OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbriviate Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.12679792

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

Early Age Heart Disease Prediction: A Comprehensive Survey

Prof. Samatha R Swamy[1], Purushotam Kumar Singh[*2], Prateek Bajpai[3], Anuj S Rakaraddi[4], Sachin D[5]

[1]Assistant Professor, Department of Information Science and Engineering, RV Institute of Technology and Management, Visvesvaraya Technological University, Bengaluru, Karnataka, India.
[2,3,4,5] Student, Department of Information Science and Engineering, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Swamy, S. R., Singh, P. K., Bajpai, P., Rakaraddi, A. S., & Sachin, D. (2024). Early Age Heart Disease Prediction: A Comprehensive Survey. *Indiana Journal of Multidisciplinary Research, 4*(3), 198-204.

Abstract: Heart disease poses a significant threat to global health, necessitating accurate and early prediction methods. This study investigates the efficacy of various machine-learning algorithms for early-age heart disease prediction. Various distinct research efforts were synthesized to provide comprehensive insights. Deep Neural Network achieved the highest accuracy of 98.15 percent, outperforming previous studies with notable sensitivity and precision. A hybrid genetic algorithm and particle swarm optimization optimized approach, GAPSO-RF, was developed to enhance heart disease prediction accuracy. GAPSO-RF employs multivariate statistical analysis, modified genetic algorithm, and particle swarm optimization for feature selection, yielding improved accuracy, specificity, sensitivity, and area under the receiver. This comprehensive analysis underscores the significance of machine learning in early-age heart disease prediction, offering advantageous insights for healthcare professionals to design effective medical interventions and preventive measures.

Keywords: Heart Disease, Health Care, Machine Learning, Naive Bayes, Decision Tree Classifier, SVM, K-Nearest Neighbor, Logistic Regression, Random Forest

INTRODUCTION

Machine Learning is a strategy for gaining knowledge from the prior research experience. The conventional methods encounter several issues to solve all the problems pointed out by various researchers. The purpose of data assessment is to evaluate genuine models, which necessitates using a dependable, robust, and trustworthy framework, including such Machine Learning methods. The Machine Learning method likes to work with immediate input from the training sample after learning the foundational patterns in the data [1]. The possible outcome of the whole training phase is an automatic framework. The proposed framework is best for static and dynamic datasets. A prediction of data is an outcome of the training and testing stage. A testing stage utilizes some information collection, i.e., datasets for training, datasets for testing, and some specific types of ML models, i.e., classification or clustering models.[2-3].

The increasing population, changing life patterns, new diseases, pandemics, and changes in the natural environment have surely changed the way we take care of our health. Maintaining good health is a leading factor related to any individual's efficiency and It has always been a big concern for all the countries of the world. Providing a good medical service across the whole country proves to be challenging for developing and poor countries. The condition of medical systems across the world is different for different countries. The ratio of healthcare workers (doctors especially) to the population is a good way to determine the strength of any healthcare system. Developed countries (the USA and Western Europe) have this ratio to a good level (25:10000), while Southeast Asian countries like India are way below the world average (India has a ratio of 12.21 to 10000, whereas the world average is closer to 10:1000) [4].

Increasing this ratio of doctors to patients is a complex process and takes a lot of time in order of years or maybe decades and medical buildings equipped with modern machinery are needed. Accurately predicting the onset of heart disease stands as a crucial and formidable challenge in contemporary medicine. It remains the leading cause of mortality in numerous developed nations, with approximately one in four individuals succumbing to it each year in the United States alone. The debilitating effects of cardiovascular disease are profound, particularly among adults and the elderly, as it impairs blood vessel function and leads to coronary artery infections.[5–8].

Upon investigating the need for medical professionals in current scenarios, which seems to be a long-term aim to achieve and the deadliness and seriousness of heart disease, this paper is developed. The overall objective of this research includes being able to predict the presence or absence of heart disease by utilizing a few important medical and common attributes. Attributes considered form the primary basis for tests and give accurate results more or less.

