Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online ***Vol 6 Issue 1, January 2024***
(https://fcihib.journals.ekb.eg)

A Comparative Study Of Artificial Intelligence Techniques For Categorization And Prediction Of Heart Diseases

Abdullah Reda Rashwan[1], Laila El Fangary[2], Safaa Azzam[2]
[1]Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh, Egypt
[2]Information Systems, Faculty of Computers and Artificial Intelligence, Helwan University, Egypt
abdullahrashwan16@gmail.com, lailaelfangary@gmail.com, eng_safaa_azam@yahoo.com

***Abstract*—Heart failure (HF) is one of the most common diseases in recent years, and a large number of people die annually around the world from it. The heart is considered one of the most important organs in the human body, so it requires high accuracy when predicting the presence of heart disease or not, as an error in prediction may cause human death, so it requires a high-accuracy method in predicting HF. Artificial intelligence (AI) plays a large and important role in many fields today, especially in the medical field, as AI helps doctors obtain a quick and accurate diagnosis of the patient's condition, which contributes to saving time during the diagnosis. It is important to predict HF using AI to help with rapid and accurate diagnosis and thus reduce the number of deaths from this disease. AI techniques increase the accuracy of predicting whether or not HF is present compared to traditional methods. Also, in rural areas where there are fewer physicians, it is very important to provide such technologies to aid in diagnosis. Many studies point to new AI-based HF prediction techniques. These technologies relied on different algorithms and datasets of different sizes and types. Each of these technologies has advantages and limitations. Therefore, this paper presents an illustrative study of the most advanced AI methods for HF prediction. This study also included a comparison between the different methods based on the most famous standards.**

***Index Terms*—Artificial Intelligence ,Machine Learning ,Deep Learning ,Feature extraction ,Feature selection ,Heart failure prediction.**

I. Introduction

Heart diseases affect human life, as the heart is one of the most important organs in the human body, but heart disease can be treated when detected early. On the other hand, detection of heart disease in the last stage leads to death due to the inaccuracy of early detection methods. According to the statistics of the World Health Organisation, HF is the most common disease in the world, and HF is the number one cause of death [1]. The chance of curing a disease is higher when it is caught early, so providing an accurate and rapid detection method is a required task to help doctors determine whether a person is sick or not. There are many factors that lead to heart disease, including high blood pressure, smoking, ageing, and heredity in the family. It is also difficult for many people to take a break and go to see a doctor for a periodic check-up, which delays the discovery of the disease and makes it difficult to treat. This problem is considered one of the most important factors that led to an increase in the number of heart diseases. There are fewer doctors in developing and rural areas, and the accuracy of heart disease prediction is high only for experienced doctors, so an automated method for predicting whether or not there is a disease is very important [2].

A person may need a follow-up of up to 24 hours to check for the presence of a disease or not, but these devices used in follow-up are very expensive, so the world is turning to artificial intelligence technology to provide a way through which doctors can be helped in the early detection of heart disease faster and with higher accuracy. Artificial intelligence has the ability to predict many diseases with high accuracy, as it simulates the capabilities of doctors based on data collected from sick and non-sick people. Machine learning (ML) and deep learning (DL) are among the most important branches of AI that help in processing the decision-making process based on previously collected data. The decision can be made based on different algorithms based on the shape and type of the data . The current traditional methods that doctors use to detect heart diseases depend entirely on the doctor's experience in terms of his ability to discover a specific pattern that indicates the presence of a disease or not. In the case of a junior doctor, he can make a huge mistake in detecting the disease early, which leads to the death of the person. Therefore, relying on ML and DL has become common and important to help diagnose with modern methods with high efficiency and speed to save time for early diagnosis and treatment [3].

The AI branches of ML and DL have recently achieved great success in the field of health, especially in detecting and diagnosing heart diseases based on different types and forms of datasets. Some of which depend on the sounds of the heart, some depend on the electrification of the heart, and some depend on factors such as age, blood pressure, whether the person smokes or not, etc. In this paper, we will provide a detailed explanation of the various methods in terms of dataset type, efficiency, disadvantages, advantages, etc. The different methods will be compared based on the most common standards among them. That is, this paper will be considered an important reference for many researchers interested in diagnosing heart diseases based on AI in order to reach the best method that can be relied upon in the future for accurate and rapid diagnosis of HF [4].

The rest of the survey is organised as follows: Details of the ML and DL algorithms used in the diagnosis of HF are in Section II, and a description of the datasets on which these algorithms are based is given in Section III. Analysis: The main findings from this survey are presented in Section IV. Section V represents the conclusion and future work.

II. An overview of artificial intelligence approaches for predicting heart diseases

A. Machine Learning Algorithms

*1) **Logistic Regression (LR)**:* A mathematical technique called LR uses past information from a dataset to predict a binary outcome (1 or 0) by examining the association between one or more characteristic variables already present. Depending on the logistic function, where numbers between 1 and 0 are given, we can classify a number greater than 0.5 as class 1 and a number less than 0.5 as class 0 as shown in figure 1. So, when it comes to the prediction of HF (1 or 0), LR can be very useful [5].

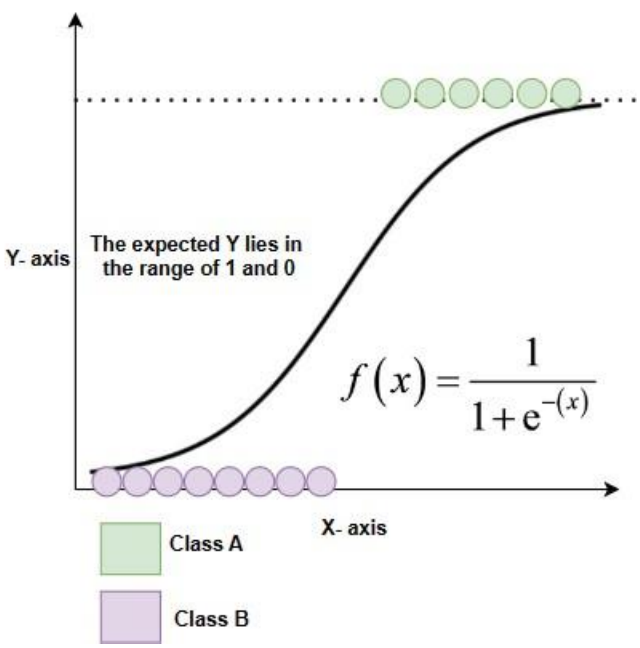

Fig. 1. Logistic Regression

*2) **Support Vector Machine (SVM)**:* SVM is employed to solve classification issues by discovers a dividing line and maximises the distance needed to distinguish between categories as shown in figure 2. When it comes to the prognosis of cardiac disease, SVM can be quite helpful [3].

*3) **Decision Tree (DT)**:* Used to solve issues related to classification. It is called this because it resembles the shape of a tree, as it consists of nodes and internal branches (shown in figure 3). In order to partition the dataset into multiple matching groups, it first measures the entropy score for each attribute [5].

