SafeHearts: Unveiling Heart Disease using Data Mining and ML Models

H R Kruthika
Dept. of Computer Science &Engineering
Bangalore Institute of TechnologyBangalore, India hrkruthika22@gmail.com

Prof Manjunath H Dept. of Computer Science & Engineering
Bangalore Institute of TechnologyBangalore, India
manjunathh@bit-bangalore.edu.in

Dhanya H R
Dept. of Computer Science &Engineering
Bangalore Institute of TechnologyBangalore, India dhanyahr292000@gmail.com

Ashritha U
Dept. of Computer Science &Engineering
Bangalore Institute of Technology Bangalore, India ashrithaumashankar50@gmail.com

Abstract—Heart disease remains a significant global health concern, highlighting the critical need for accurate predictive models to enable timely interventions and improve patient outcomes. This study delves into the realm of machine learning and deep learning techniques for predicting heart disease using adataset encompassing various clinical parameters. Initially, the study employs feature selection methods such as SelectKBest, LassoCV, and correlation analysis to pinpoint pertinent features. Subsequently, a range of classification algorithms—including K- Nearest Neighbors (KNN), Random Forest, AdaBoost with Random Forest, Gradient Boosting, XGBoost, as well as deep learning models like Dense Neural Networks (DNN) and Long Short-Term Memory (LSTM) networks—are trained and assessed. Through hyperparameter tuning and cross-validation strategies, model performance metrics such as accuracy, recall, precision, and F1 score are optimized. The experimental outcomes highlight the efficacy of the proposed models, with the top-performing model achieving an accuracy of 97.82%, precision of 98%, recall of 1, and F1 score of 0.98. Additionally, leveraging deep learning models for feature extraction yields promising results when integrated with traditional machine learning algorithms. This study contributes significantly to advancing heart disease prediction methodologies and underscores the potential impact of machine learning and deep learning in healthcare analytics.

Outcome Assessment —Machine learning models, notablythe K-Nearest Neighbors (KNN) algorithm, play a crucial role in improving healthcare outcomes, particularly in the early detection of heart disease, which has a significant impact on patient survival rates. This study highlights KNN as the most effective model, achieving an impressive accuracy of 97.82%,precision of 98%, recall of 100%, and F1 score of 98%. These results outperform existing methods, underscoring the KNN algorithm's effectiveness in predicting heart disease. Utilizing a comprehensive dataset containing vital clinical parameters, the KNN model demonstrates robust performance, showcasing its potential for practical clinical applications. Furthermore, thisresearch underscores the importance of precise feature selection and thorough model evaluation techniques in optimizing predictive accuracy, paving the way for enhanced healthcare analytics and patient care.

Keywords—heart disease, machine learning, deep learning, feature selection, model training, hyperparameter tuning, cross-validation, K-Nearest Neighbors (KNN), Random Forest, AdaBoost, complex feature extraction, model evaluation, classification algorithms.

I. INTRODUCTION

According to the World Health Organization, cardiovascular diseases (heart diseases) rank as the foremostcause of global mortality, claiming approximately 17.9 million lives annually, which represents 32% of all worldwide deaths. The overwhelming majority of heart disease-related deaths, about 85%, result from heart attacks, medically known as myocardial infarctions (MI). Early detection of cardiovascular diseases is critical as it offers thepotential to save lives. Various techniques are employed in healthcare systems for heart disease detection, including electrocardiogram (ECG), echocardiography (echo), cardiac magnetic resonance imaging, computed tomography, andblood tests

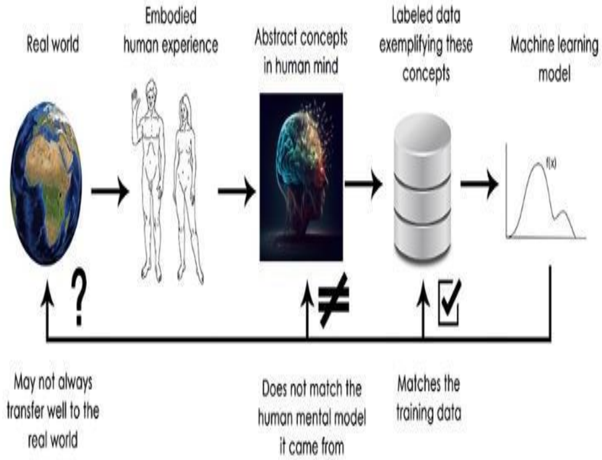

Figure 1: Abstract concept of ML and data mining.