- We have created a machine learning model so that based on his/her previous medical records, he/she

***Corresponding Author:** Purushotam Kumar Singh

can determine whether they are symptomatic of any heart disease or not.

- Models made are trained with various machine learning algorithms. Multiple algorithms used in this research are measured for their accuracy and suitable algorithms are used in the end result. Our basic aim will be to increase the accuracy of models that already exist and to detect them at a very early age as possible.

The main benefit of this research was to use modern machine learning techniques to construct an intuitive medical prediction system for the diagnosis of heart disease. Different types of ML classifier algorithms were trained, including logistic regression (LR), K-nearest neighbors (K-NN), support vector machine (SVM), Decision Tree and Random Forest, to select the best predictive model for accurate heart disease detection at an early stage.

FILTERS

EXPORT DATA in CSV format:
Right-click here & Save link

Last updated: 2023-01-12

Location	Medical doctors (per 10,000)	Medical doctors (number)	Generalist medical practitioners (number)	Specialist medical practitioners (number)	Medica
United States of America					
2020	35.55	1194267	135708	1058559	

Figure 1: The Ratio of Doctors Per 10000 Population in the USA

FILTERS

EXPORT DATA in CSV format:
Right-click here & Save link

Last updated: 2023-01-12

Location	Medical doctors (per 10,000)	Medical doctors (number)	Medical doctors not further defined (number)
India			
2020	7.265	1014538	1014538

Figure 2: Ratio of Doctors Per 10000 Population in India

LITERATURE SURVEY

Researches nowadays are focused on FS, prediction, and increasing heart-disease prediction accuracy. This section overviews the recently published related research. By utilizing active learning techniques, the study in Ibrahim et. al [8] aims to generalize the concepts of the outcomes and findings, allowing for their interpretation in new cases without the requirement of being a valid instance of the training, test, or validation samples. This approach seeks to enhance the efficiency and accuracy of heart disease diagnosis by leveraging the insights gained from previous data to inform predictions and decisions in real-time scenarios. Ultimately, the goal is to develop a robust and adaptable diagnostic framework that can effectively identify and address heart disease in diverse patient populations.

Mohamed *et al.* [9] developed an efficient, hybrid genetic algorithm (GA) and particle swarm optimization (PSO) approach based on random forest (RF) for optimizing the FS process to select the crucial features that increase the accuracy of heart-disease diagnosis. It outperformed the existing state-of-the-art methods on the same datasets. Furthermore, a comparative analysis is performed between GAPSO-RF and conventional GA and found that the conventional GA was outperformed by the proposed approach. It outperformed the existing state-of-the-art methods on the same datasets. A variety of ML methods, including Stochastic Gradient Descent, KNN, Naive Bayes, Support Vector Machine, AdaBoost, Decision Tree Classifier, J48, JRIP, and others, were utilized to classify and predict heart disease source data[18].

In research [19], the researcher proposed an efficient method utilizing the Back-Propagation (BP) feature extractor of an Artificial Neural Network (ANN) from an online heart disease database for identifying the presence of heart disease. In a subsequent study [14], machine learning methods incorporating ANN were employed to predict heart disease cases. The researchers developed an automated application to anticipate heart disease sensitivity based on primary symptoms such as disease period, gender, heartbeat rate, and history. The results were able to depict that the ANN model exhibited the highest accuracy and precision among various ML models for heart disease forecasting. In research [20], a hybrid algorithm was proposed to accurately predict and classify within different age groups, the risk of heart disease. This study introduced an automated model for precise data analysis using ANN. The model proposed, accurately predicts heart conditions in the early stages. It has been shown that an individual's risk profile using machine learning techniques such as decision trees, k-nearest neighbors, naive Bayes, and genetic algorithms can be more effective when incorporating features and variations of the methods mentioned above.

