International Journal of Computer Applications (0975 – 8887)
Volume 186 – No.63, January 2025

A Systematic Review of Machine Learning Models for Cardiac Disease Prediction

Sunanda Budihal
Assistant Professor
Dept. of Computer Science
Government First Grade College, Bagalkot-587103, Karnataka, India

Sheetalrani R. Kawale
Assistant Professor
Dept. of Computer Science
Karnatak State Akkamahadevi Women's University, Vijayapura 586108, Karnataka, India

ABSTRACT
Heart disease is one of the leading causes of mortality worldwide, making its early detection and prediction crucial for saving lives. Machine learning (ML) algorithms have the potential to revolutionize the healthcare system by enhancing diagnostic accuracy and improving patient outcomes. This study reviews previous research that applied Deep Learning (DL) and ML techniques to predict heart disease. From the study it has seen that most of the work have used supervised ML algorithms, which includes Support Vector Machines (SVM), Gradient Boosting Classifier (GB), Decision Trees (DT), Random Forest (RF), and Logistic Regression (LR), have been employed on the UCI Machine Learning Repository (Heart) dataset to predict cardiac conditions. The accuracy of these algorithms varies, with studies reporting success rates between 88% and 95%. This review explores the factors influencing these outcomes, contributing to a better understanding of ML-based heart disease prediction models.

Keywords
Heart disease prediction, machine learning, deep learning, supervised learning, UCI heart dataset.

1. INTRODUCTION
Globally, cardiovascular diseases (CVDs) are the leading cause of disease burden and mortality. In India, there is a steady rise in cardiovascular disorders and associated risk factors. The increasing prevalence of heart-related conditions can be attributed to behavioral lifestyle factors, age, and genetic predisposition [1]. Usually, the term "heart disease" encompasses a range of disorders affecting the heart's ability to function efficiently, including heart failure, arrhythmia, coronary artery disease (CAD), and heart valve disease. Among these, CAD is the most prevalent. This condition arises due to plaque buildup within the coronary arteries, leading to a narrowing or blockage that restricts blood flow to the heart muscle. Such blockages can result in severe complications like heart attacks, heart failure, and arrhythmias. In some cases, medical interventions such as angioplasty or bypass surgery are required to restore proper blood flow.

Arrhythmia, the second most common heart condition, occurs due to irregular electrical activity in the heart, which disrupts its normal rhythm. This can cause the heart to beat too fast, too slow, or irregularly. Millions of people worldwide suffer from arrhythmias, which manifest through symptoms such as chest pain, weakness, dizziness, shortness of breath, palpitations, and fatigue [2], [3]. While some arrhythmias are harmless, others can be life-threatening. Lifestyle choices, physical inactivity, and excessive consumption of processed foods have contributed significantly to the rising incidence of heart disease. Early detection is crucial, as advanced stages can lead to life-threatening conditions like heart attacks. Sophisticated diagnostic techniques can aid in early detection, yet one of the major challenges remains patient reluctance to participate in research studies. Moreover, these studies are often expensive and time-consuming, limiting their widespread implementation. Unlike traditional clinical methods, data-driven techniques can analyze patterns of heart disease using patient health records, offering a cost-effective alternative [4].

In recent years, artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), has revolutionized auxiliary diagnostics, leading to significant advancements in automated disease detection [5-8]. One of the key advantages of ML is its ability to diagnose heart disease with high accuracy while reducing the need for multiple invasive clinical trials [9]. By analyzing key patient characteristics and medical data, ML algorithms can provide precise predictions, aiding in early diagnosis. However, despite these advancements, physician validation remains an essential aspect of the diagnostic and treatment process. ML applications hold great potential in enhancing healthcare efficiency and improving medical professional's productivity. As the era of big data continues to expand, the integration of advanced ML algorithms is expected to play a crucial role in automated cardiac disease prediction, offering promising opportunities for the future of healthcare [10-14]. Hence, a detailed explanation of different ML models currently used for heart disease prediction are discussed in detail in next section.

2. MACHINE LEARNING MODELS
This study conducted review on several supervised ML techniques which are currently being used for early detection of heart disease. In ML, the labelled training dataset is used to train the algorithms to train the core supervised ML algorithm. After that, a non-labelled testing dataset is incorporated into this qualified model to group the data into comparable categories [15]. This section provides a summary of various ML supervised learning approaches used for heart-disease classification.