*4) **K-Nearest Neighbor (KNN)**:* KNN is utilised for regression and classification. It locates the ’k’ closest points of data in the dataset to the point of data where the desired value needs to be located. The mean of all ”k” points of data is then applied to this specific point of data as shown in figure 4 [2].

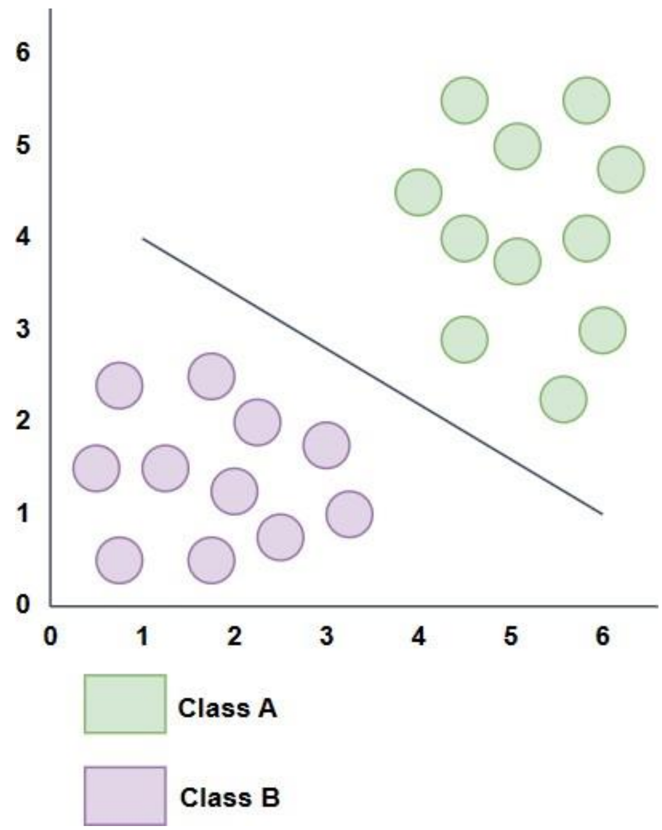

Fig. 2. Support Vector Machine

*5) **Random Forest (RF)**:* The RF algorithm is utilised for regression and classification. Multiple decision trees are formed via RF. In classification, RF employs a vote method to select the class, whereas in regression, RF takes the mean of the output from each individual tree as shown in figure 5 [4].

*6) **Na¨ıve Bayes (NB)**:* NB is suitable for classification issues. NB makes the assumption that each classifier is independent of the others, and that the presence of a given characteristic within a class is unrelated to that of other characteristics. Even though there is a relationship, the features will still individually influence the likelihood (shown in figure 6) [3].

Ekta et al. [2] proposed a new method for the early prediction of HF. More than one model has been trained based on different ML algorithms, including RF, NB, and KNN. They relied on a dataset containing 507 records. RF gave the best accuracy, as it classified 407 true records from the total size of the dataset. Although it gave the best result, the size of the dataset is small, so this model must be tested on a larger dataset to ensure the validity of the results.

Likewise, Ying et al. [4] proposed a new system for diagnosing HF. Choice operator techniques were used to identify attributes useful in diagnosis. They relied on RF to classify these attributes and then tell if a person was sick or not. This system achieved an accuracy of 80.1%. Despite this, however, the efficiency is considered small because the aim of the trend

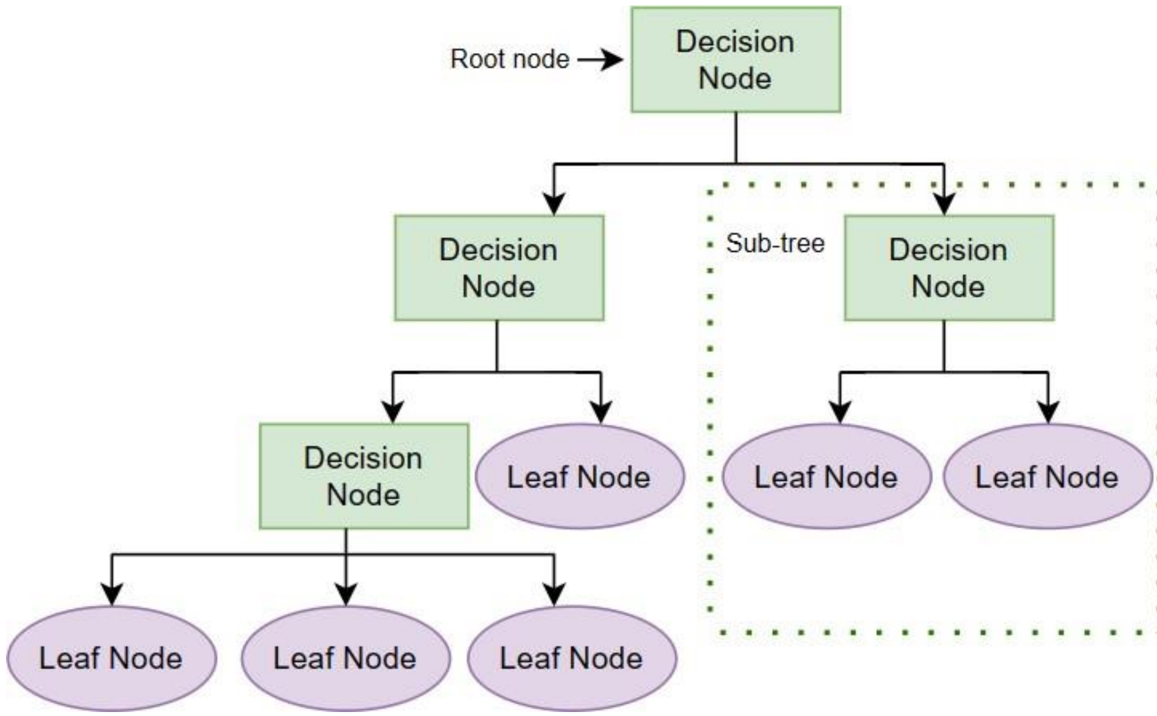

Fig. 3. Decision Tree

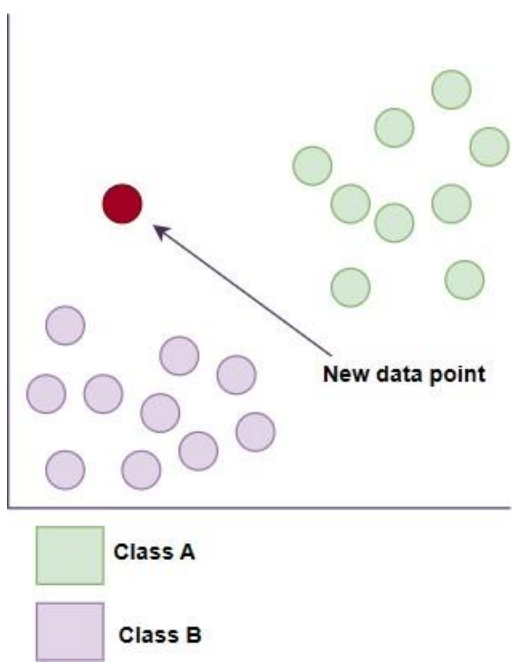

Fig. 4. K-Nearest Neighbor

towards diagnosis using AI instead of traditional means is rapid and accurate detection, so we still need a system that gives more efficiency than that.

whereas Rachana R Sanni and H.S.Guruprasad relied on the clinical HF dataset containing 377,650 visualisations to predict HF [6]. While training the model, they relied on choosing important attributes and then classifying the person as having HF or not depending on those attributes. More than one model was tested based on different ML algorithms, namely DT, RF, KNN, and LR. When compared, DT had the highest accuracy of 85.33%. Despite this, we still need to increase efficiency to be able to rely on this method for early detection of HF [5].