***Corresponding Author:** Purushotam Kumar Singh

Table 1: Summary of the Related Work and Comparison of Various Existing Machine Learning Methods

Reference	Methods and Techniques	Key findings	Limitations
[8]	Active learning	Improving accuracy and F-score of 57.4 ± 4% and 62.2 ± 3.6%, respectively.	Discretizing the numeric values of features, categorization, and binning levels using advanced metaheuristic algorithms for fine-tuning the predictive models' parameters.
[9]	Hybrid genetic algorithm (GA) and particle swarm optimization (PSO) approach based on random forest (RF).	Improved accuracy of 95.6% and 91.4% on the Cleveland and Statlog datasets, respectively	Classifiers should be evaluated to have a more extensive evaluation of the results, High computational cost and temporal complexity
[12]	Machine Learning methods	Higher Accuracy	Limited features dataset
[13]	The statistical model X2 was used with DNN and ANN	Predictions were aligned using clinical data parameters	Accuracy and time improvement can be done
[14]	Naive Bayes and SVM were used as classifiers	Classification of heart disease dataset, cause of heart disease, diabetes	Features selection and classification perform slower
[15]	NB + (Backward Elimination / Optimize Selection / Forward Selection).	(NB+ forward selection) outperforms other feature selection approach.	The classifier algorithm needs to improve.
[16]	Long short-term memory (LSTM) neural network.	The result shows that the ADAM optimizer is significantly better than SGD.	The feature selection algorithm needs to improve. Moreover, the classification accuracy still needs to increase.
[17]	Fuzzy classifiers.	The results outperform the other methods compared.	The authors did not mention the parameters for genetic algorithms. Moreover, the classification accuracy still needs to increase.
[18]	Feature selection algorithm (FCMIM), SVM.	Improved accuracy results for heart disease dataset.	Perform better only for small datasets.
[19]	Machine Learning,iPSCs, and omics	Better throughput	Only a few parameters were implemented.
[20]	Machine Learning method,i.e. logistic regression, SVM.	Accuracy (85%), sensitivity (89%), and Specificity (81 %).	Accuracy can be improved.

Table 2: Heart Disease Dataset Attributes and Information

1	Age	Age in years	Numeric values
2	Sex/Gender	Gender type, i.e., Male or Female	0: Male, 1: Female
3	CP	Chest Pain Level	Four types of Chest pain [1: typical, 2: atypical, 3: pain, 4: asymptotic]
4	Trestbps	Blood pressure value at the time of rest	less than or greater than 120 Mg/DL
5	Chol	Represents the level of Serum Cholesterol	Numeric values
6	FBS	Represents the level of sugar in fasting blood	Numeric values
7	Restecg	Represents the level of Resting electrocardiographic	Numeric Values
8	Thalach	Maximum heart rate level	Numeric values
9	Exang	Exercise-induced level	Yes / No.
10	Oldpeak	ST level during the workout, compared with the results of rest taken	Numeric values.
11	Slope	level of peak exercise in the ST segment	Three values (1: up, 2: flat, 3: down)
12	CA	Represents the number of fluoroscopy vessels	Four values (0 to 3)
13	Thal	Used for Defect classes	[6(non-fixable), 7(reversible), otherwise: 3]
14	Class	Representing the Target	1: High and 0: Low chances

MATERIALS AND METHODS

Data Acquisition and Preprocessing

The dataset utilized in this research for early-age heart disease prediction is sourced from the UCI Machine Learning database and a shortened version of the Kaggle dataset. It comprises 76 characteristics, including a class attribute, for a total of 1025 patients from various healthcare institutions. However, only a subset of 14 features is employed in this study. The

***Corresponding Author:** Purushotam Kumar Singh

dataset is structured with 14 columns, featuring 13 independent components and one dependent target variable, which categorizes individuals into those with and without heart disease. Notably, there are no values that are missing in the dataset. The highlight of the study involves utilizing a subset of related features for the precise prediction of heart disease in early-age individuals.[22]. Using a Kaggle dataset containing heart disease reports across various age groups is justified for predicting early-age heart disease due to its rich data source encompassing diverse demographics. This dataset allows for a comprehensive analysis of risk factors, symptoms, and prevalence specific to younger populations, providing a baseline for comparison and training machine learning models.