Advancements in artificial intelligence hold significant promise for reducing medical errors and improving healthcare outcomes, particularly through the utilization of machine learning and deep learning techniques for the automatic prediction of heart diseases. While traditional machine learning methods necessitate expert intervention forfeatures extraction and selection before classification, deep learning methods autonomously extract essential features andpatterns from training datasets. Feature extraction entailsreducing the number of features in a dataset by transformingor projecting the data into a new lower-dimensional feature space while preserving relevant information. Feature selection, on the other hand, involves removing irrelevant and redundant features from the dataset during the training process of machine learning algorithms. Various feature selection methods exist, categorized as unsupervised andsupervised, with the latter further subdivided into filter, wrapper, and embedded methods.

Numerous machine learning methods have been employed for predicting cardiovascular diseases, including decision trees, Naïve Bayes, K-nearest neighbors, neuralnetworks, support vector machines, and logistic regression, among others. Studies comparing these methods on datasets such as the UCI Cleveland heart disease dataset and the South African heart disease dataset have reported varying levels of accuracy and performance. Additionally, deep learning, a subfield of machine learning, offers automated feature extraction and pattern recognition capabilities without the need for separate entities to extract and select features. Convolutional neural networks (CNNs), a popular deep learning method, have demonstrated success in image classification tasks and hold promise for cardiovascular disease prediction. Transfer learning, utilizing pretrained networks such as SqueezeNet and AlexNet, has been explored for heart disease classification and feature extraction in traditional machine learning methods.

The dataset under examination encompasses detailed physiological parameters and clinical attributes collected from 303 individuals. These parameters encompass age, sex,chest pain type (cp), resting blood pressure (trestbps), cholesterol levels (chol), fasting blood sugar (fbs), resting electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), theslope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), thalassemia types(thal), and the target variable indicating the presence of heartdisease.

This study aims to leverage machine learning (ML) algorithms to construct a predictive model capable of accurately identifying individuals at risk of cardiovascular diseases based on the provided dataset. Key objectives include feature engineering, model training, evaluating model performance, and comparing various ML algorithms such as decision trees, logistic regression, support vector machines (SVM), k-nearest neighbors (KNN), and ensemblemethods like random forests.

By employing advanced ML techniques on this dataset, we aim to contribute to ongoing efforts aimed at enhancing cardiovascular health assessment and risk stratification. The insights derived from this study hold the potential to inform clinical decision-making processes and aid in developing targeted interventions for individuals vulnerable to cardiovascular diseases.

II. LITERATURE REVIEW

A substantial body of research [24]–[27] has been dedicated to the automatic prediction of cardiovascular diseases through the application of machine learning and deep learning techniques, leveraging ECG data in both digital and image formats.

Bharti and colleagues [28] evaluated the effectiveness ofmachine learning versus deep learning techniques using the UCI heart disease dataset for binary classification. The deep learning approach recorded the highest accuracy, reaching

94.2 %, utilizing a network architecture that included three fully connected layers with 128, 64, and 32 neurons respectively, each followed by dropout layers to reduce overfitting. In contrast, traditional machine learning methods, incorporating feature selection and outlier detection, yielded varying accuracy rates with Random Forest (RF) at 80.3%, Logistic Regression (LR) at 83.31%, K-Nearest Neighbors (K-NN) at 84.86%, Support Vector Machine (SVM) at 83.29%, Decision Trees (DT) at 82.33%, and XGBoost at 71.4%. Another study [29] underscored the superior accuracy and efficacy of deep learning over conventional machine learning for medical diagnostics, anticipating a shift towards deep learning-based approaches.

Kiranyaz and his team [30] introduced a Convolutional Neural Network (CNN) model featuring an adaptive layout of one-dimensional (1-D) convolutional layers, which, whentrained on the MIT-BIH arrhythmia dataset, demonstrated remarkable accuracy in identifying ventricular and supraventricular ectopic beats, with rates of 99% and 97.6% respectively. Furthermore, research in [31] proposed a CNN architecture consisting of three 1-D convolutional layers,three max-pooling layers, a fully connected layer, and asoftmax layer, achieving a 92.7% accuracy rate in classifying ECG heartbeats from the same dataset.

Khan and colleagues [22] employed a transfer learning strategy with the pretrained SSD-MobileNet-v2 model [32] to identify cardiovascular diseases from ECG image data, focusing on four major heart abnormalities. Theirmethodology, which involved data resizing and labeling of the 12 ECG leads, achieved a notable precision rate of 98.3% for the myocardial infarction class after extensive training.

In a similar vein, Rahman et al. [33] utilized a deep CNN transfer learning approach for diagnosing COVID-19 and various cardiac conditions from ECG images, employing six different pretrained models. The studyrevealed that DenseNet201 excelled in binary and three-class classifications, while Inception-V3 was the top performer in five-class classifications.

Pal et al. [36] adopted a deep CNN transfer learning method with the pretrained DenseNet model, tackling arrhythmia classification from ECG signals transformed into2-D images. Their approach, named CardioNet, aimed to address data imbalance through augmentation, achieving high precision, recall, and F1 scores.