2.1 MACHINE LEARNING CLASSIFICATION TECHNIQUES
Classification is a kind of supervised ML technique that uses a prior dataset to predict future cases. This section provides a concise overview of the most popular categorization methods for the prediction of heart disease.

2.1.1 Decision Tree
A Decision Tree (DT) is a supervised learning method that is most used for classification, but it may also be utilized for regression tasks. Similar to a tree, a DT consists of root, branch, and leaf nodes, where each node denotes a quality or attribute,

each branch a decision or rule, and each leaf a conclusion. DT methods are used to separate the features. Like a flowchart, a DT is a very particular form. DTs are selected because they provide quick and dependable results in comparison to other ML techniques; minimal data preparation is required; and they are simple to understand. [16], [17]. Both classification and regression can be performed using DT using Classification and Regression Trees (CART algorithm).

2.1.2 Logistic Regression

A classification technique called Logistic Regression (LR) is used to estimate discrete values, usually binary ones like 0 and 1, yes or no. Because it forecasts the likelihood that an instance will belong to a class, it is crucial for binary classification issues such as spam identification and illness diagnosis. The output of logistic functions ranges from 0 to 1, making them perfect for classification tasks. Its efficiency, interpretability, and simplicity make it useful in a wide range of fields. When the event probability and features are linear, logistic regression performs well. In tasks involving binary classification, logistic regression is employed. Binary categorization is accomplished by logistic regression. It forecasts the possibility of class membership despite its name. In this linear approach, probability is modelled by a logistic function [36]. To implement LR as a binary classification, a threshold value is determined, which defines the separation of the data into two groups [18], [19].

2.1.3 Naïve Bayes

The Naïve Bayes (NB) classifier applies the Gaussian distribution rule to the problem at hand to carry out the classification analysis [20], [21]. The number of pre and posterior conditional probability classes. It is also very easy to use, quickly forecasted, and very expandable.

2.1.4 Random Forest

To get the optimal outcomes, the Random Forest (RF) combines multiple DTs. Tree learning is the main use of a bootstrap aggregate. Based on the complete set of data, many random samples are generated. Each model is created from the Bootstrap samples using row sampling. Each model is trained independently throughout the process. The decision is made by a majority vote using the totality of the models.

2.1.5 Support Vector Machine

Depending on the type of problem, a Support Vector Machine (SVM) applies mathematical models to solve regression and classification issues. The SVM algorithm uses statistically based techniques to identify the most efficient hyperplane. It converts your data using the kernel approach, and then uses these modifications to determine the best boundary's is a powerful algorithm for regression and classification. It looks for the hyperplane that maximizes the margin while classifying data the best. SVM's kernel approach allows it to manage nonlinear feature interactions and perform well in high dimensional domains. Powerful and well-known for its accuracy in high-dimensional domains, this classification method. SVM generalizes effectively to a variety of datasets and is resistant to overfitting. Among other domains, it finds use in image recognition, text categorization, and bioinformatics. Its applications include text classification, picture recognition, and bioinformatics—fields where accuracy is critical.

2.1.6 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is an ancient and user-friendly statistical learning algorithm. The number of nearest neighbors to be used is specified by the K value [22], [23]. KNN facilitates simple and adaptable regression and classification by using the majority class of KNN. There are no data distribution presumptions with non-parametric KNN. It performs well for a variety of applications and functions best with unequal decision boundaries. As an instance-based, or lazy, learning method, KNN only approximates the function locally; all computation is postponed until the function is evaluated. It uses a similarity metric (like distance functions) to categorize new examples. Because of its ease of use and efficiency when processing non-linear data, KNN is frequently utilized in recommendation systems, anomaly detection, and pattern recognition applications.

2.2 DATA PREPROCESSING

The quality of the dataset and the preprocessing methods have an impact on the predictive model's accuracy and performance in addition to the algorithms employed. The actions taken on the dataset prior to implementing ML algorithms are referred to as preprocessing. Because it prepares the dataset and puts it in a format that the algorithm can understand, the preprocessing step is crucial. Errors, missing information, redundancies, noise, and many other issues might make a dataset unfit for direct use by a machine learning system. The dataset's size is an additional factor. There exist datasets with numerous features that complicate the process of algorithm analysis, pattern recognition, and forecast accuracy. These issues can be resolved by examining the dataset and applying the appropriate data preprocessing methods. Depending on the kind of dataset, data preprocessing procedures can involve feature selection, data normalization, data cleaning, data transformation, imputation of missing values, and other processes [24].