Dataset intuition reveals that it can inform targeted interventions and preventive measures to mitigate heart disease incidence in early-age individuals. After downloading the dataset, available on the Kaggle website, and processed from UCI's dataset, it is important to clean the data by addressing any missing values or noise present. The data is imported from downloaded CSV files using the Panda's library in Python and stored in RAM as a data frame optimized for handling two-dimensional array data. There are no null values present in the dataset. Thirteen attributes are used for prediction, while the last attribute, "target," indicates whether or not an individual is suffering from heart disease. The statistics of the processed dataset can be obtained using the heart Data. Info () and heart Data. Describe () commands.

Data Analysis and Visualizations:

We will generate and understand the correlation between our attributes and the target class. A correlation matrix will be generated and plotted using Matplotlib (Fig.4). For a correlation matrix, the more positive the value of correlation, the more increase in the value of one variable causes the other to increase i.e. more directly proportional. The higher a negatively correlated variable gets, the lower the value of the target becomes. Using the correlation matrix, we discover the target to be positively correlated to chest pain significantly. This arrival should also be obvious as the greater amount of chest pain means greater risk in the heart. Max Heart rate is also significantly correlated to Target for the reason that healthier hearts do not need to become elevated much for blood supply. It simply means a higher heart rate and a higher risk of heart disease. A positive correlation can be observed between those with thalassemia. Since it is a 3-valued ordinal, where 3 indicates normal, 6 to 7 defects. Hence being in the normal category is better. The presence of a negative correlation between the target and angina can also be observed. Exercise causes muscles to crave more oxygen, in turn boosting heartbeat, while narrowed-down arteries would act as a blockage.