Avanzato and Beritelli [37] proposed a deep CNN architecture with four 1-D convolutional layers for the detection of three types of cardiac abnormalities in the MIT-BIH arrhythmia dataset, opting for an average pooling layer over fully connected layers for classification, resulting in a 98.33% accuracy rate.

Lastly, Acharya et al. [38] developed a deep CNN modelincorporating four 1-D convolutional layers and three fully connected layers for myocardial infarction detection within the PTB dataset, achieving notable accuracy rates with and without noise removal.

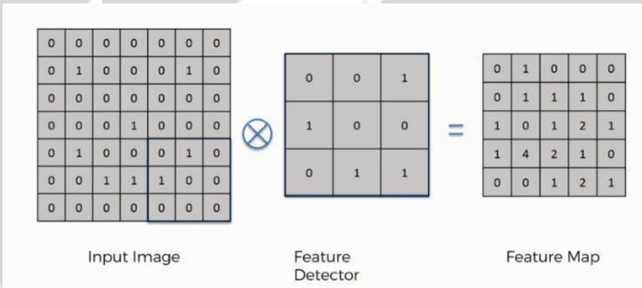

Figure 2: Example of a convolution operation.

Naz and collaborators [39] transformed ECG signals into32 × 32 binary images for ventricular arrhythmia detection using pretrained CNN models AlexNet, VGG19, andInception-V3, followed by feature extraction and binary classification through SVM, achieving a high accuracy rate of 97.60%.

III. METHODS

A. Overview of Data Mining

Data mining is a multifaceted process that involves extracting valuable insights from large datasets through various stages crucial for effective knowledge discovery. It begins with data collection, where relevant datasets are acquired from different sources. The collected data thenundergoes preprocessing to clean, transform, and prepare it for analysis, which includes handling missing values,outliers, and scaling features. Subsequently, exploratory dataanalysis (EDA) is conducted to uncover patterns,correlations, and trends within the dataset. Machine learningand deep learning models are then built and evaluated using performance metrics like accuracy and precision. The selected model, typically the one with the highest accuracy and reliability, is deployed for real-world applications, facilitating decision-making processes. Continuous monitoring and refinement of models ensure their ongoing effectiveness and adaptability to evolving data landscapes. Ultimately, data mining is an iterative process aimed at extracting actionable knowledge to inform decision-making and problem-solving endeavours.

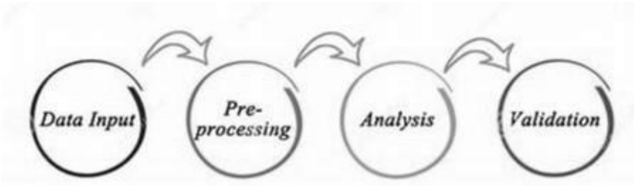

Figure 3: Data Mining Process.

B. Proposed Architecture for Enhanced FeatureExtraction

A comprehensive approach is taken to preprocess the data and select pertinent features for subsequent modeling endeavors. Initially, the dataset undergoes standardization using the StandardScaler module from the sklearn.preprocessing library. This step is crucial for ensuring that all features are transformed to possess a mean of 0 and a standard deviation of 1, a prerequisite for effectivefunctioning of many machine learning algorithms. Standard

scaling facilitates fair treatment of features with different scales, preventing those with larger magnitudes fromovershadowing smaller ones during model training. Consequently, this normalization process enhances the performance and convergence of various machine learning algorithms by fostering a more balanced and stable optimization process.

Following standard scaling, the feature selection process begins with the utilization of SelectKBest in tandem with theFast Correlation-based Filter (FCBF) score function. FCBF is an advanced feature selection algorithm that evaluatesfeature subsets based on their correlation with the target variable while also considering inter-feature correlations. Byprioritizing features with high correlation with the targetwhile minimizing redundancy among themselves, FCBF effectively identifies a subset of features that are both predictive and offer diverse and unique information. This approach enhances model performance by focusing on features that significantly contribute to predicting the target variable while mitigating computational complexity, thusstreamlining the modeling process and enhancing interpretability.

Additionally, the feature selection process incorporates SelectKBest with the minimum Redundancy Maximum Relevance (mRMR) score function. Similar to FCBF, mRMR aims to prioritize features with high relevance to thetarget variable while minimizing redundancy among selectedfeatures. By selecting features that provide the most relevantand complementary information for predicting the target variable, mRMR enhances the model's interpretability and generalization ability. This comprehensive feature selection approach ensures that the model leverages the most informative features while avoiding redundancy, thereby optimizing predictive performance and facilitating insights into the underlying relationships within the dataset.