Also in this study, Rani et al. [3] developed a new system based on various ML algorithms, including RF, NB, and SVM. This system is trained and tested on Cleveland (CL) HF dataset, which contains 14 attributes, including age, blood pressure, etc. When comparing the algorithms used, RF had the best accuracy of 86.60%. One of the shortcomings of this system is that it needs to use an optimizer at the attribute selection stage to increase efficiency.

The same ClHF dataset was used by Ahmed et al. [7] for the diagnosis of human HF. They did not rely on all of the full features in this dataset, but more than one technique was used to determine which attributes are most reliable for predicting HF. A set of ML algorithms including DT, SVM, and RF. Comparing these different algorithms, RF came out with the highest efficiency of 94.9%.

Table I shows more details about different research studies that used ML algorithms to predict HF. The general workflow of the HF prediction system is shown in figure 7.

B. *Deep Learning Algorithms*

1) ***Convolutional Neural Network (CNN)****:* CNN has been very successful lately, especially in image processing. Where the most important attributes are extracted from an image, and then based on these attributes, this image can be classified. In

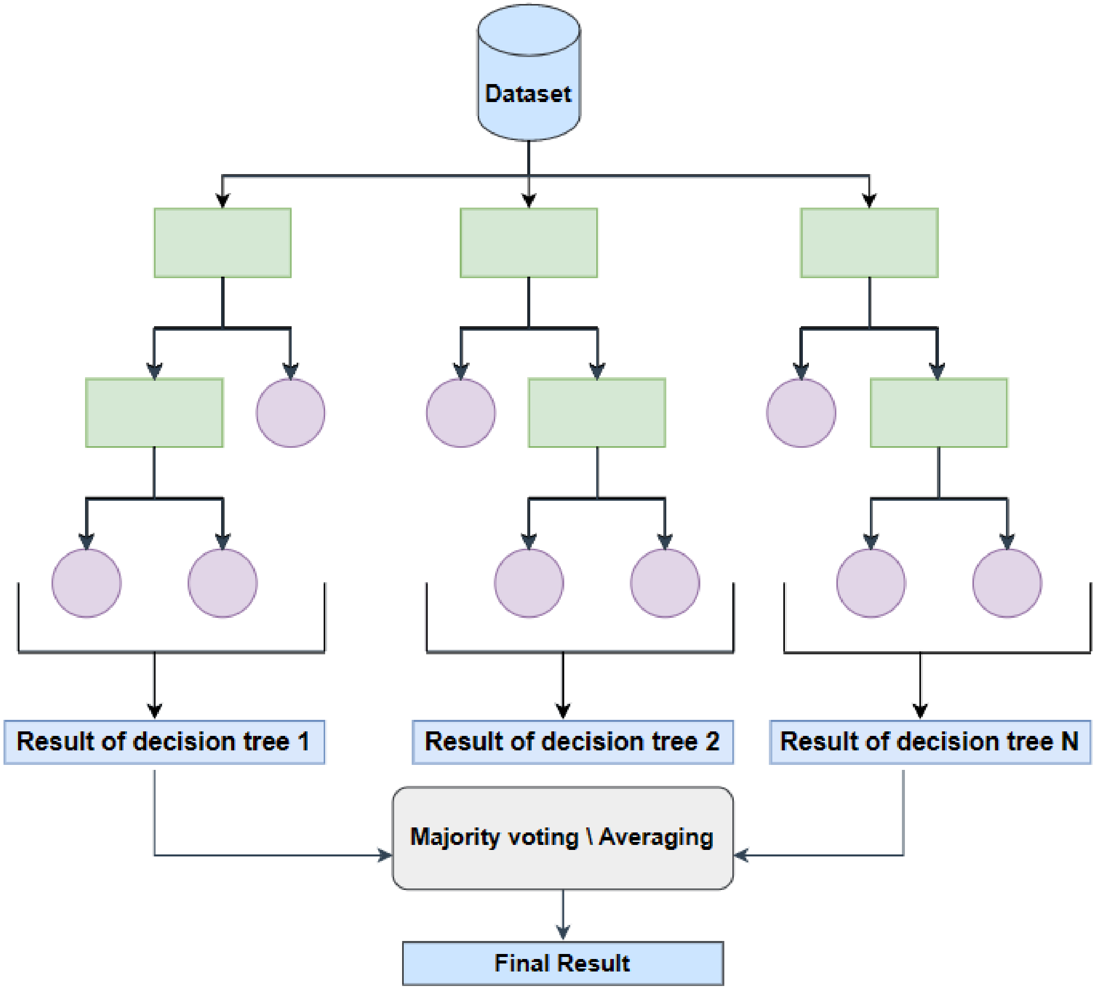

Fig. 5. Random Forest

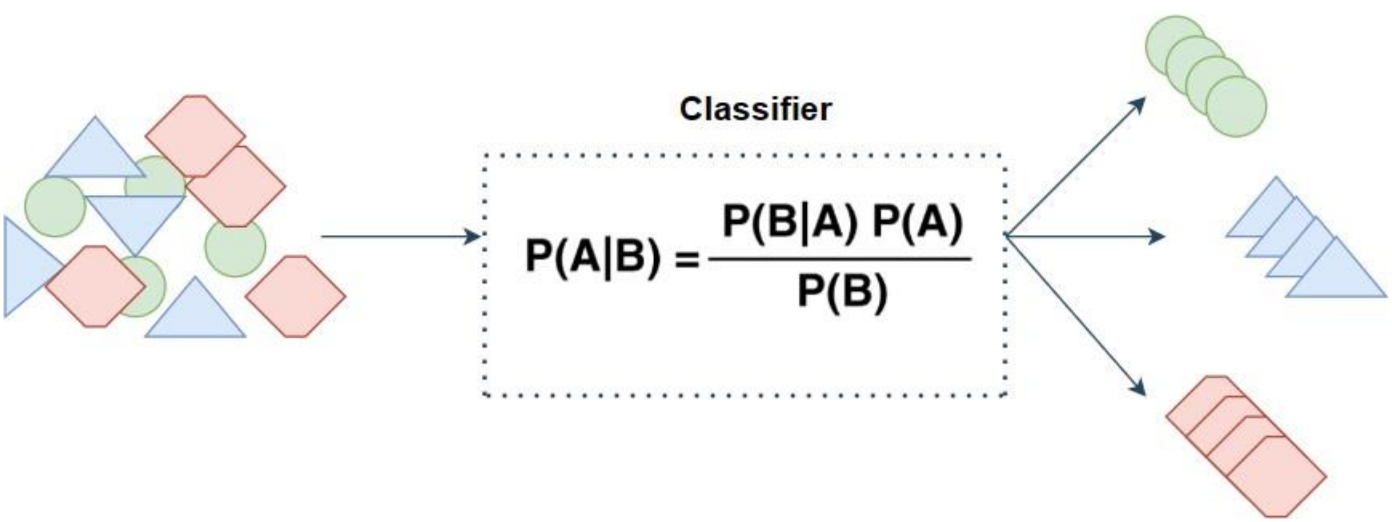

Fig. 6. Na¨ıve Bayes

TABLE I
DIFFERENT RESEARCHERS' APPROACHES TO HEART FAILURE PREDICTION

Reference	Year	Dataset	Techniques	Accuracy
[8]	2017	Two datasets: CLHF and Human Hospital (HH)	Enhanced RF	CLHF: 91.6% , HH: 97%
[9]	2018	CLHF dataset	Various ML algorithms, including RF, NB, KNN, and LR	LR achieved a high accuracy of 89%
[10]	2019	UCIHF dataset	Various ML algorithms, including RF, LR, DT, and NB	SVM achieved a high accuracy of 84.85%
[11]	2019	Stat-log HF and Framingham HF datasets	Various ML algorithms, including LR, SVM, and NB	LR achieved a high accuracy of 86.32% and 81.48% on log HF and Framingham HF datasets respectively
[12]	2019	UCIHF dataset	Enhanced NB, Minimum Sequencing Optimization	89.77%
[13]	2020	CLHF dataset	Various ML algorithms, including RF, LR, DT, KNN, and NB	SVM achieved a high accuracy of 81%
[14]	2020	UCIHF dataset	Various ML algorithms, including SVM, DT, KNN, and LR	KNN achieved a high accuracy of 87%
[15]	2020	CLHF dataset	Various ML algorithms, including RF, DT, KNN, and NB	KNN achieved a high accuracy of 90.78%
[16]	2021	UCIHF dataset	Various ML algorithms, including RF, LR, DT, KNN, and NB	RF achieved a high accuracy of 88%
[17]	2021	CLHF dataset	Hyberied ML algorithms (RF + DT)	RF + DT model achieved an accuracy of 88.70%
[18]	2021	Comprehensive HF dataset (two datasets)	Various ML algorithms, including RF, LR, NB, KNN, DT and SVM	RF achieved a high accuracy of 94%