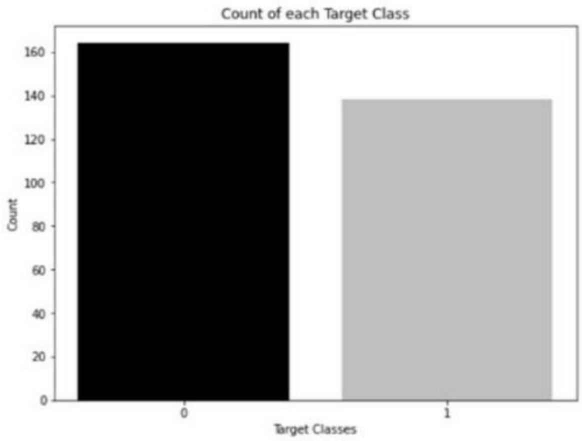

Figure 3: Categorization of a dataset based on Target Class

Checking Data Distribution: In the context of early-age heart disease prediction, understanding the data pattern is crucial for ensuring precise predictions about heart disease levels. Within the heart disease dataset, positive cases account for 54%, while negative cases represent 45%. However, to prevent overfitting issues, dataset balancing may be a necessity. Figure 3 illustrates the representation of the heart disease dataset's statistics, with "1" indicating data associated with heart disease and "0" representing data without heart disease. Specifically, the dataset comprises 165 candidates with heart disease and 138 without any heart disease, thus depicting the distribution of positive and negative heart disease cases. This balanced representation of data will assist machine learning methods in identifying the most effective patterns for the prediction process in the context of early-age heart disease.

```
corr = heartData.corr()
corr.style.background_gradient(cmap='coolwarm')
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1.000000	-0.103240	-0.071966	0.271121	0.219823	0.121243	-0.132696	-0.390227	0.088163	0.208137	-0.169105	0.271551	0.072297	-0.229324
sex	-0.103240	1.000000	-0.041119	-0.078974	-0.198258	0.027200	-0.055117	-0.049365	0.139157	0.084687	-0.026666	0.111729	0.198424	-0.279501
cp	-0.071966	-0.041119	1.000000	0.038177	-0.081641	0.079294	0.043581	0.306839	-0.401513	-0.174733	0.131633	-0.176206	-0.163341	0.434854
trestbps	0.271121	-0.078974	0.038177	1.000000	0.127977	0.181767	-0.123794	-0.039264	0.061197	0.187434	-0.120445	0.104554	0.059276	-0.138772
chol	0.219823	-0.198258	-0.081641	0.127977	1.000000	0.026917	-0.147410	-0.021772	0.067382	0.064880	-0.014248	0.074259	0.100244	-0.099966
fbs	0.121243	0.027200	0.079294	0.181767	0.026917	1.000000	-0.104051	-0.008866	0.049261	0.010859	-0.061902	0.137156	-0.042177	-0.041164
restecg	-0.132696	-0.055117	0.043581	-0.123794	-0.147410	-0.104051	1.000000	0.048411	-0.065606	-0.050114	0.086086	-0.078072	-0.020504	0.134468
thalach	-0.390227	-0.049365	0.306839	-0.039264	-0.021772	-0.008866	0.048411	1.000000	-0.380281	-0.349796	0.395308	-0.207888	-0.098068	0.422895
exang	0.088163	0.139157	-0.401513	0.061197	0.067382	0.049261	-0.065606	-0.380281	1.000000	0.310844	-0.267335	0.107849	0.197201	-0.438029
oldpeak	0.208137	0.084687	-0.174733	0.187434	0.064880	0.010859	-0.050114	-0.349796	0.310844	1.000000	-0.575189	0.221816	0.202672	-0.438441
slope	-0.169105	-0.026666	0.131633	-0.120445	-0.014248	-0.061902	0.086086	0.395308	-0.267335	-0.575189	1.000000	-0.073440	-0.094090	0.345512
ca	0.271551	0.111729	-0.176206	0.104554	0.074259	0.137156	-0.078072	-0.207888	0.107849	0.221816	-0.073440	1.000000	0.149014	-0.382085
thal	0.072297	0.198424	-0.163341	0.059276	0.100244	-0.042177	-0.020504	-0.098068	0.197201	0.202672	-0.094090	0.149014	1.000000	-0.337838
target	-0.229324	-0.279501	0.434854	-0.138772	-0.099966	-0.041164	0.134468	0.422895	-0.438029	-0.438441	0.345512	-0.382085	-0.337838	1.000000

Figure 4: Maximum Positive Correlated Features Are Cp and Thalach and Maximum Negative Correlated Features Are Exang and Old Peak.

Recommended Techniques:

1) Logistic Regression: For our model, we use Logistic regression to know the factors of risk for heart disease and forecast the probability of disease occurrence based on risk factors. This model is most frequently applied for classification, primarily two-category issues (that is, the output is categorized as two types, each representing one category), and can indicate the probability of occurrence of each classification event(Fig.5). Logistic Regression has an accuracy of 82%.

```
logistic_model = LogisticRegression()
logistic_model.fit(X_train.values, y_train.values)
logistic_model_prediction=logistic_model.predict(X_test.values)
print(accuracy_score(y_test.values,logistic_model_prediction))
print(classification_report(y_test.values,logistic_model_prediction))

0.8131868131868132
              precision    recall  f1-score   support

           0       0.85      0.71      0.77        41
           1       0.79      0.90      0.84        50