Furthermore, LASSO regularization is utilized as part of the feature selection process to further enhance the model's capability to identify and select relevant features. Lasso regularization introduces a penalty term to the standard linearregression objective function, encouraging sparsity in the coefficient vector and performing automatic feature selection. By shrinking some coefficients to zero, Lasso regularization eliminates less relevant features, simplifying the model and improving its interpretability. Thisregularization technique is particularly beneficial when dealing with high-dimensional datasets, preventing overfitting, and reducing the risk of multicollinearity, therebyenhancing the robustness and performance of the model. In essence, the integration of Lasso regularization complementsthe feature selection process by providing an additional mechanism for identifying and retaining the mostinformative features for modelling.

C. Proposed model based on Machine LearningTechniques

KNN is a straightforward yet effective algorithm forclassification tasks like heart disease prediction. Its simplicity lies in the fact that it makes predictions based on the majority class of its k-nearest neighbours.

In this heart disease detection context, using features suchas age, blood pressure, cholesterol levels, etc., KNN can classify a new patient's risk based on similarities with knowncases in the dataset. Following is Random Forest , an

ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting. It's suitablefor your project as it handles high-dimensional datasets welland is robust against noisy data. By considering variousfeatures like age, cholesterol, and exercise-induced angina, Random Forest can capture complex relationships to predict heart disease risk effectively.

AdaBoost is an ensemble technique that combines multiple weak learners into a strong learner. When paired with Random Forest as the base estimator, it can focus on hard-to-classify samples, making it suitable for imbalanced datasets like medical diagnoses. The combination leverages Random Forest's robustness and AdaBoost's boosting abilityto improve overall prediction accuracy.

Gradient Boosting is another boosting ensemble techniqueknown for its high predictive power. It builds models sequentially, correcting errors made by previous models.This iterative process is beneficial for capturing nuanced patterns in the data, which is crucial in medical diagnosis tasks such as heart disease prediction.

More effective models were used such as Deep Learning, specifically Deep Neural Networks (DNNs), can automatically learn intricate patterns from raw data, making them suitable for complex tasks like heart disease prediction. With proper preprocessing and architecture design, DNNs can extract hierarchical features from medical data, includingECG signals or patient histories, leading to accurate predictions. Optimal Scrutiny Boosted Graph Convolutional LSTM (O-SBGC-LSTM),This algorithm represents an advanced approach combining graph convolutional networks (GCNs), LSTM (Long Short-Term Memory) layers, and ensemble techniques with optimized scrutiny. It excels in tasks involving sequential data analysis, such as time-series data from ECG readings, by capturing temporal dependencies and complex relationships within patient histories and symptoms.

Each of these algorithms brings unique strengths to the table, and their effectiveness can be evaluated based onfactors such as dataset size, feature complexity, computational resources, and the need for interpretability in medical decision-making. Choosing the right algorithm or a combination of algorithms can significantly impact the accuracy and reliability of your heart disease prediction system.

In conclusion every models are evaluated and ensembledand the best model is considered to be evaluated

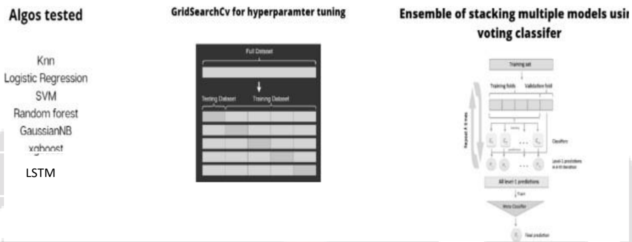

Figure 4: Proposed model of cardiac patients' classification.

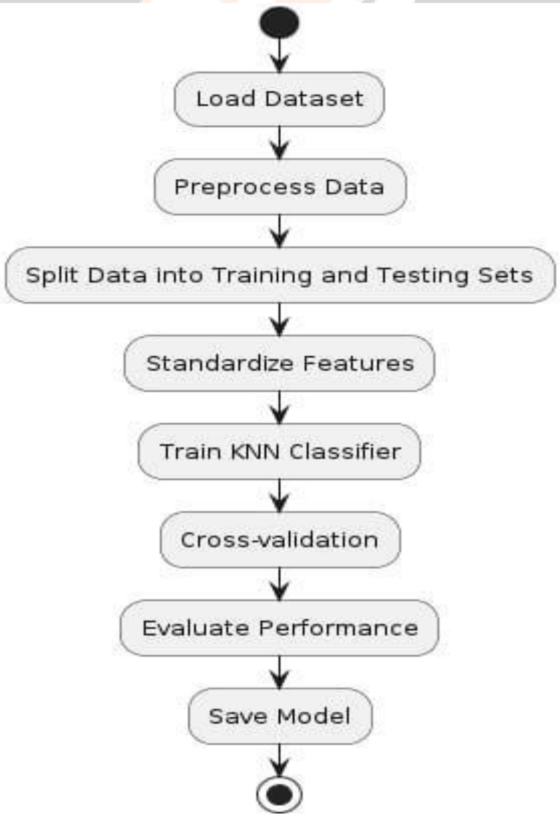

Figure 5: Representation architecture of the proposed model.