[19]	2021	UCIHF dataset	Various ML algorithms, including RF, LR, NB, KNN, DT and SVM	RF achieved a high accuracy of 95.60%
[20]	2021	CLHF dataset	Various ML algorithms, including RF, LR, NB, KNN, DT and SVM	RF achieved a high accuracy of 95.60%
[21]	2022	Comprehensive HF dataset	Various ML algorithms, including RF, LR, and SVM	RF achieved a high accuracy of 92.9%
[22]	2022	Three datasets(Mendeley HF, IEEE HF, CLHF)	Various ML algorithms, including RF, NB, and SVM	96.75%, 93.39%, and 88.24% on Mendeley, IEEE, and CLHF datasets respectively
[23]	2023	UCIHF datasets	Various ML algorithms, including RF, LR, NB, KNN, DT, and SVM with neural network methods	LR achieved a high accuracy of 86%
[24]	2023	Comprehensive HF dataset (five HF datasets)	Enhanced DT	DT achieved an accuracy of 87.25%
[25]	2023	CLHF datasets	Various ML algorithms, including RF, LR, DT, KNN, and SVM	LR and SVM achieved a high accuracy of 88.52%
[26]	2023	UCIHF datasets	Various ML algorithms, including RF, LR, NB, KNN, and SVM	Enhanced RF achieved a high accuracy of 98.49%

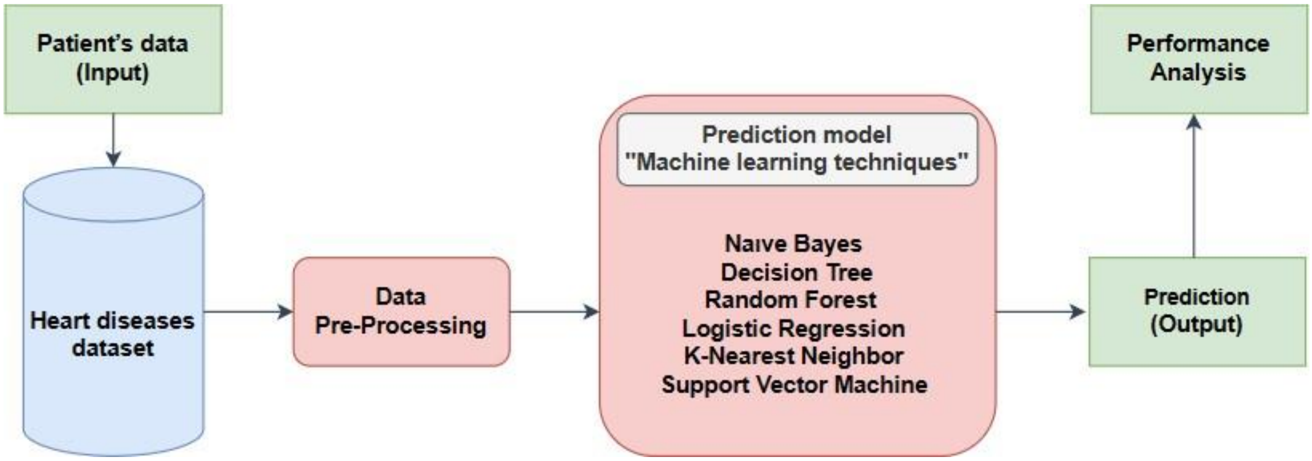

Fig. 7. General heart disease prediction system process

2023, Arushi Jain et al. [27] used a dataset of images related to cardiac disease to suggest a new technique based on CNN to predict HF. In the data preparation step, they reduced the size of the images to handle the complexity. The accuracy of the suggested method was 95.74 percent. Despite the high accuracy, we need to test this proposed method on larger datasets, as the dataset approved for this method contains images of 500 patients and was divided into a part for training, a part for testing, and another part for validation (300, 100, 100).

*2) **Long Short Term Memory Networks (LSTMs)**:* Muhannad Al-Khodari and Louay Fariwan [28] relied on heart sounds to diagnose HF. CNN was used with LSTM, with some steps in the data preparation stage. This system has been tested on the PhysioNet dataset with an accuracy of 87.31%. But this accuracy needs to be increased in order to be able to rely on this method of diagnosis based on this data.

*3) **Generative Adversarial Networks (GANs)**:* To increase the volume of data, Raniya R. Sarra et al. [29] used GAN to generate more examples of people with and without HF based on two different models for characterising a person with or without HF: CNN and LSTM. The proposed method was tested on three datasets, including the CLHF dataset, which achieved 99.3% and 99.1% on LSTM and CNN with GAN, respectively. However, various DL algorithms, in addition to those suggested in this research, can promptly and precisely detect HF. It is possible to deploy the suggested DL algorithms of this research on cloud-based infrastructure to test this method on real data.