2.3 PERFORMANCE EVALUATION

Researchers assess prediction models using the measures listed below and present the outcomes of their performance. Without getting into the intricate nuances and mathematical formulas, we give a brief explanation of each technique.

1. Accuracy: The percentage of accurate results is displayed by this indicator.
2. Precision: The relevance of the outcome is shown by this parameter.
3. Sensitivity or Recall: Evaluates the pertinent findings that are returned.
4. F-Measure: Integrates recall and precision.
5. Receiver Operation Characteristic (ROC): This is a performance graph for the classifier. Both the appropriately and improperly classified cases are displayed [27].

All the exiting approaches for classification of CVD employed accuracy as the most common performance evaluation parameter. Because of this, the main objective of this overview article is to classify, compare, and evaluate earlier work according to the correctness.

3. HEART DISEASE DATASET

In the realm of heart disease prediction using ML, studies have employed a range of models and datasets, showcasing the effectiveness of various algorithms in predicting heart conditions. Most studies focus on datasets like the Cleveland Heart Disease dataset and UCI (University of California, Irvine, CA) heart disease dataset and IEEE data set, utilizing classifiers such as DTs, RF, SVM, and ensemble methods like XGBoost (XGB). These studies often report high accuracy rates, typically ranging from 88% to 95%, with some models demonstrating the ability to handle complex healthcare data effectively.

The heart disease dataset from the UCI [25] Centre for ML and intelligent systems is the dataset that is utilized in most research papers. Every database has fourteen features, but varies in the quantity of records. Due to its greater record count and lower percentage of missing characteristics compared to other datasets, the Cleveland dataset is the most popular among machine learning researchers. The num field indicates whether the patient has cardiac disease. The 14 attributes/features as they appear in the dataset are displayed in Table 1 along with the description of each attribute.

Table 1. Dataset Attributes

Number	Attribute	Description
1	age	Age in years
2	gender	Male or Female
3	cp	Chest Pain type
4	trestbps	Resting Blood Pressure in mmHg
5	chol	Serum cholesterol in mg/dl
6	fbs	Fasting Blood Sugar
7	restecg	Resting Electrocardiographic results
8	thalach	Maximum heart rate achieved
9	exang	Exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	The slope of the peak exercise ST segment
12	ca	Number of major vessels (0-3) colored by fluoroscopy
13	thal	Thallium heart scan
14	num	Diagnosis of heart disease

Studies on the prediction of heart disease have been conducted up to this point. On datasets of heart patients, numerous data mining and ML algorithms have been used and proposed; varying outcomes have been obtained for various techniques. However, still heart disease is a major issue that we deal with nowadays. Here are a few recent research publications in the field. M. Kavitha et.al [26], discussed the problem of heart disease, which is a significant cause of mortality worldwide. Early prediction of heart disease can save many lives, but it is a critical challenge for regular clinical data analysis. The paper proposed a novel ML approach to predict heart disease using data mining techniques such as regression and classification. The authors used the Cleveland heart disease dataset and applied ML techniques such as RF and DT. They also designed a hybrid model of RF and DT to predict heart disease. According to the experimental findings, the hybrid model's heart disease prediction accuracy rate was 88.7%. The researchers also discussed the future work of applying DL algorithms for heart disease prediction and classifying it as a multi-class problem to identify the disease's level

A. A. Ahdal et al. [27], used ML techniques to predict heart disease by processing and analyzing medical data. The study found that several ML approaches can be used to predict heart disease accurately, and some encouraging results were achieved and validated. The paper also identified various risk factors for heart disease and discussed many of the heart disease factors and symptoms that a person could have. Overall, the paper provided insights into the use of ML techniques for predicting heart disease and highlighted the importance of accurate prediction in the medical field.