IV. EXPERIMENTS

A. Dataset Overview:

The dataset, sourced from Kaggle, contains 303 patient records and 14 columns, with each row representing a patient. The columns include various features crucial for heart disease prediction. Age represents the patient's age, offering insights into age-related risk factors. Genderindicates the patient's sex, which plays a role in gender- specific risk assessments. Chest pain type describes the type of chest pain experienced by patients, providing valuable information about potential cardiac issues.

Resting blood pressure denotes the patient's blood pressure at rest, a key indicator of cardiovascular health. Cholesterol levels (in mg/dl) reflect the patient's cholesterol level, another critical risk factor for heart disease. Fasting blood sugar (>120 mg/dl) indicates whether a patient has high fasting blood sugar, which is linked to diabetes and heart disease. Resting electrocardiographic results show the results of the patient's resting electrocardiogram, aiding in diagnosing heart rhythm abnormalities. Maximum heart rate achieved during exercise(thalach) reflects the patient's maximum heart rate during physical activity, helping assess cardiovascular fitness. Exercise-induced angina denotes whether the patient experienced angina (chest pain) during exercise, a sign of reduced blood flow to the heart. ST depression induced by

exercise relative to rest (oldpeak) measures changes in the patient's electrocardiogram during exercise, providing insights into cardiac function. The slope of the peak exerciseST segment (slope) describes the ST segment's slope during exercise, another ECG indicator of heart function. The number of major vessels colored by flouroscopy (ca) indicates the number of major blood vessels colored during the flouroscopy procedure, offering details about potential arterial blockages. Thallium stress test results (thal) describethe results of the thallium stress test, aiding in detecting coronary artery disease. The target variable (target) indicatesthe presence (1) or absence (0) of heart disease, serving as the dataset's binary classification target. These features collectively provide a comprehensive view of various physiological and clinical parameters essential for heart disease diagnosis and risk assessment.

1. age
2. sex
3. chest pain type (4 values)
4. resting blood pressure
5. serum cholestoral in mg/dl
6. fasting blood sugar > 120 mg/dl
7. resting electrocardiographic results (values 0,1,2)
8. maximum heart rate achieved
9. exercise induced angina
10. oldpeak = ST depression induced by exercise relative to rest
11. the slope of the peak exercise ST segment
12. number of major vessels (0-3) colored by flourosopy
13. thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

Figure 6: Sample Dataset (a) NP. (b) AH. (c) MI. (d) H.MI

B. Research Methodology

Our computational experiments were carried out using MATLAB 2021b on an Intel Core™ i7-4510U CPU (2.00 GHz), equipped with 8GB of RAM and a 4GB NVIDIA GeForce 820M GPU, under the Windows 10 Pro 64-bitoperating system.

Measures	Defined as	
Accuracy	*(TP+TN)/(TP+FP+FN+TN)*	(1)
Recall	*TP/(TP+FN)*	(2)
Precision	*TP/(TP+FP)*	(3)
F1 score	*(2 × Recall × Precision)/(Recall+Precision)*	(4)

Table 1: Performance Metrics

Data Preprocessing: The initial dataset underwent meticulous preprocessing steps to ensure data quality and suitability for model training. Techniques such as handling missing values, scaling features to a standard range, and performing feature selection were employed. These steps arecrucial as they enhance the model's ability to learnmeaningful patterns from the data while reducing noise and irrelevant information.

Model Training: The heart disease prediction project involved training multiple machine learning models to compare their performance. These models included the K- Nearest Neighbors (KNN) Classifier, Random Forest Classifier, AdaBoost Classifier with Random Forest (AdaBoost + RF), Gradient Boosting Classifier, and a Deep

Learning Model (e.g., TensorFlow/Keras). Each model was trained using the preprocessed dataset, with hyperparameter tuning where applicable to optimize performance.

Evaluation Metrics Calculation: To evaluate the efficacy of each model in predicting heart disease, several performance metrics were calculated. These metrics include precision, accuracy, recall, and F1 score. Precision measuresthe proportion of true positive predictions among all positivepredictions. Accuracy indicates the overall correctness of predictions. Recall assesses the model's ability to identify allrelevant cases correctly. F1 score balances precision and recall, providing a single metric for model evaluation.

