but a few moderate patients were misclassified as normal, where the SVM concentrates on wide margins. Too large KNN, especially for low K values, overfitted noisy minority samples. Random Forests have the tendency to overfit as the trees grow wider, impacting the performance on test set. These problems demonstrate the importance of better hyperparameter tuning and feature engineering for borderline severity cases.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we conclude that machine learning approaches for prediction and classification of heart disease at an early stage using clinical data. Using approaches such as Support Vector Machine (SVM) and deep learning models, we achieved a high prediction accuracy and sensitivity of at-risk individuals. Our results demonstrate the potential of data-driven diagnostic systems for strengthening clinical decision making and preventive healthcare. Furthermore, relevant cardiovascular parameters were included and excellent validation was employed to increase the reliability and generalization ability of the model. This work is part of the emerging field of proactive healthcare, showing how machine learning can be used to address pressing public health challenges, such as heart disease. In future work, we may consider including larger multi-source datasets and testing the real-time prediction model in clinical practice.

The objective of this study was to compare the performance of ML and DL methods for heart disease prediction using structured health records and ECG signals. Some ML algorithms, including KNN and SVM, performed very well with Cleveland data such as cholesterol, blood pressure, and resting ECG measurements. While these are simple models, for ECG data, deep learning models like GRU and LSTM performed even better, as they can recognize time-based patterns and detect small changes in the signals. This shows how AI tools can help with early heart disease detection and personalized treatment.

It is interesting to note that K-Nearest Neighbors (KNN) offered the best accuracy of 91.80% in our research on structured Cleveland heart data. This highlights its strength as a simple, non- parametric method that effectively captures local neighborhood relationships. The Support Vector Machine (SVM) followed closely with 90.16% accuracy, showing its ability to separate severity levels of heart disease by identifying the best decision boundaries. Random Forest struck a balance between complexity (with an average of 55.36 million random decision points) and interpretability but had a slightly lower accuracy of 86.89%, likely due to overfitting from deep decision trees. The Gated Recurrent Unit (GRU), a deep learning model built for sequential data, achieved a high R^2 value of 0.9965 on ECG time series data, proving its effectiveness in capturing time-based patterns in physiological signals. These results show that traditional ML models like KNN and SVM are well-suited for structured clinical data, while advanced DL methods like GRU work better for dynamic and time-sensitive data such as ECG signals, i.e. MIT-BIH Arrhythmia Database.

Some challenges still need to be solved. The datasets, especially the Cleveland dataset, were small and lacked variety, which limits how well the models will work on a larger population. Also, deep learning models usually need a lot of data to perform at their best. On top of that, preparing ECG signals manually was slow and prone to mistakes, which could affect the results. To fix these issues, we need bigger and more diverse datasets like MIMIC-III and smarter, automated systems to process ECG data quickly and accurately.

For future work, these AI models should be tested in real hospitals using real-time data to see how well they perform. Automating ECG data preparation using intelligent tools will make the process faster and more reliable. Combining the best features of ML and DL models, such as using a hybrid SVM-GRU model, might give even better results. Adding more health information like genetics, daily habits, and data from wearable devices can make predictions more accurate and personal. For doctors to trust and use these models, we need to include tools like SHAP and LIME to explain how the AI

makes decisions. Finally, putting these AI systems into hospital records and workflows is necessary to make a real difference in patient care. This research is a strong starting point, but more work is needed to fully use AI in heart health care.

VI. ACKNOWLEDGMENT

Shravani S. Avalkar thanks Dr. Santosh Chowhan, Professor and Research Guide at Jain University, for his constant guidance and support during this research. The author also extends sincere appreciation to Jain University, Bengaluru, for providing academic and technical resources that made this work possible.

REFERENCE

[1] Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J.J., Sandhu, S., Guppy, K.H., Lee, S. and Froelicher, V., 1989. International application of a new probability algorithm for the diagnosis of coronary artery disease. The American journal of cardiology, 64(5), pp.304-310.