G. N. Ahmad et.al [28], stated that heart disease is a serious health condition and a major concern for human beings. The paper compared five ML algorithms, namely LR, DT, NB, KNN, and SVM to predict the existence of heart disease. The algorithms were evaluated based on their accuracy scores, sensitivity, specificity, and F1 score. After analyzing the results, the SVM algorithm was found to be the best model for forecasting the existence of heart disease

K. K. Baseer et al. [29], compared ML models for diagnosing heart diseases with and without hyperparameter tuning. This paper discussed the use of ML models such as LR, RF, SVM, and KNN for the prediction of heart diseases. The study focused on the importance of hyperparameter tuning, which involved finding the best combination of hyperparameters for each model to achieve accurate results. The paper proposed use of GridSearchCV and RandomizedSearchCV methods for hyperparameter tuning. The work concluded that the accuracy of the models depends on the type of model selected for building and the effectiveness of hyperparameter tuning

M. S. Manoj et.al [30], proposed a novel approach for accurate heart attack detection using Exploratory Data Analysis (EDA) based heart attack detection using ML. The presented approach considered LR algorithm, Multiple Layer Perceptron (MLP), CatBoost-Regression (CBR) algorithm, and RF regression algorithm for comparative analysis to create an effective model using Hybrid SMLT technique. The work concluded that the proposed approach achieved a high accuracy during the deployment of work into the cloud and can be further optimized to work in IoT network to test various real-time applications.

A. Khan et.al [31], focused on utilizing ML techniques to enhance doctor's perception and improve patient diagnosis and treatment in the healthcare sector. The study highlighted that the RF algorithm is the most suitable for CVD classification and prediction, with the potential for global implementation in healthcare settings. CVD prediction is crucial for saving live worldwide. ML algorithms like DT, RF, LR, NB, and SVM were applied in the study. Among these algorithms, the RF algorithm showed the highest accuracy, sensitivity, and Receiver Operating Characteristic (ROC) curve for CVD prediction

N. Narayanan et.al [32], demonstrated effectiveness of ML algorithms in early prediction of cardiac issues for lifestyle changes and appropriate medical interventions CVD is a leading cause of death worldwide, with ML being used to predict and detect heart disease efficiently. The research focuses on using supervised ML classifiers to identify important features for predicting heart disease. The paper introduced an oversampling technique called Synthetic Minority Over-sampling Technique (SMOTE) to handle unbalanced datasets in predicting cardiac diseases. By applying ML techniques with SMOTE oversampling, the RF method achieved high accuracy

R. Hoque et.al [33], focused on employing SVM algorithm to predict heart disease using clinical data. The study utilized a dataset comprising 14 features and 300 patient records to train and validate the SVM model. Various preprocessing techniques, including data normalization and splitting into training and test sets, were applied to enhance model performance. The SVM was tested with both linear and non-linear kernels to determine the optimal approach for classification. The results demonstrated that the non-linear SVM outperformed the linear model, achieving high accuracy in predicting heart disease. This underscored the potential of SVM as a reliable tool for early heart disease detection, which can facilitate timely and effective clinical interventions.

Y. Rimal et.al [34] reviewed around 18 different ML models amongst 8 models have trained by researchers and remaining are auto trained. These models were tested on an open-source heart disease dataset, and their accuracy, mean squared error (MSE), and R2 scores were analyzed. Traditional models like SVM, LR, and neural networks achieved around 80% accuracy, while models like Gaussian classifiers, KNN, and MLP scored about 76%. AutoML models outperformed these, with the generalized linear model achieving 88% accuracy, and the gradient boosting and distributed RF models both scoring 87%. The Extra-Trees (ET) model also performed well, with an 82% accuracy.

D. M. K. Selvi et al. [35] focused on leveraging AI, ML and DL techniques for early prediction and treatment of CVD. Among the methodologies explored, ensemble models soft voting ensemble classifier, which combined multiple algorithms to enhance prediction performance. The researchers have reported that their algorithm have achieved 93.44% of accuracy with the Cleveland dataset and achieved 95% accuracy with the IEEE Dataport dataset. The ensemble approach outperformed individual models like LR (90.16%) and AdaBoost (90%) by leveraging the strengths of multiple algorithms. The authors optimized these models using GridSearchCV and validated them with five-fold cross-validation to ensure reliability and robustness in early heart disease prediction.