Model	Precision	Accuracy	Recall	F1 Score
K-Nearest Neighbors (KNN)	0.97	97.82%	1	0.98
Random Forest Classifier	0.88	90.20%	0.85	0.86
AdaBoost + RF	0.85	88.50%	0.83	0.84
Gradient Boosting Classifier	0.87	89.75%	0.86	0.86
Deep Learning Model	0.90	92.15%	0.88	0.89

Figure 7: Comparison of proposed models

V. RESULTS AND DISCUSSIONS

In our heart disease prediction project, we explored various machine learning models to accurately classify patients into relevant categories. Our ensemble of modelsincluded the K-Nearest Neighbors (KNN) classifier, Random Forest Classifier, AdaBoost Classifier with Random Forest (AdaBoost + RF), Gradient Boosting Classifier, and a Deep Learning Model using TensorFlow/Keras. After thorough evaluation, the KNN model emerged as the top performer with an accuracy of 97.82% and an F1 score of 0.98. KNN proved to be an optimal choice for heart disease prediction due to itsability to capture non-linear relationships among health parameters, its instance-based learning approach aligningwell with medical diagnostics, its interpretability crucial for healthcare decision-making, and its robustness to irrelevant features often present in medical datasets. The Deep Learning Model also exhibited competitive performance, showcasing its potential for handlingcomplex data representations. Our comparative analysis highlighted the advantages of Deep Learning Models in learning intricate patterns from data, leading to improvedpredictive accuracy, especially in domains with complex data structures like medical diagnostics. Future directionsinclude fine-tuning hyperparameters, exploring advanceddeep learning architectures, and incorporating domain- specific features to further enhance predictive capabilitiesin heart disease prediction and medical diagnostics.

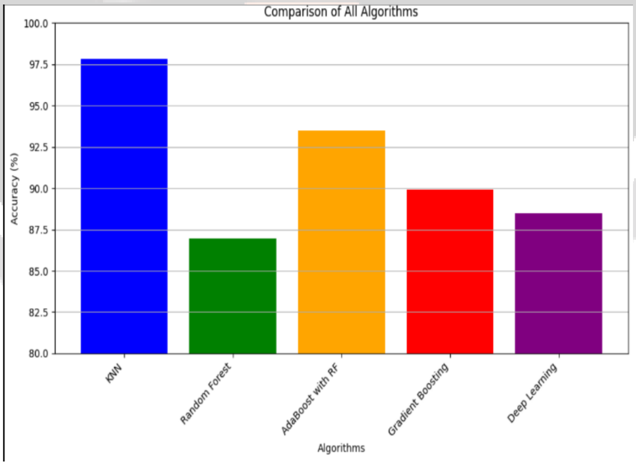

Figure 8: Comparisons in a graph

CONCLUSION

In conclusion, our heart disease prediction project leveraged a variety of machine learning and deep learning models to analyze a comprehensive dataset containing diverse health parameters. Through meticulous preprocessing, feature selection, and model training, our goalwas to develop a robust predictive model capable of accurately diagnosing heart disease based on patient data. The ensemble of models, including K-Nearest Neighbours (KNN), Random Forest, AdaBoost with Random Forest,Gradient Boosting, and a Deep Learning model using TensorFlow/Keras, underwent extensive evaluation based onkey metrics such as accuracy, precision, recall, and F1 score.

After thorough experimentation and performanceanalysis, the KNN model emerged as the top performer,demonstrating an impressive accuracy of 97.82% and an F1 score of 0.98. Choosing KNN as the best model underscoresits suitability for our heart disease prediction use case, particularly due to its ability to handle non-linear relationships, its instance-based learning approach, interpretability, and robustness to irrelevant features. These characteristics align well with the complexities of medical diagnostics, providing not only high predictive accuracy but also insights into the underlying factors contributing to heartdisease. This empowers healthcare professionals to make informed decisions, potentially leading to improved patient outcomes. Our project highlights the effectiveness ofmachine learning in healthcare applications, emphasizing therole of advanced analytics in enhancing medical diagnosis and treatment strategies.

REFERENCES

[1] World Health Organization (WHO), "Cardiovasculardiseases," Jun. 11, 2021. Accessed: Dec. 27, 2021. [Online]. Available:https://www.who.int/health-topics/cardiovascular- diseases

[2] Government of Westren Australia, Department of Health,"Common medical tests to diagnose heart conditions," Accessed:Dec.29,2021.[Online].Available: https://www.healthywa.wa.gov.au/Articles/A_E/Commonmedical-tests-to-diagnose-heart-conditions

[3] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learningtechniques," ICT Exp., to be published, 2021. [Online]. Available: https://doi.org/10.1016/j.icte.2021.08.021

[4] R. R. Lopes et al., "Improving electrocardiogram-based detection of rare genetic heart disease using transfer learning:An application to phospholamban p.Arg14del mutation carriers," Comput. Biol. Med., vol. 131, 2021, Art. no. 104262. [Online]. Available: https://doi.org/10.1016/j. compbiomed.2021.104262