    accuracy                           0.81        91
   macro avg       0.82      0.80      0.81        91
weighted avg       0.82      0.81      0.81        91
```

Figure 5: Logistic Model Result

2) K-nearest neighbors:

K-nearest neighbors is an algorithm having capabilities of both classification and regression. When the input data point is provided, with the help of all the training data, it calculates the distance, using Euclidean, Manhattan, etc. distance metrics. It finds k examples closest to the input. Based on a majority vote, it predicts the target result for the input. It is a non-parametric algorithm. (Fig.6). K-nearest neighbors have an accuracy of 64%.

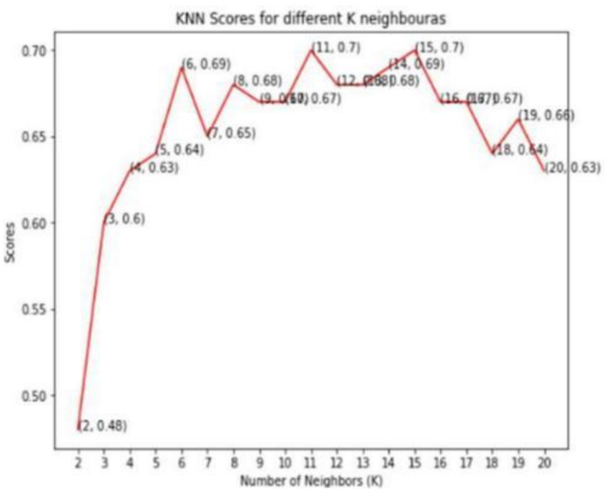

Figure 6: Classification Report of KNN

3) SVM: It belongs to the supervised machine learning algorithm. That is an algorithm, that can be used for both classification and regression challenges. SVM is mostly used in classification problems. In the SVM algorithm, we set each data object as a point in the n-dimensional space (where several attributes are represented by n) by the value of each element which is the value of a particular combination. Then, we do the splitting by finding a hyper-plane that separates the two sections very well(Fig.7). Linear Kernel has an accuracy of 79%.

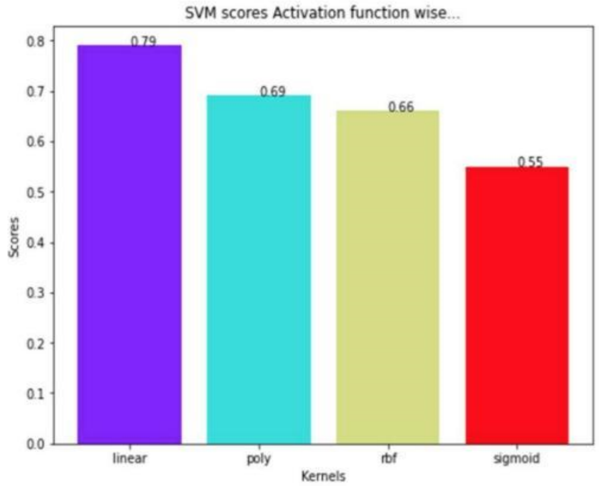

Figure 7: SVM Scores Against Various Kernels

4) Decision Tree: A decision tree contains a flowchart-like structure. In the structure of DT (Decision Tree) in a test, an attribute is represented using an internal node. The result of the test is represented using the branch. To represent a class label leaf node is used. To represent classification rules, root to leaf path is used. This is the type of analysis in which for A visual and analytical root, closely related influence diagrams are used, where calculation is done for potential attributes of advantage(Fig.8). The Decision Tree has an accuracy of 74%.

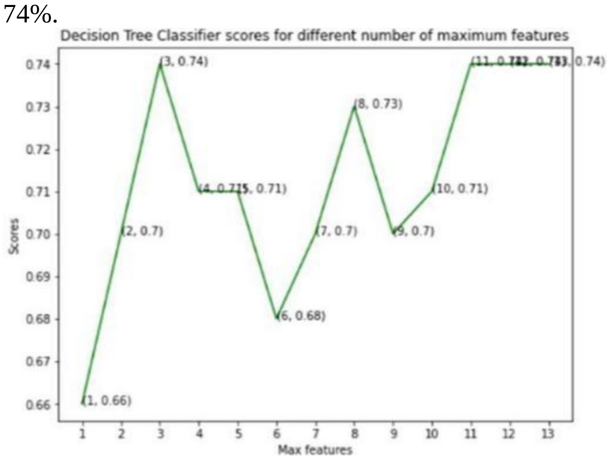

Figure 8: Decision Tree Result Graph

5) Random Forest Classifier: It is based on the decision Tree. It has a group of various Decision trees. Various decision trees are used internally. When we try to make a classification for the given Input data we feed same same data into all Decision trees. Now we collect all the votes from Decision trees and the majority of votes enter results for input. (Fig.9). Random Forest Classifier has an accuracy of 82% with 100 estimators. An increase in the number of estimators will increase the calculation complexities significantly.

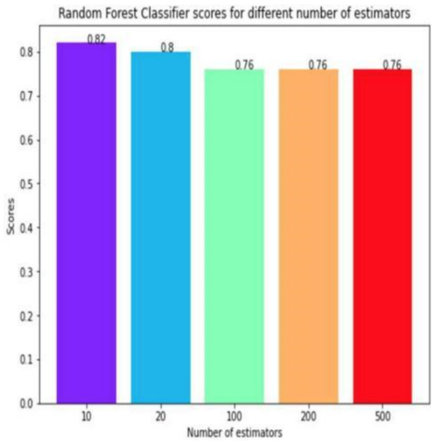

Figure 9: Random Forest Result Image

RESULT AND DISCUSSION

This study investigates hyperparameters' impact on the predictive performance of five distinct machine learning models—Logistic Regression, k-nearest Neighbors, Support Vector Machine, Decision Tree, and Random Forest—in the context of early-age heart disease prediction. Through trials, various hyperparameter techniques are employed in the prediction performance comparison of the five algorithms. Data is divided into sets of training and testing, in proportions of 70% and 30%, respectively, and a five-fold cross-validation method is employed. Earlier this has been demonstrated to yield a generalized model while mitigating the risk of overfitting, crucial for effective early-age heart disease prediction.