4. COMPARISON OF DIFFERENT ML TECHNIQUES FOR HEART DISEASE PREDICTION

This segment provides a tabular comparison between all the research papers described above. The comparison is made based on accuracy and can be seen in table 2. The table has six elements which of few are as follows

Table 2. Comparative Analysis of Machine Learning Approaches for Cardiac Disease Prediction

Author	Title	Dataset Used	Model/Classifier	Validation Parameters	Results	Limitations	Future Work
M. Kavitha et.al, [26] (2021) IEEE	Heart Disease Prediction Using Hybrid Machine Learning Model	Cleveland Heart Disease dataset	DT, RF, and a hybrid DT+RF model	Accuracy	Hybrid model, i.e., DT+RF: 88.7%	Focused only on a single dataset, limiting generalizability.	Explore DL techniques for better performance.
A. Garg et al. [27] (2022) IOP	An Integrated Machine Learning Techniques for Accurate Heart Disease Prediction	Cleveland Heart Disease dataset	DT, RF, SVM	Accuracy, F1-score	Ensembled model: 90%	Lack of discussion on computational complexity and real-world application.	Develop a computationally efficient ensemble framework.
Ahmad, G. N., et al. [28] (2022) IEEE	Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning with GridSearchCV	Cleveland and other datasets	SVM, RF, with/without GridSearchCV	Accuracy, Precision, Recall	SVM with GridSearch CV: 93%	High computational cost due to hyperparameter tuning.	Explore lightweight optimization techniques.
K. K. Baseer et al. [29] (2023) IEEE	Medical Diagnosis of Human Heart Diseases with/without Hyperparameter Tuning	Cleveland Dataset	KNN, SVM, DT	Accuracy, AUC-ROC	Tuned models achieved 92%	Computationally intensive hyperparameter tuning.	Include real-time datasets for validating model scalability.
M. S. Manoj et al. [30]. (2023) IEEE	Design and Analysis of Heart Attack Prediction System Using ML	Likely Cleveland dataset	NB, DT, Ensemble Methods	Accuracy, F1-score	Ensemble with RF achieved 91%	Requires validation on larger datasets for real-world use.	Explore real-time data streams and edge computing for prediction.
A. Khan et al. [31] (2023) IEEE	A Novel Study on ML Algorithm-Based Cardiovascular Disease Prediction	UCI datasets	Gradient Boosting, XGBoost	Accuracy	XGBoost achieved 94%	Limited focus on dynamic datasets and real-time applications.	Implement real-time monitoring systems integrated with healthcare facilities.
N. Narayanan et al [32] (2024) Elsevier	Efficient ML Techniques for Cardiac Disease Prediction Using SMOTE	Cleveland Dataset	LR, SVM, Neural Networks	Accuracy, Precision, Recall	Neural Networks achieved 95%	Computational overhead due to SMOTE application.	Investigate adaptive sampling techniques to handle data imbalance.
R. Hoque et al [33] (2024)	Heart Disease Prediction Using SVM	UCI Heart Disease dataset	SVM	Accuracy, F1-score	SVM achieved 92%	Lacks interpretability of SVM model.	Combine SVM with explainable AI (XAI) for

IJSR							better insights.
Y. Rimal et al [34] (2024) Springer	Comparative Study of Ensemble Learning and AutoML for Heart Disease Prediction	Cleveland Dataset	Ensemble methods (e.g., XGBoost) vs. AutoML	Accuracy, Execution Time	AutoML achieved 94%	AutoML dependency affects reproducibility.	Explore AutoML frameworks on more complex and dynamic datasets.
D. M. K. Selvi et al. [35] (2024) IEEE	Revolutionizing Cardiovascular Care: The Role of AI, ML, and DL in Early Heart Disease Prediction and Treatment	Cleveland Dataset (UCI Repository) IEEE Dataport Dataset	LR, RF, KNN, NB, Gradient Boosting, AdaBoost. Soft Voting Ensemble Classifier	GridSearchCV for Hyperparameter Optimization Five-fold Cross-Validation	Soft Voting Ensemble Classifier: 93.44% (Cleveland dataset) 95% (IEEE Dataport dataset)	Data imbalance in datasets. Privacy concerns over patient data. Scalability issues in clinical use.	Improved data augmentation. IoT-enabled real-time monitoring. Explainable AI for clinician trust.

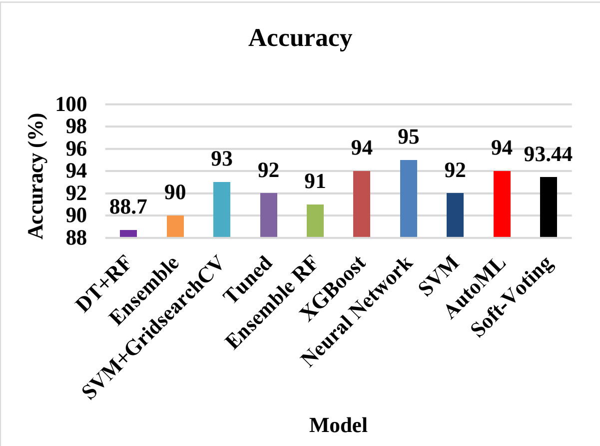

Figure 1. Accuracy achieved by existing approaches.