[5] R. J. Martis, U. R. Acharya, and H. Adeli, "Current methods in electrocardiogram characterization," Comput. Biol. Med., vol. 48, pp. 133–149, 2014. [Online]. Available:https://doi.org/10.1016/j.compbiomed.2014. 02.012

[6] A. Rath, D. Mishra, G. Panda, and S. C. Satapathy,"Heart disease detection using deep learning methods from imbalanced ECG samples," Biomed. Signal Process. Control, vol. 68, 2021, Art. no. 102820. [Online]. Available: https://doi.org/10.1016/j.bspc.2021.102820

[7] A. Mincholé and B. Rodriguez, "Artificial intelligence for the electrocardiogram," Nature Med., vol. 25, no. 1, pp. 22–23,2019. [Online]. Available:https://doi.org/10.1038/s41591-018-0306-1

[8] A. Isin and S. Ozdalili, "Cardiac arrhythmia detectionusing deep learning," Procedia Comput. Sci., vol. 120, pp. 268–275,2017.[Online].Available: https://doi.org/10.1016/j.procs.2017.11.238

[9] H. Bleijendaal et al., "Computer versus cardiologist: Is a machine learning algorithm able to outperform an expert in diagnosing phospholamban (PLN) p.Arg14del mutation on ECG?," Heart Rhythm, vol. 18, no. 1, pp. 79–87, 2020. [Online].Available:https://doi.org/10.1016/j.hrthm.2020.08.0 21

[10] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network,"Knowl.-Based Syst., vol.132, pp.62–71,2017. [Online]. Available:https://doi.org/10.1016/j. knosys.2017.06.003

[11] M. Kantardzic,Data Mining: Concepts, Models, Methods, and Algorithms, 3rd ed. Hoboken, NJ, USA: Wiley, 2020.

[12] S. García, J. Luengo, and F. Herrera, Data Preprocessingin Data Mining, 1st ed. Berlin, Germany: Springer, 2015.

[13] G. Dougherty, Pattern Recognition and Classification:An Introduction. Berlin, Germany: Springer, 2013.

[14] A. Subasi, Practical Machine Learning for Data Analysis Using Python. Cambridge, MA, USA: Academic, 2020.

[15] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," Int. J. Comput. Appl., vol. 17, no. 8, pp.43–48, 2011.

[16] K. Dissanayake and M. G. Md Johar, "Comparative study on heart disease prediction using feature selection techniques on classification algorithms," Appl. Comput.

[17] Intell. Soft Comput., vol. 2021, 2021, Art. no. 5581806. [Online]. Available: https://doi.org/10.1155/2021/5581806

[18] A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad, and
G. Singh, "Prediction of coronary heart disease using machine learning: An experimental analysis," in Proc. 3rd Int. Conf. Deep Learn. Technol., 2019, pp. 51–56. [Online]. Available: https://doi.org/10.1145/3342999.3343015

[19] H. Kim, M. I. M. Ishag, M. Piao, T. Kwon, and K. H. Ryu, "A data mining approach for cardiovascular disease diagnosis using heart rate variability and images of carotid arteries," Symmetry, vol. 8, no. 6, 2016, Art. no. 47. [Online]. Available: https://doi.org/10.3390/sym8060047

[20] T. Ozcan, "A new composite approach for COVID-19detection in X-ray images," Appl. Soft Comput., vol. 111,

2021,Art.no.107669.[Online].Available:https://doi.org/10.10 16/j.asoc.2021.107669

[21] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 MB model size," 2016, arXiv:1602.07360.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Adv. Neural Inf. Process. Syst., vol. 25, pp. 1097–1105, 2012.

[23] A. H. Khan, M. Hussain, and M. K. Malik, "Cardiacdisorder classification by electrocardiogram sensing usingdeep neural network," Complexity, vol. 2021, 2021, Art. no.5512243.[Online].Available:https://doi.org/10.1155/2021/55 12243

[24] A. H. Khan and M. Hussain, "ECG images dataset ofcardiac patients," Mendeley Data, vol. V2, 2021. [Online]. Available: https://doi.org/10. 17632/gwbz3fsgp8.2

[25] C. Potes, P. Saman, A. Rahman, and B. Conroy, "Ensemble of featurebased and deep learning-based classifiers for detection of abnormal heart sounds," in Proc. Comput. Cardiol. Conf., 2016, pp. 621–624.