*4) **Recurrent Neural Networks (RNNs)**:* S. Chitra and V. Jayalakshmi [30] relied on the UCIHF dataset to predict HF. To extract the most important features from these data, they used the KNN algorithm, and then based on these features, they diagnosed a patient or not. RNN has been used to diagnose abnormal people (heart patients) based on these features. The proposed method achieved an efficiency of 96%. From this study, we find that when applying more than one algorithm to obtain the final model, the efficiency increases. Also, the features that we depend on greatly affect the final efficiency of the model, so more than one method must be applied to choose the essential features and then choose among them the best algorithm that gives the best efficiency.

*5) **Self Organizing Maps (SOMs)**:* Adisha Rath et al. [31] developed a new method for predicting HF based on the electrocardiogram (ECG) using the PTB and MIT datasets. More than one DL model was tested, and based on the results, it turned out that SOM with an autoencoder gave the best results on the two datasets. scored 99.2% and 98.4% on the PTB and MIT datasets, respectively. The proposed system can be tested on real data by developing a platform using this approach that allows anyone to obtain a diagnosis of their heart condition and then use this real data to increase the efficiency of the system.

*6) **Radial Basis Function Networks (RBFNs)**:* Abhijit Reddy Peravulu et al. [32] relied on the UCIHF dataset to predict HF. These data have a wide range of features, of which only the 14 most important were selected for diagnosing a sick or non-ill person. To build the model, various DL algorithms were tested, including CNN and RBFN, which achieved accuracy of 98.49% and 98.75%, respectively. In order to trust this method, these algorithms must be tested on other datasets to ensure the validity of the results, and then we can rely on this method for diagnosis based on the required features.

*7) **Multilayer Perceptrons (MLPs)**:* Ali Al-Batayneh and Sarah Manak [33] relied on the CLHF dataset to predict HF based on the features in this data, with some steps to adjust the data in the data preparation stage. To predict HF, the MLP was trained on this dataset after it was prepared. A set of ML algorithms, including DT, KNN, LR, RF, and SVM, was also trained to compare the results with the MLP algorithm. When compared, it turned out that MLP had the best accuracy of 84.61%. Although this model had the best accuracy compared to the ML algorithms, efficiency needs to be increased in order to be able to rely on it for diagnosis. Therefore, the optimizer technique can be used to select the best hyperparameters for this network. Also, a technique must be suggested by which the best features that give the best accuracy are selected, and then this method must be tested on other datasets to verify the results.

*8) **Deep Belief Networks (DBNs):*** Syed Arsalan Ali et al. [34] also used the same CLHF dataset to develop a method that predicts HF. However, in contrast to the previous study [33], a Ruzzo-Tompa strategy was used to select the best features that give the best accuracy, and a genetic algorithm with stacking was used to obtain the best hyperparameters for their network. To predict HF based on the most important features chosen, DBN was trained on these features with an accuracy of 94.61%. In spite of this, the time taken for the diagnosis must also be calculated for this proposed method, as it was not specified in this study. This proposed method must also be tested on examples from other datasets to verify the results, and then we can rely on it in future diagnostics.

III. Description of the datasets

In this study, more than one dataset was mentioned, so in this section, information about these datasets and how to access them will be provided.

*A. **Cleveland HF (CLHF)***

The 303 HF patients in the CLHF data were measured for 14 different attributes (age, fasting blood sugar, and so on). The people were divided into groups according to their severity of cardiac disease. This [35] referrer can be used to access this dataset.

*B. **University of California Irvine HF (UCIHF)***

One of the most important datasets used in predicting HF in many studies There are 76 characteristics in this database; however, only a portion of 14 of them are used in all published research. This dataset contains attributes such as age, blood pressure, and so on. This referrer [36] can be used to find out the rest of the features and download this dataset.

*C. **Stat-log HF***

This data is similar to the CLHF dataset but differs slightly, as it contains 13 attributes (age, fasting blood sugar, and so on) taken from 270 objects to classify a person with HF or not. You can get this dataset and learn more about the remaining features by using this referral link, [37].

*D. **Framingham HF***

One of the most significant datasets utilised in the diagnosis of HF patients, the Framingham HF dataset constitutes more than 4240 examples, and each example contains 15 features (age, current smoker, and so on). Use this referral link, [38], to obtain this dataset and discover more about the other features.

*E. **Mendeley HF***

This dataset on heart illness was obtained from one of India's multispecialty clinics. It is one of the HF datasets now available for studies because it has 1000 objects over 14 common characteristics (age, fasting blood sugar, and so on). This dataset can be used to create predictive AI models to classify people who have HF. To learn about the remaining features and obtain this dataset, use this referral [39].

*F. **IEEE HF (Comprehensive dataset)***

This dataset contains an aggregated set of datasets for the classification of HF. After merging five datasets, they got 1190 objects with 11 attributes (age, fasting blood sugar, and so on). In order to further investigate HF diagnosis using different AI techniques, these datasets were collected and compiled into a single source. Reference [40] can be used to access this source and obtain this data.

*G. **PhysioNet HF***

This data is so far considered one of the most important datasets used in many research studies in order to reach the highest accuracy in diagnosing HF. This dataset was compiled to diagnose people with HF based on heart sounds. This data collection consists of five subsets, A to E, which collectively contain 3,126 recordings with durations ranging from 5 seconds to a little over 120 seconds. You can visit this source [41] and get these datasets.

*H. **PTB ECG***

The diagnosis of HF by ECG is common in many studies. So this data was based on the diagnosis of HF based on the ECG. To make up this dataset, 549 entries from 290 people were recorded. To access this source and get this dataset, go to reference [42].

*I. **MIT ECG***

This dataset also depends on diagnosing HF based on the ECG. This dataset contains 48 recordings collected from 47 people, each lasting half an hour. This data is available through this reference [43] for access and use.

IV. Findings and main results

*A. **Performance Analysis***

Comparing the algorithms in Table I, we found that the RF algorithm gave the highest accuracy in the CLHF dataset among all the ML algorithms that were applied in different research projects, including SVM, RF, LR, DT, KNN, and NB. Table II and Figure 8 illustrate this comparison.

The comparison shown in the digram 9 shows the accuracy that researchers have achieved through ML algorithms to diagnose HF diseases in previous years until now.

For the DL algorithms used in HF diagnosis, we find that LSTM and CNN with GAN gave the highest accuracy, followed by RBFN, which gave an accuracy of 98.75%. This comparison is shown in Table III and Diagram 10.

After this comparison, the highest accuracy ML algorithm was chosen and compared based on the most common standards with the highest accuracy DL algorithm, as shown in Table IV and Diagram 12.