In Figure 1, the accuracy achieved by existing approaches is shown graphically. A common trend among these studies is the use of feature selection, which helps improve the accuracy of predictions by focusing on the most influential factors. For example, some studies combine RF with feature selection techniques, achieving notable improvements in accuracy but also recognizing challenges such as the risk of overfitting and a lack of model interpretability. Moreover, hyperparameter tuning, such as using GridSearchCV, has proven beneficial in optimizing model performance, although it often leads to increased computational costs. While high accuracy is achieved, many studies also highlight limitations in their models. These include the generalizability of models trained on a single dataset, computational complexity, and the lack of real-time applicability. As such, real-time data integration and edge computing are frequently suggested as avenues for future research. Furthermore, adaptive sampling techniques like SMOTE are proposed to address issues like class imbalance, although these methods add extra computational overhead. Moving forward, the need for XAI is emphasized to make models more interpretable, especially with complex models like SVM, which can lack transparency. In addition, there is a call for incorporating deep learning methods, as they have the potential to enhance performance but may require more computational resources. Moreover, AutoML is gaining attention for its potential to optimize models with less human intervention, though its reliance on specific frameworks could limit the reproducibility of results. In this review paper we came to know that ML models have shown promising results in predicting heart disease, challenges such as computational overhead, data generalization, and model interpretability remain. Study can be extending to focus on integrating real-time data, automating hyperparameter tuning, exploring DL techniques, and improving the interpretability and scalability of models for broader use in healthcare settings.

5. CONCLUSION

ML has emerged as a transformative tool in the diagnosis and prediction of cardiac diseases, offering an efficient, cost-effective alternative to traditional diagnostic methods. This review has examined various ML models, including DT, LR, SVM, KNN, and Neural Networks, highlighting their strengths, limitations, and comparative accuracy in predicting cardiac conditions. While many studies have reported high accuracy rates ranging from 88% to 95%, the successful integration of ML into real-world clinical practice remains a challenge due to the need for validation from medical professionals. Ensuring the reliability, interpretability, and generalizability of these models across diverse populations is essential for their widespread adoption in healthcare settings. Looking ahead, future research should focus on leveraging larger and more diverse datasets, refining ML methodologies, and exploring advanced techniques such as DL and explainable AI to enhance model interpretability. Additionally, real-time data analysis, IoT-enabled health monitoring systems, and cloud-based AI platforms can further strengthen the predictive capabilities of ML-driven models. Addressing challenges such as data privacy, computational efficiency, and regulatory compliance will be crucial in bridging the gap between ML research and clinical applications. As technology advances, the collaboration between data scientists and healthcare professionals will be key in ensuring that ML-based cardiac disease prediction systems are not only highly accurate but also practical, ethical, and seamlessly integrated into modern healthcare workflows.

6. REFERENCES

[1] The Cardiovascular study in India. [Online]. Available: https://www.statista.com/study/86148/cardiovascular-disease-in-india/. Accessed: Jan. 28, 2024.

[2] A. F. A. Iswisi, O. Karan, and J. Rahebi, "Diagnosis of Multiple Sclerosis Disease in Brain Magnetic Resonance Imaging Based on the Harris Hawks Optimization Algorithm," *BioMed Research International*, vol. 2021, pp. 1–12, Dec. 2021, doi: 10.1155/2021/3248834.

[3] H. Al-Safi, J. Munilla and J. Rahebi, "Harris Hawks Optimization (HHO) Algorithm based on Artificial Neural Network for Heart Disease Diagnosis," *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkur, Karnataka, India, 2021, pp. 1-5, doi: 10.1109/ICMNWC52512.2021.9688348.

[4] T. Nguyen and Z. (Amy) Wang, “Cardiovascular Screening and Early Detection of Heart Disease in Adults With Chronic Kidney Disease,” *The Journal for Nurse Practitioners*, vol. 15, no. 1, pp. 34–40, Jan. 2019, doi: 10.1016/j.nurpra.2018.08.004.