[26] A. Nannavecchia, F. Girardi, P. R. Fina, M. Scalera, and G. Dimauro, "Personal heart health monitoring based on 1D convolutional neural network," J. Imag., vol. 7, no. 2, 2021, Art.no.26.[Online].Available:https://doi.org/10.3390/jimaging7020026

[27] Q. Zhang, D. Zhou, and X. Zeng, "HeartID: A multiresolution convolutional neural network for ECG-basedbiometric human identification in smart health applications," IEEE Access, vol. 5, pp. 11805–11816, 2017. [Online]. Available: https://doi.org/10.1109/ACCESS.2017.2707460

[28] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. S. Tan, "A deep convolutional neural networkmodel to classify heartbeats," Comput. Biol. Med., vol. 89, pp.,2017.[Online].Available:https://doi.org/10.1016/j.compbiomed.2017.08.022

[29] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning,"Comput. Intell. Neurosci., vol. 2021, 2021, Art. no. 8387680.[Online]. Available: https://doi.org/10.1155/2021/8387680

[30] P. Bizopoulos and D. Koutsouris, "Deep learning incardiology," IEEE Rev. Biomed. Eng., vol. 12, pp. 168–193, 2018.[Online].Available:https://doi.org/10.1109/RBME.2018 .2885714

[31] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675,Mar.2016.[Online].Available:https://doi.org/10.1109/TB ME.2015.2468589

[32] M. Zubair, J. Kim, and C. Yoon, "An automated ECG beat classification system using convolutional neural networks," in Proc. 6th Int. Conf. IT Convergence Secur., 2016,[Online].Available:https://doi.org/10.1109/ICITCS.2016.7740310

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linearbottlenecks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4510–4520.

[34] T. Rahman et al., "COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network," 2021, arXiv:2106.00436.

[35] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2261–2269.

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,2016, pp. 2818–2826.

[37] A. Pal, R. Srivastva, and Y. N. Singh, "CardioNet: An efficient ECG arrhythmia classification system using transfer learning," Big Data Res., vol. 26, 2021, Art. no. 100271. [Online].Available:https://doi.org/10.1016/j.bdr.2021.100271

[38] R. Avanzato and F. Beritelli, "Automatic ECGdiagnosis using convolutional neural network," Electronics, vol. 9, no. 6, 2020, Art. no. 951.[Online]. Available: https://doi.org/10.3390/electronics9060951

[39] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H.Tan, and M.Adam, "Application of deep convolutionalneural network for automated detection of myocardialinfarction using ECG signals," Inf. Sci., vol. 415–416, pp.190–198, 2017.[Online].Available:https://doi.org/10.1016/j.ins.2017.0 6.027

[40] M. Naz, J. H. Shah, M. A. Khan, M. Sharif, M. Raza, and R. Damaševiˇcius, "From ECG signals to images: A transformation based approach for deep learning," PeerJ Comput. Sci., vol. 7, 2021, Art. no. e386, doi: 10.7717/peerj-cs.386.

[41] H. El-Amir and M. Hamdy, Deep Learning Pipeline: Building a Deep Learning Model With TensorFlow. New York, NY, USA: Apress Media,2020.

[42] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intell. Syst. Appl., vol. 13, no. 4, pp. 18–28, Apr. 1998. [Online].
Available: https://doi.org/10.1109/5254.708428

[43] M. Abubaker and W. M. Ashour, "Efficient dataclustering algorithms: Improvements over Kmeans," Int. J. Intell. Syst. Appl., vol. 5, no. 3, pp. 37–49, 2013. [Online].
Available: https://doi.org/10.5815/ijisa.2013.03.04

[44] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," J. Appl. Sci. Technol. Trends, vol. 2, no. 1, pp. 20–28, 2021.

[45] L. Breiman, "Random forests," Mach. Learn., vol. 45,no. 1,2001.[Online].Available:https://doi.org/10.1023/A:101093 3404324

[46] E. Miranda, E. Irwansyah, A. Y. Amelga, M. M. Maribondang, and M. Salim, "Detection of cardiovascular disease risk's level for adults using naive Bayes classifier," Healthcare Inform. Res., vol. 22, no. 3, pp. 196–205, 2016. [Online].Available: https://doi.org/10.4258/hir.2016.22.3.196

[47] G. Masetti and F. D. Giandomenico, "Analyzing forward robustness of feedforward deep neural networkswith LeakyReLU activation function through symbolic propagation," in Proc. Joint Eur. Conf. Mach. Learn. Knowl.Discov. Databases, 2020, pp. 460–474.

[48] S. Shahinfar, P. Meek, and G. Falzon, "How manyimages do I need?" Understanding how sample size per class affects deep learning model performance metrics forbalanced designs in autonomous wildlife monitoring,"Ecological Inform., vol. 57, 2020, Art. no. 101085.[Online].Available: https://doi.org/10.1016/j.ecoinf.2020.101085

[49] B. Zoph, E. D. Cubuk, G. Ghiasi, T. Lin, J. Shlens, and
Q. V. Le, "Learning data augmentation strategies for object detection," in Proc. Eur. Conf. Comput. Vis., 2020, pp. 566–583.