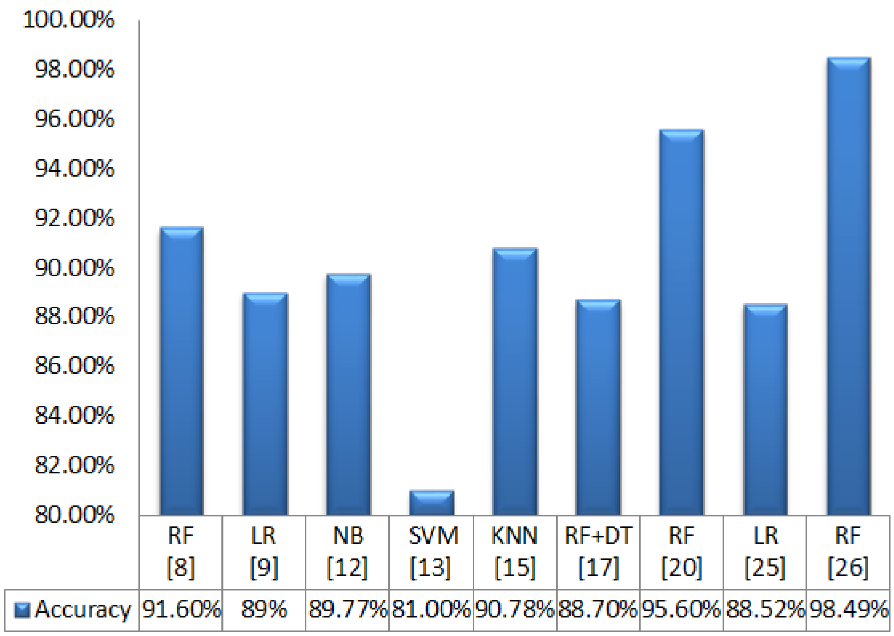

Fig. 8. Comparison of different ML algorithms

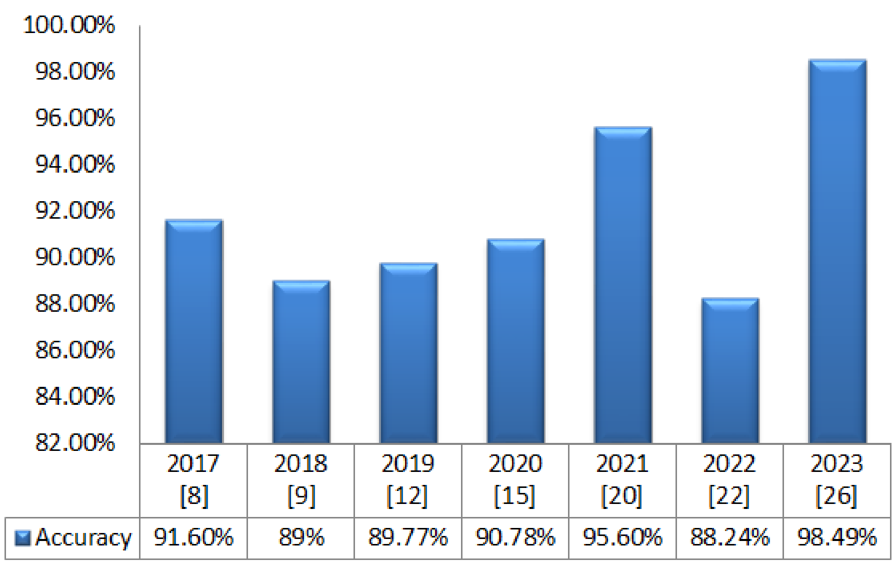

Fig. 9. A comparison of previous studies from 2017 to 2023 using machine learning algorithms

B. *Major Findings*

- There are three most common methods for detecting HF diseases using AI techniques: either through heart sounds or electrocardiograms or through a set of data that is entered, such as age, blood pressure, and so on. In this survey, what the researchers have achieved with each method has been presented.
- Based on these methods, a set of datasets was reached for each method. Where details about this dataset and how to access it were presented to be used in improving the diagnosis of HF diseases.
- Based on these datasets, a set of ML and DL algorithms have been shown and compared.
- When comparing the ML and DL algorithms in the CLHF dataset, we found that the RF algorithm gave the best accuracy among the ML algorithms. While LSTM and CNN with GAN gave the best accuracy compared to other DL algorithms on the same dataset.
- Finally, when we compare RF (ML model) with LSTM and CNN (DL model), the DL model is superior in terms of accuracy. A host of other criteria are shown in Table IV.
- When we study the most efficient method, we find that to reach high diagnostic efficiency, we must rely on a large and balanced dataset between the number of patients and non-patients. Data should be appropriately prepared

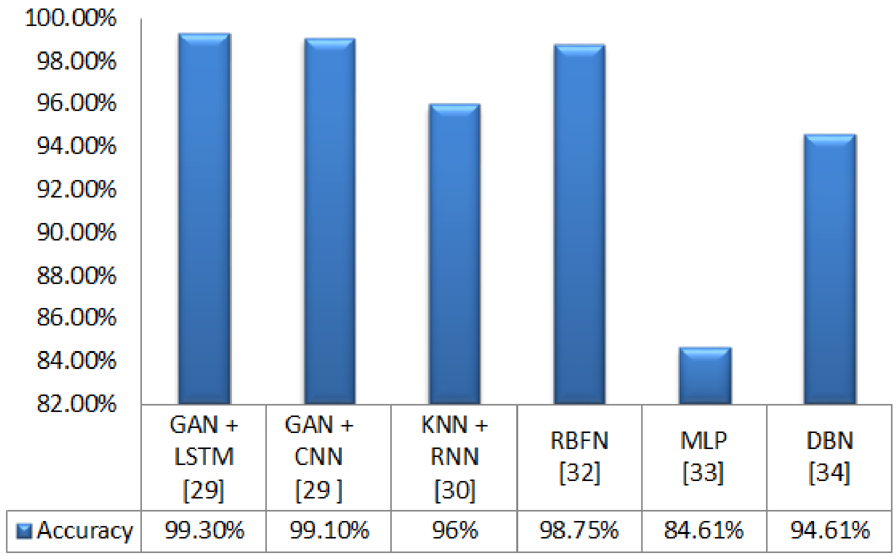

Fig. 10. Comparison of different DL algorithms

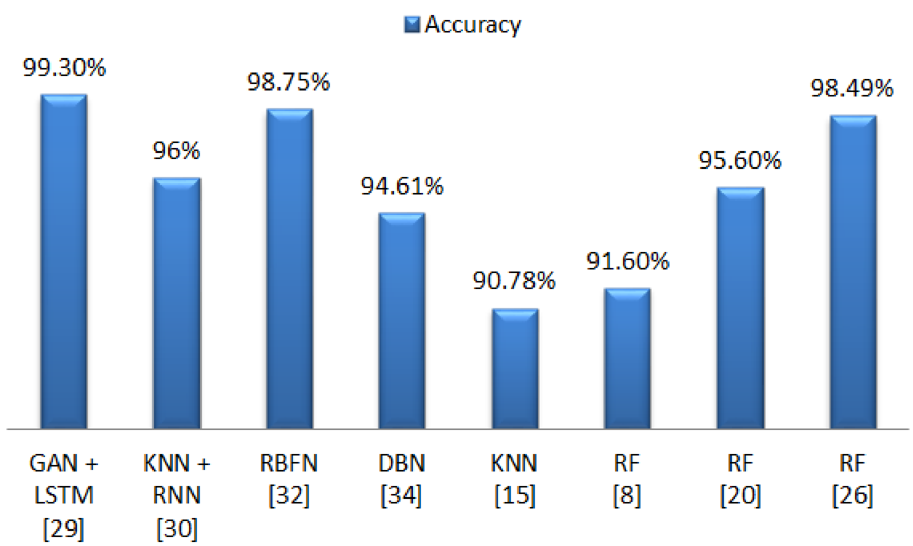

Fig. 11. Comparison of DL and ML algorithms

according to its nature. We must rely on a suitable method to select or extract the most important features that we can rely on for the correct diagnosis. Finally, we have to choose the appropriate algorithm for classification based on these features.