[5] R. Liu, C. Ren, M. Fu, Z. Chu, and J. Guo, “Platelet Detection Based on Improved YOLO_v3,” *Cyborg and Bionic Systems*, vol. 2022, pp. 1–9, Sep. 2022, doi: 10.34133/2022/9780569.

[6] A. A. Mohamed, Aybaba Hançerlioğullari, Javad Rahebi, M. K. Ray, and S. Roy, “Colon Disease Diagnosis with Convolutional Neural Network and Grasshopper Optimization Algorithm,” *Diagnostics*, vol. 13, no. 10, pp. 1728–1728, May 2023, doi: 10.3390/diagnostics13101728.

[7] J. Rahebi, “Fishier mantis optimiser: a swarm intelligence algorithm for clustering images of COVID-19 pandemic,” *International Journal of Nanotechnology*, vol. 20, no. 1/2/3/4, pp. 25–49, 2023, doi: 10.1504/ijnt.2023.131111.

[8] A. Shalchi and Javad Rahebi, “Human retinal optic disc detection with grasshopper optimization algorithm,” *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 24937–24955, Mar. 2022, doi: 10.1007/s11042-022-12838-8.

[9] R. Sameni and G. D. Clifford, “A Review of Fetal ECG Signal Processing Issues and Promising Directions,” *The Open Pacing, Electrophysiology & Therapy Journal*, vol. 3, pp. 4-20, 2010, doi: 10.2174/1876536x01003010004.

[10] H. Al-Safi, J. Munilla, and J. Rahebi, “Patient privacy in smart cities by blockchain technology and feature selection with Harris Hawks Optimization (HHO) algorithm and machine learning,” *Multimedia Tools and Applications*, Feb. 2022, doi: 10.1007/s11042-022-12164-z.

[11] J. Zou, Q. Li, S. Yang, J. Zheng, Z. Peng, and T. Pei, “A dynamic multiobjective evolutionary algorithm based on a dynamic evolutionary environment model,” *Swarm and Evolutionary Computation*, vol. 44, pp. 247–259, Feb. 2019, doi: 10.1016/j.swevo.2018.03.010.

[12] E. K. P. Chong and S. H. Zak, An Introduction to Optimization. Hoboken, NJ, USA: John Wiley & Sons, 2004.

[13] T. Rashid, Make Your Own Neural Network. Scotts Valley, CA, USA: CreateSpace Independent Publishing Platform, 2016.

[14] J. Anderson, L. Rainie, and A. Luchsinger, “Artificial intelligence and the future of humans,” *Pew Research Center*, vol. 10, pp. 12, 2018.

[15] “Identification of Cardiovascular Diseases Using Machine Learning | Semantic Scholar.” [Online]. Available: https://www.semanticscholar.org.

[16] K. Purushottam, K. Saxena, and R. Sharma, “Efficient Heart Disease Prediction System,” *Procedia Computer Science*, vol. 85, pp. 962–969, 2016, doi: 10.1016/j.procs.2016.05.288.

[17] M. Shouman, T. Turner, and R. Stocker, “Using Decision Tree for Diagnosing Heart Disease Patients,” *Proceedings of the 9-th Australasian Data Mining Conference (AusDM'11), Ballarat, Australia*, vol. 121, pp. 23-29, 2011.

[18] E. Maini, B. Venkateswarlu, B. Maini, and D. Marwaha, “Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India,” *Medical Journal Armed Forces India*, Jan. 2021, doi: 10.1016/j.mjafi.2020.10.013.

[19] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, “Developing prediction models for clinical use using logistic regression: an overview,” *Journal of Thoracic Disease*, vol. 11, no. S4, pp. S574–S584, Mar. 2019, doi: 10.21037/jtd.2019.01.25.

[20] V. Radpour and F. S. Gharehchopogh, “A Novel Hybrid Binary Farmland Fertility Algorithm with Naïve Bayes for Diagnosis of Heart Disease,” *Sakarya University Journal of Computer and Information Sciences*, vol. 5, no. 1, pp. 90–103, Apr. 2022, doi: 10.35377/saucis...978409.

[21] R. TR, U. K. Lilhore, P. M, S. Simaiya, A. Kaur, and M. Hamdi, “Predictive Analysis of Heart Diseases with Machine Learning Approaches,” *Malaysian Journal of Computer Science*, pp. 132–148, Mar. 2022, doi: 10.22452/mjcs.sp2022no1.10.