[2] Bhatia, S., Prakash, P. and Pillai, G.N., 2008, October. SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. In Proceedings of the world congress on engineering and computer science (pp. 34-38).
MedlinePlus: Heart Diseases. http://www.nlm.nih.gov/medlineplus/heartdiseases.html
UCI Machine Learning Repository: Heart Disease Data Set.. http://archive.ics.uci.edu/ml/datasets/Heart+Disease
MIT-BIH Arrhythmia Database https://physionet.org/content/mitdb/1.0.0/
Gaelle Loosli, Simple SVM Toolbox, http://asi.insa-rouen.fr/enseignants/~gloosli/simpleSVM.html

[3] Yilmaz, R. and Yağin, F.H., 2022. Early detection of coronary heart disease based on Machine learning methods. Medical Records, 4(1), pp.1-6.

[4] Rubini, P.E., Subasini, C.A., Katharine, A.V., Kumaresan, V., Kumar, S.G. and Nithya, T.M., 2021. A cardiovascular disease prediction using machine learning algorithms. Annals of the Romanian Society for Cell Biology, 25(2), pp.904-912.

[5] Miao, K.H. and Miao, J.H., 2018. Coronary heart disease diagnosis using deep neural networks. international journal of advanced computer science and applications, 9(10).

[6] Chang, V., Bhavani, V.R., Xu, A.Q. and Hossain, M.A., 2022. An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics, 2, p.100016.

[7] Shukur, B.S. and Mijwil, M.M., 2023. Involving machine learning techniques in heart disease diagnosis: a performance analysis. International Journal of Electrical and Computer Engineering, 13(2), p.2177.

[8] Ramprakash, P., Sarumathi, R., Mowriya, R. and Nithyavishnupriya, S., 2020, February. Heart disease prediction using deep neural network. In 2020 international conference on inventive computation technologies (ICICT) (pp. 666-670). IEEE.

[9] Supani, A., Nurmaini, S., Partan, R.U. and Suprapto, B.Y., 2024, February. The Classification of Coronary Artery Disease Using A Machine Learning Approach: A Preliminary Study. In Proceedings of the 7th FIRST 2023 International Conference on Global Innovations (FIRST-ESCSI 2023) (Vol. 232, p.4). Springer Nature. (https://www.kaggle.com/datasets/tanyachi99/zalizadeh-sani-dataset-2csv).

[10] Hu, J., Hao, G., Xu, J., Wang, X. and Chen, M., 2024. Deep learning-based coronary artery calcium score to predict coronary artery disease in type 2 diabetes mellitus. Heliyon, 10(6).

[11] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S. and Singh, P., 2021. Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience, 2021(1), p.8387680.

[12] Ramesh, T.R., Lilhore, U.K., Poongodi, M., Simaiya, S., Kaur, A. and Hamdi, M., 2022. Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, pp.132-148.

[13] Padmaja, B., Srinidhi, C., Sindhu, K., Vanaja, K., Deepika, N.M. and Patro, E.K.R., 2021. Early and accurate prediction of heart disease using machine learning model. Turkish Journal of Computer and Mathematics Education,

12(6), pp.4516-4528.

[14] Muhammad, L.J., Al-Shourbaji, I., Haruna, A.A., Mohammed, I.A., Ahmad, A. and Jibrin, M.B., 2021. Machine learning predictive models for coronary artery disease. SN Computer Science, 2(5), p.350.

[15] Arooj, S., Rehman, S.U., Imran, A., Almuhaimeed, A., Alzahrani, A.K. and Alzahrani, A., 2022. A deep convolutional neural network for the early detection of heart disease. Biomedicines, 10(11), p.2796.

[16] Ayon, S.I., Islam, M.M. and Hossain, M.R., 2022. Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. IETE Journal of Research, 68(4), pp.2488-2507.

[17] Setiawan, N.A., Venkatachalam, P.A. and Hani, A.F.M., 2020. Diagnosis of coronary artery disease using artificial intelligence-based decision support system. arXiv preprint arXiv:2007.02854.

[18] Nazlı, B., Gültepe, Y. and Altural, H., 2020. Classification of Coronary Artery Disease Using Different Machine Learning Algorithms. Int. J. Educ. Manag. Eng, 10(4), pp.1-7.

[19] Anderies, A., Tchin, J.A.R.W., Putro, P.H., Darmawan, Y.P. and Gunawan, A.A.S., 2022. Prediction of heart disease UCI dataset using machine learning algorithms. Engineering, MAthematics and Computer Science Journal (EMACS), 4(3), pp.87-93.

[20] Garavand, A., Salehnasab, C., Behmanesh, A., Aslani, N., Zadeh, A.H. and Ghaderzadeh, M., 2022. Efficient model for coronary artery disease diagnosis: a comparative study of several machine learning algorithms. Journal of Healthcare Engineering, 2022(1), p.5359540.