TABLE II
ACCURACY OF DIFFERENT ML ALGORITHMS

Reference	Algorithm	Accuracy
[26]	RF	98.49%
[19]	RF	95.6%
[8]	RF	91.6%
[15]	KNN	90.78%
[12]	NB	89.77%
[9]	LR	89%
[17]	RF+DT	88.7%
[13]	SVM	81%

TABLE III
ACCURACY OF DIFFERENT DL ALGORITHMS

Reference	Algorithm	Accuracy
[29]	GAN + LSTM	99.3%
[29]	GAN + CNN	99.10%
[32]	RBFN	98.75%
[30]	KNN + RNN	96%
[34]	DBN	94.61%
[33]	MLP	84.61%

TABLE IV
COMPARISON OF THE HIGHEST MODEL IN DL AND ML, BASED ON THE MOST FAMOUS CRITERIA

Reference	Algorithm	Accuracy	Specificity	Sensitivity
[29]	GAN + LSTM	99.3%	99.2%	99.3%
[29]	GAN + CNN	99.10%	99.10%	99.10%
[26]	RF	98.49%	96.8%	100%

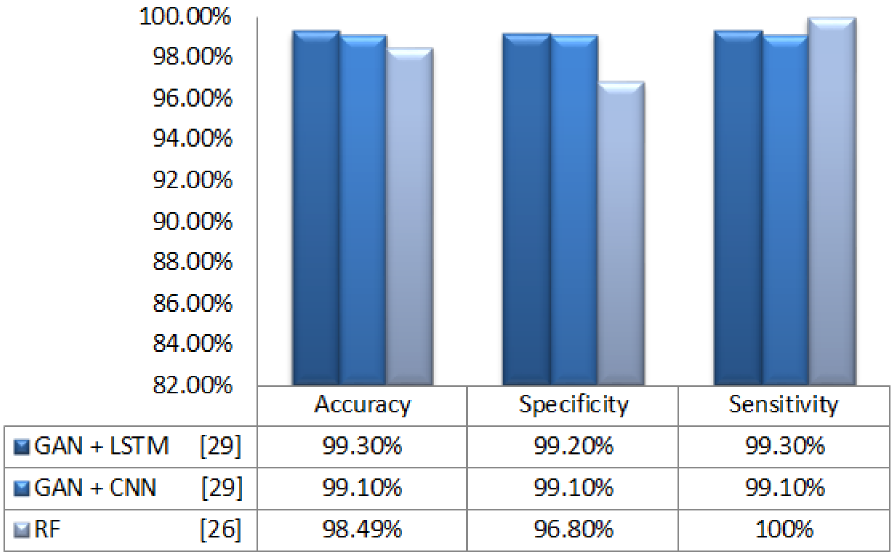

Fig. 12. Comparison of the highest model in DL and ML

V. Conclusion and Future work

HF is one of the most common diseases around the world, as this disease causes the deaths of many people every year. There is no doubt that early and accurate detection of this disease helps us treat it effectively. Since AI has become involved in all fields, especially in the field of health, to detect disease, many researchers have used AI methods to reach a method that helps us in the early and rapid detection of HF. But we still need to find the best method that gives us the best efficiency in diagnosis because the problem is related to the field of health. In this survey, a group of methods based on AI were presented to diagnose HF diseases to provide the reader with all the information about these methods to reach the best method and then improve it to reach the best and fastest reliable method in the future for diagnosing HF diseases. In this survey, an accurate comparison was provided between all the most efficient methods based on the most common criteria among them and on the same dataset. Information has been provided on groups of datasets and how to access them, as they can be relied upon to develop a method that helps us detect HF diseases.

In the future, a group of datasets mentioned in this survey will be combined to obtain large and balanced data between the number of patients and non-patients and rely on them to develop a new method that is more efficient than the methods mentioned in this survey for the detection of HF diseases. Based on these data, the best technique will be developed through which the most reliable features can be extracted or selected and then used to classify whether a person has HF or not.

References

[1] World Health Organization (WHO). Cardiovascular diseases (cvds). https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2017. Online; accessed 26 April 2023.

[2] Ekta Maini, Bondu Venkateswarlu, Baljeet Maini, and Dheeraj Marwaha. Machine learning–based heart disease prediction system for indian population: An exploratory study done in south india. *medical journal armed forces india*, 77(3):302–311, 2021.

[3] M Ganga, N Janakiraman, Arun Kumar Sivaraman, Rajiv Vincent, A Muralidhar, and Priya Ravindran. An effective denoising and enhancement strategy for medical image using rl-gl-caputo method. In *Advances in Parallel Computing (Smart Intelligent Computing and Communication Technology)*, volume 38, pages 402–408. IOS Press, 2021.

[4] Ying Zhou, Si Chen, Zhenqi Rao, Dong Yang, Xiang Liu, Nianguo Dong, and Fei Li. Prediction of 1-year mortality after heart transplantation using machine learning approaches: A single-center study from china. *International Journal of Cardiology*, 339:21–27, 2021.

[5] Rachana R Sanni and HS Guruprasad. Analysis of performance metrics of heart failured patients using python and machine learning algorithms. *Global transitions proceedings*, 2(2):233–237, 2021.

[6] Davide Chicco and Giuseppe Jurman. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, 20(1):1–16, 2020.

[7] Hager Ahmed, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A Ali. Heart disease identification from patients' social posts, machine learning solution on spark. *Future Generation Computer Systems*, 111:714–722, 2020.

[8] Shan Xu, Zhen Zhang, Daoxian Wang, Junfeng Hu, Xiaohui Duan, and Tiangang Zhu. Cardiovascular risk prediction method based on cfs subset evaluation and random forest classification framework. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pages 228–232. IEEE, 2017.

[9] Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun. A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile information systems*, 2018:1–21, 2018.

[10] Saba Bashir, Zain Sikander Khan, Farhan Hassan Khan, Aitzaz Anjum, and Khurram Bashir. Improving heart disease prediction using feature selection approaches. In *2019 16th international bhurban conference on applied sciences and technology (IBCAST)*, pages 619–623. IEEE, 2019.

[11] Jyostna Devi Bodapati, Venkata Rama Krishna Sajja, Nirupama Bhat Mundukur, and Naralasetti Veeranjaneyulu. Robust cluster-then-label (rctl) approach for heart disease prediction. *Inge´nierie des Syste`mes d Inf.*, 24(3):255–260, 2019.