CONCLUSION

Predicting early-age heart disease presents a formidable challenge in the medical field, demanding significant time and effort from healthcare professionals. This study employs a variety of algorithms, including LR, KNN, SVM, DT, and Random Forest, to address this challenge. Heart Disease UCI Kaggle dataset is utilized to evaluate model performance. The Random Forest and Logistic Regression emerge as the optimal hyperparameter for testing accuracy, producing consistently high accuracy results of 82%. For early-age heart disease prediction, simplifying predictive models by prioritizing common attributes like smoking status, alcohol consumption, and exercise frequency can enhance accessibility and reduce reliance on costly medical tests. This approach makes assessments more user-friendly and inclusive, aiding in early detection and prevention efforts alternative to having attributes not available to any normal person until he/she takes some medical tests which will cost them more money, time and for medical professionals, equipment.

REFERENCES

1. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using Machine Learning classification in e-healthcare," IEEE Access, Vol. 8, 2020, pp.107562–107582.
2. M. Mullen, A. Zhang, G. K. Lui, A. W. Romfh, J. W. Rhee, and J. C. Wu, "Race and Genetics in Congenital Heart Disease: Application of iPSCs, Omics, and Machine LearningTechnologies," Frontiers in Cardiovas-cular Medicine, Vol. 8, 2021, pp. 37–51.
3. Ramesh TR, Umesh Kumar Lilhore, Poongodi M, Sarita Simaiya ,Amandeep Kaur ,Mounir Hamdi, "Predictive Analysis of Heart Diseases with Machine Learning Approaches" Malaysian Journal of Computer Science, 2022, pp. 132-148, 2022.
4. https://www.who.int/data/gho/data/themes/topics/health-workforce
5. Ahmed, H.; Younis, E.M.G.; Hendawi, A.; Ali, A.A. Heart disease identification from patients' social posts, machine learning solution on Spark. Future Gener. Comput. Syst. 2020, 111, 714–722.
6. Hao, Y.; Usama, M.; Yang, J.; Hossain, M.S.; Ghoneim, A. Recurrent convolutional neural network based multimodal disease risk prediction. Futur. Gener. Comput. Syst. 2018, 92, 76–83.
7. Ali, F.; El-Sappagh, S.; Islam, S.R.; Kwak, D.; Ali, A.; Imran, M.; Kwak, K.-S. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Inf. Fusion 2020, 63, 208–222.
8. Ibrahim M. El-Hasnony 1 , Omar M. Elzeki, Ali Alshehri and Hanaa Salem "Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction" Sensors, 2022.
9. Mohamed G. El-Shafiey, Ahmed Hagag, El-Sayed A. El-Dahsha, Manal A. Ismail "A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest" Springer, 2022.
10. Kartik Budholiya , Shailendra Kumar Shrivastava, Vivek Sharma "An optimized XGBoost based diagnostic system for effective prediction of heart disease", Journal of King Saud University – Computer and Information Sciences, ScienceDirect, 2022.