[22] 22. I. K. A. Enriko, M. Suryanegara, and D. Gunawan, “Heart disease prediction system using k-Nearest neighbor algorithm with simplified patient’s health parameters,” *Journal of Telecommunication, Electronic and Computer Engineering*, Dec. 2016.

[23] “Top 6 Machine Learning Classification Algorithms,” *GeeksforGeeks*, Feb. 26, 2024. https://www.geeksforgeeks.org/top-6-machine-learning-algorithms-for-classification/ Accessed: Jan. 28, 2024.

[24] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, “Big data preprocessing: methods and prospects,” *Big Data Analytics*, vol. 1, no. 1, Nov. 2016, doi: 10.1186/s41044-016-0014-0.

[25] A. Garg, B. Sharma, and R. Khan, “Heart disease prediction using machine learning techniques,” *IOP Conference Series: Materials Science and Engineering*, vol. 1022, p. 012046, Jan. 2021, doi: 10.1088/1757-899x/1022/1/012046.

[26] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2021, pp. 1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

[27] A. A. Ahdal, M. Rakhra, S. Badotra and T. Fadhaeel, "An integrated Machine Learning Techniques for Accurate Heart Disease Prediction," *2022 International Mobile and Embedded Technology Conference (MECON)*, Noida, India, 2022, pp. 594-598, doi: 10.1109/MECON53876.2022.9752342.

[28] G. N. Ahmad, H. Fatima, S. Ullah, A. Salah Saidi and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV," in *IEEE Access*, vol. 10, pp. 80151-80173, 2022, doi: 10.1109/ACCESS.2022.3165792.

[29] K. K. Baseer, S. B. A. Nas, S. Dharani, S. Sravani, P. Yashwanth and P. Jyothirmai, "Medical Diagnosis of

Human Heart Diseases with and without Hyperparameter tuning through Machine Learning," *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2023, pp. 1-8, doi: 10.1109/ICCMC56507.2023.10084156.

[30] M. S. Manoj, K. Madhuri, K. Anusha and K. U. Sree, "Design and Analysis of Heart Attack Prediction System Using ML," *2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS)*, Raichur, India, 2023, pp. 01-06, doi: 10.1109/ICICACS57338.2023.10099819.

[31] A. Khan, M. Qureshi, M. Daniyal, and K. Tawiah, “A Novel Study on Machine Learning Algorithm-Based Cardiovascular Disease Prediction,” *Health & Social Care in the Community*, vol. 2023, p. e1406060, Feb. 2023, doi: 10.1155/2023/1406060.

[32] N. Narayanan and None Jayashree, “Implementation of Efficient Machine Learning Techniques for Prediction of Cardiac Disease using SMOTE,” *Procedia Computer Science*, vol. 233, pp. 558–569, Jan. 2024, doi: 10.1016/j.procs.2024.03.245.

[33] R. Hoque, None Masum Billah, None Amit Debnath, S. Hossain, and N. Bin, “Heart Disease Prediction using SVM,” *International Journal of Science and Research Archive*, vol. 11, no. 2, pp. 412–420, Mar. 2024, doi: 10.30574/ijsra.2024.11.2.0435.

[34] Y. Rimal, Siddhartha Paudel, N. Sharma, and Abeer Alsadoon, “Machine learning model matters its accuracy: a comparative study of ensemble learning and AutoML using heart disease prediction,” *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35025–35042, Sep. 2023, doi: 10.1007/s11042-023-16380-z.

[35] D. M. K. Selvi, J. Aswini, C. Balakrishnan, K. Suganya, B. G. Sheena and S. S. R, "Revolutionizing Cardiovascular Care: The Role of AI, ML, and DL in Early Heart Disease Prediction and Treatment," *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)*, Bengaluru, India, 2024, pp. 1-6, doi: 10.1109/ICECCC61767.2024.10593941.

[36] “World Health Statistics 2021,” *Who.int*, 2021. https://www.who.int/publications/b/58471 (accessed Jan. 28, 2025).

[37] A. Kumar Dubey, K. Choudhary, and R. Sharma, “Predicting Heart Disease Based on Influential Features with Machine Learning,” *Intelligent Automation & Soft Computing*, vol. 30, no. 3, pp. 929–943, 2021, doi: 10.32604/iasc.2021.018382.