[12] Anjan Nikhil Repaka, Sai Deepak Ravikanti, and Ramya G Franklin. Design and implementing heart disease prediction using naives bayesian. In *2019 3rd International conference on trends in electronics and informatics (ICOEI)*, pages 292–297. IEEE, 2019.

[13] Jian Ping Li, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan,

and Abdus Saboor. Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*, 8:107562–107582, 2020.

[14] Archana Singh and Rakesh Kumar. Heart disease prediction using machine learning algorithms. In *2020 international conference on electrical and electronics engineering (ICE3)*, pages 452–457. IEEE, 2020.

[15] Devansh Shah, Samir Patel, and Santosh Kumar Bharti. Heart disease prediction using machine learning techniques. *SN Computer Science*, 1:1–6, 2020.

[16] Rony Chowdhury Ripan, Iqbal H Sarker, Syed Md Minhaz Hossain, Md Musfique Anwar, Raza Nowrozy, Mohammed Moshiul Hoque, and Md Hasan Furhad. A data-driven heart disease prediction model through k-means clustering-based anomaly detection. *SN Computer Science*, 2:1–12, 2021.

[17] M Kavitha, G Gnaneswar, R Dinesh, Y Rohith Sai, and R Sai Suraj. Heart disease prediction using hybrid machine learning model. In *2021 6th international conference on inventive computation technologies (ICICT)*, pages 1329–1333. IEEE, 2021.

[18] Hana H Alalawi and S Alsuwat Manal. Detection of cardiovascular disease using machine learning classification models. *International Journal of Engineering Research & Technology (IJERT) ISSN*, pages 2278–0181, 2021.

[19] Rahul Katarya and Sunit Kumar Meena. Machine learning techniques for heart disease prediction: a comparative study and analysis. *Health and Technology*, 11:87–97, 2021.

[20] Pooja Rani, Rajneesh Kumar, Nada MO Sid Ahmed, and Anurag Jain. A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments*, 7(3):263–275, 2021.

[21] Ru¨stem YILMAZ and Fatma Hilal YAG˘ IN. Early detection of coronary heart disease based on machine learning methods. *Medical Records*, 4(1):1–6, 2022.

[22] Bhanu Prakash Doppala, Debnath Bhattacharyya, Midhunchakkaravarthy Janarthanan, and Namkyun Baik. A reliable machine intelligence model for accurate identification of cardiovascular diseases using ensemble techniques. *Journal of Healthcare Engineering*, 2022, 2022.

[23] Vijay Mane, Yash Tobre, Swapnil Bonde, Arya Patil, and Parth Sakhare. Heart disease prediction using machine learning and neural networks. In *Computer Vision and Robotics: Proceedings of CVR 2022*, pages 205–228. Springer, 2023.

[24] Mert Ozcan and Serhat Peker. A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics*, 3:100130, 2023.

[25] Huating Sun and Jianan Pan. Heart disease prediction using machine learning algorithms with self-measurable physical condition indicators. *Journal of Data Analysis and Information Processing*, 11(1):1–10, 2023.

[26] Kritika Tiwari, Sajid Ansari, and Govind Kushwaha. Earlier heart disease prediction system using machine learning. In Praveen Kumar Shukla, Krishna Pratap Singh, Ashish Kumar Tripathi, and Andries Engelbrecht, editors, *Computer Vision and Robotics*, pages 191–204. Springer Nature Singapore, 2023.

[27] Arushi Jain, Annavarapu Chandra Sekhara Rao, Praphula Kumar Jain, and Yu-Chen Hu. Optimized levy flight model for heart disease prediction using cnn framework in big data application. *Expert Systems with Applications*, 223:119859, 2023.

[28] Mohanad Alkhodari and Luay Fraiwan. Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings. *Computer Methods and Programs in Biomedicine*, 200:105940, 2021.

[29] Raniya R Sarra, Ahmed M Dinar, Mazin Abed Mohammed, Mohd Khanapi Abd Ghani, and Marwan Ali Albahar. A robust framework for data generative and heart disease prediction based on efficient deep learning models. *Diagnostics*, 12(12):2899, 2022.

[30] S Chitra and V Jayalakshmi. Prediction of heart disease and chronic kidney disease based on internet of things using rnn algorithm. In *Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1*, pages 467–479. Springer, 2022.

[31] Adyasha Rath, Debahuti Mishra, Ganapati Panda, Suresh Chandra Satapathy, and Kaijian Xia. Improved heart disease detection from ecg signal using deep learning based ensemble model. *Sustainable Computing: Informatics and Systems*, 35:100732, 2022.

[32] Rakibul Islam, Abhijit Reddy Beeravolu, Md Al Habib Islam, Asif Karim, Sami Azam, and Sanzida Akter Mukti. A performance based study on deep learning algorithms in the efficient prediction of heart disease. In *2021 2nd International Informatics and Software Engineering Conference (IISEC)*, pages 1–6. IEEE, 2021.

[33] Ali Al Bataineh and Sarah Manacek. Mlp-pso hybrid algorithm for heart disease prediction. *Journal of Personalized Medicine*, 12(8):1208, 2022.

[34] Syed Arslan Ali, Basit Raza, Ahmad Kamran Malik, Ahmad Raza Shahid, Muhammad Faheem, Hani Alquhayz, and Yogan Jaya Kumar. An optimally configured and improved deep belief network (oci-dbn) approach for heart disease prediction based on ruzzo–tompa and stacked genetic algorithm. *IEEE Access*, 8:65947–65958, 2020.

[35] Matthias Pfisterer Andras Janosi, William Steinbrunn and Robert Detrano. Heart disease cleveland uci. https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci, 1988. Online; accessed 22 April 2023.

[36] Matthias Pfisterer Andras Janosi, William Steinbrunn and Robert Detrano. Uci heart disease data. https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data, 1988. Online; accessed 20 April 2023.

[37] Dheeru Dua and Casey Graff. UCI machine learning repository. https://archive.ics.uci.edu/ml/datasets/statlog+(heart), 2017. Online; accessed 22 April 2023.

[38] Ashish Bhardwaj. Framingham heart study dataset. https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset, 1948. Online; accessed 26 April 2023.

[39] Bhanu Prakash Doppala and Debnath Bhattacharyya. Cardiovascular_disease_dataset. https://data.mendeley.com/datasets/dzz48mvjht/1, 2021. Online; accessed 27 April 2023.

[40] Manu Siddhartha. Heart disease dataset (comprehensive). https://dx.doi.org/10.21227/dz4t-cm36, 2020. Online; accessed 27 April 2023.

[41] Moukadem A and Dieterlen A. Classification of heart sound recordings - the physionet computing in cardiology challenge 2016. https://physionet.org/content/challenge-2016/1.0.0/, 2016. Online; accessed 28 April 2023.

[42] Goldberger and A Amaral. Components of a new research resource for complex physiologic signals. https://www.physionet.org/content/ptbdb/1.0.0/, 2000. Online; accessed 28 April 2023.

[43] Goldberger and A Amaral. Components of a new research resource for complex physiologic signals. https://physionet.org/content/mitdb/1.0.0/, 2000. Online; accessed 29 April 2023.