11. GHULAB NABI AHMAD, HIRA FATIMA, SHAFIULLAH, ABDELAZIZ SALAH SAIDI, (Member, IEEE), AND IMDADULLAH, (Senior Member, IEEE)," Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV", IEEE Access, 2022.
12. U. K. Lilhore, S. Simaiya, D. Prasad, and D. K. Verma, "Hybrid Weighted Random Forests Method for Prediction & Classification of Online Buying Customers," Journal of Information Technology Management, vol. 13, no. 2, pp. 245–259, 2021.
13. "The Heart Disease Dataset from the UCI Machine Learning Repository contains four databases. Accessed on July 10th, 2021. Available at:

https://archive.ics.uci.edu/ml/datasets/heart+Disease.” (2021)

14. S. M. Nagarajan, V. Muthukumaran, R. Murugesan, R. B. Joseph, M. Meram, and A. Prathik, “Innovative feature selection and classification model for heart disease prediction,” Journal of Reliable Intelligent Environments, pp. 1–11, 2021.
15. Purnomo A, Barata MA, Soeleman MA, Alzami F (2020) ”Adding feature selection on Na¨ıve Bayes to increase accuracy on classification heart attack disease”. J Phys Conf Ser 1511:012001–012001.
16. Luo M, Wu K (2020) Heart rate prediction model based on neural network. IOP Conf Ser Mater Sci Eng 715:012060–012060
17. Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G (2020) Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evol Intel 13:185–196.
18. S. Sharma and M. Parmar, “Heart diseases prediction using deep learning neural network model,” International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 9, no. 3, pp. 124–137, 2020.
19. D. Singh and J. S. Samagh, ”A comprehensive review of heart disease prediction using Machine Learning,” Journal of Critical Reviews, vol. 7, no. 12, pp. 281–285, 2020.
20. U. K. Lilhore, S. Simaiya, K. Guleria, and D. Prasad, ”An Efficient Load Balancing Method by Using Machine Learning-Based VM Distribution and Dynamic Resource Mapping,” Journal of Computational and Theoretical Nanoscience, vol. 17, no. 6, pp. 2545–2551, 2020.
21. Bo Jin , Chao Che, Zhen Liu, Shulong Zhang, Xiaoming Yin, and Xiaopeng Wei, “Predicting the Risk of Heart Failure WithEHR Sequential Data Modeling”, IEEE Access Volume 6 2018.
22. https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
23. Swamy, S. R., & KS, N. P. (2022, December). Hybrid machine learning model for early discovery and prediction of polycystic ovary syndrome. In 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE) (pp. 1-8). IEEE.
24. Archana, R., Vaishnavi, C., Priyanka, D. S., Gunaki, S., Swamy, S. R., & Honnavalli, P. B. (2022, May). Remote Health Monitoring using IoT and Edge Computing. In 2022 International Conference on IoT and Blockchain Technology (ICIBT) (pp. 1-6). IEEE.
25. Swamy, S. R., Prasad, K. N., & Tripathi, P. (2020, October). Smart home lighting system. In 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC) (pp. 75-81). IEEE.
26. Hukkeri, S., Malage, R. V., Swamy, S. R., & Honnavali, P. B. (2021). Estimation of Engagement of Learners in MOOCs using Smart Visual Processing.
27. Singh, N., Shyam, A., Swamy, S. R., & Honnavalli, P. B. (2021). Differential Privacy in NoSQL Systems. In Data Science and Security: Proceedings of IDSCS 2021 (pp. 374-384). Springer Singapore.

***Corresponding Author:** Purushotam Kumar Singh