[21] Cenitta, D., Arjunan, R.V. and Prema, K.V., 2022. Ischemic heart disease multiple imputation technique using machine learning algorithm. Engineered Science, 19(6), pp.262-272.

[22] Bilal, M., Aslam, N., Naeem, A. and Abid, M.K., 2023. Using Machine Learning Models for The Prediction of Coronary Arteries Disease. VFAST Transactions on Software Engineering, 11(2), pp.149-159.

[23] Ambrish, G., Ganesh, B., Ganesh, A., Srinivas, C. and Mensinkal, K., 2022. Logistic regression technique for prediction of cardiovascular disease. Global Transitions Proceedings, 3(1), pp.127-130.

[24] Omkari, D.Y. and Shaik, K., 2024. An integrated two-layered voting (TLV) framework for coronary artery disease prediction using machine learning classifiers. IEEE Access.

[25] Prusty, S., Patnaik, S. and Dash, S.K., 2022. Comparative analysis and prediction of coronary heart disease. Indonesian Journal of Electrical Engineering and Computer Science, 27(2),pp.944-953.

[26] Pasha, S.M. and Ankalaki, S., 2020. Diabetes and Heart Disease Prediction Using Machine Learning Algorithms. International Journal of Emerging Trends in Engineering Research, 8(7).

[27] Nagavelli, U., Samanta, D. and Chakraborty, P., 2022. Machine learning technology- based heart disease detection models. Journal of Healthcare Engineering, 2022(1), p.7351061.

[28] Pal, M. and Parija, S., 2021, March. Prediction of heart diseases using random forest. In Journal of Physics: Conference Series (Vol. 1817, No. 1, p. 012009). IOP Publishing.

[29] Nadeem, M.W., Goh, H.G., Khan, M.A., Hussain, M. and Mushtaq, M.F., 2021. Fusion-Based Machine Learning Architecture for Heart Disease Prediction. Computers, Materials & Continua, 67(2).

[30] Budholiya, K., Shrivastava, S.K. and Sharma, V., 2022. An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University-Computer and Information Sciences, 34(7), pp.4514-4523.

[31] Ayatollahi, H., Gholamhosseini, L. and Salehi, M., 2019. Predicting coronary artery disease: a comparison between two data mining algorithms. BMC public health, 19, pp.1-9.

[32] Sadr, H., Salari, A., Ashoobi, M.T. and Nazari, M., 2024. Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models. European Journal of Medical Research, 29(1), p.455.

[33] Viega, M.T., Marvin, E., Jayadi, R. and Mauritsius, T., 2020. Heart disease prediction system using data mining classification techniques: Naïve bayes, knn, and decision tree. International Journal of Advanced Trends in Computer Science and Engineering, 9(3), pp.3028-3035.

[34] Sayadi, M., Varadarajan, V., Sadoughi, F.,

Chopannejad, S. and Langarizadeh, M., 2022. A machine learning model for detection of coronary artery disease using noninvasive clinical parameters. Life, 12(11), p.1933.

[35] Shorewala, V., 2021. Early detection of coronary heart disease using ensemble techniques. Informatics in Medicine Unlocked, 26, p.100655.

[36] Bora, N., Gutta, S. and Hadaegh, A., 2021. Using machine learning to predict heart disease (Doctoral dissertation, California State University San Marcos).

[37] Rahimi, K., Bennett, D., Conrad, N., Williams, T.M., Basu, J., Dwight, J., Woodward, M., Patel, A., McMurray, J. and MacMahon, S., 2014. Risk prediction in patients with heart failure: a systematic review and analysis. JACC: Heart Failure, 2(5), pp.440-446.

[38] Zulkiflee, N.F. and Rusiman, M.S., 2021. Heart Disease Prediction Using Logistic Regression. Enhanced Knowledge in Sciences and Technology, 1(2), pp.177-184.

[39] Khanna, D., Sahu, R., Baths, V. and Deshpande, B., 2015. Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease. International Journal of Machine Learning and Computing, 5(5), p.414.

[40] He, Q., Maag, A. and Elchouemi, A., 2020, November. Heart disease monitoring and predicting by using machine learning based on IoT technology. In 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA) (pp. 1-10). IEEE.