Diagnosis of Heart Failure using AI Techniques

Roopali Gupta
Department of information technology
IIMT college of engineering,
Greater Noida, India
roopaligoyal.gupta@gmail.com

Jyoti Rawat
VertexPlus Technologies Limited,
Gopalpura Road Jaipur,
Rajasthan, India
drjyotirawat19@gmail.com

Indrajeet Kumar
CSE Department
Graphic Era Hill University
Dehradun, India
erindrajeet@gmail.com

Teekam Singh
CSE Department
Graphic Era Deemed to be University
Dehradun, India
tsingh@ma.iitr.ac.in

Manvi Bohra
Graphic Era Hill University
Dehradun, India
manvibohra321@gmail.com

Devendra kumar
ABES Engineering College,
Ghaziabad, Uttar Pradesh 201009
devendra.arya@gmail.com

***Abstract*— Human body contains many organs, but one of the prime organs is the heart, which is a network of vessels which help the blood to flow into our body. Cardiovascular disease is any disturbance in the heart's normal functioning or performance. In this work, closely worked with heart disease prediction and for that, the team will be looking into the heart disease dataset from where work will be derived various perceptions that help us to know the value of each feature and their connections. Also, here the aim is to detect the possibility of a person be affected by a savior heart disease or not. This research paper presents various algorithms of machine learning by which a projection of heart disease can be made and required steps can be taken as early as possible for prevention. This paper consists of five algorithms (SVM, Decision Tree, K-NN, Ensemble Method and ANN) which are applied to the dataset that is publicly available. On applying Ensemble Method (combination of K-NN, Decision Tree, and SVM) on the dataset an accuracy of 89.88% was attained. ANN was able to outperform all the other algorithms with accuracy of 95.33%.**

Keywords*— Heart Disease, Machine Learning, SVM, K-NN, Decision Tree (DT), Ensemble Model, ANN.

I. INTRODUCTION

Heart is a principal organ for a human to live. Major functions of heart are to circulate the blood throughout the body, maintain the blood pressure and many more. Heart failure and heart disorders are the biggest cause of death worldwide. As per WHO (World health organization), 17.9 million people die each year from cardiovascular disease [2]. Major causes of these heart diseases are the regular habits of an individual like their caffeine and alcohol intake, smoking, lack of physical activities etc. Most of the times cardiovascular diseases can be prevented by choosing a healthier lifestyle and taking other measure, for which making an early and accurate prediction of the cardiovascular diseases is extremely important.

In this work we will be comparing accuracy of different ML (Machine learning) algorithms and based on this comparison will be seeing which one is the best among all [27, 28, 29]. Machine learning is a technique which can be used when huge amount of data is available. It learns from the data i.e., it looks at the trends and then based on these trends further predictions can be made. Many types of ML algorithms are present. ML analyses the characteristics of dataset and checks if a person has a heart disease [11,30, 31].

The dataset is split into testing and training portion. In this work we have implemented SVM, K-NN, Decision Tree, Ensemble model and ANN. The main goal of this paper is to identify a well-functioning model for the prediction of heart disease. Further section 2 of the paper consists of literature review followed by methodology in section 3, result in section 4 and section 5 having conclusion and section 6 are the reference. Table 1 consist of all the information collected from different sources for the review of heart disease prediction methods.

Table 1. Literature collected from various studies of heart disease prediction.

Author	Year	ummary	Dataset	Classifier	Accuracy
Rajdhan et al. [1]	2020	Compared different machine learning and found the best one among those for the prediction.	UCI repository	NB, DT, Logistic Regression and RF	90.16 %
Shah et al. [2]	2020	Predicted that weather a patient will have a heart disease or not by the help of ML algorithms.	UCI repository	NB, DT, K-NN, Random Forest Tree	90.78 %
Singh et al. [3]	2020	Implemented four ML algorithms and found their prediction accuracy then choose the best one among them for prediction.	UCI repository	Linear Regression, DT, SVM and K-NN	87.00 %
Kavitha et al. [4]	2021	Created a hybrid model by combining two different ML algorithms.	Cleaveland Heart Disease Dataset	DT, RF and Hybrid model (RF+DT)	88.00 %
Bharti et al. [5]	2021	Machine Learning algorithms is used along with deep learning to enhance the results of the prediction.	UCI repository	Combination of ML and DL	94.20%

979-8-3503-2919-3/23/$31.00 ©2023 IEEE

2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET) | 979-8-3503-2919-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSEIET58677.2023.10303428

Katarya et al. [6]	2020	Used supervised machine learning for heart disease prediction.	UCI repository	ANN, DT, RF, SVM, NB and K-NN	--
Ramprakah et al. [7]	2020	Deep Neural Network is used with X2 statistical model on the dataset for the prediction.	UCI repository	Deep neural networks, X2-statistical model and machine learning	--
Sharma et al. [8]	2020	Demonstration of Talos hyper-parameterized optimizer to know that weather it is more efficient than other ML algorithms or not.	UCI repository	Logistic Regression, SVM, NB, K-NN Hyper-parameter optimization, RF	90.78 %
Yadav et al. [9]	2020	Used four different types of trees algorithm and performed three experiments to make the conclusion	UCI repository	M5P, RF Tree, Reduced Error Pruning and Random Forest ensemble method	--
Waigi et al. [10]	2020	Implemented ML classification algorithm to predict the risk of heart disease.	Kaggle	Decision Tree	73.00 %
Princy et al. [11]	2020	To predict the cardiac disease, they have implemented six supervised ML classifiers and found the best one among them to predict the disease.	Kaggle	NB, DT, Logistic Regression, Random Forest, SVM, K-NN	--
Terrada et al. [12]	2020	Worked with ANN and Adaptive Boosting to enhance the accuracy.	UCI repository	Machine learning algorithms and ANN	--
Mohan et al. [13]	2019	A hybrid model is made which consists of two ML algorithms and gives better result when compared with them.	Cleaveland Heart Disease Dataste	Random forest and Linear Regression	--
Repaka et al. [14]	2019	Used Naive Bayes ML algorithm and compared it with the result of another algorithm and gets NB as the best one.	UCI repository	NB, Sequential Minimal Optimization, Bayes Net and Multi-Layer Perception	89.77 %
Tarawneh et al. [15]	2019	Applied different ML algorithms and developed a hybrid model with the ones which were selected most of the times.	UCI repository	SVM, NB, DT, K-NNN	89.20 %
Khan et al. [16]	2020	Cloud computing is used with Machine Learning where the data is collected from the cloud and then algorithms are applied. When the required criteria are matched then the trained model is stored in the cloud.	Internet of medical things (IoMT) enabled devices	SVM	93.33 %
Jindal et al. [17]	2021	Two ML algorithms were implemented on the dataset and they found out that these two works better then Naïve bayes algorithm.	UCI repository	Logistic regression and K-NN	88.50 %
Arroyo et al. [18]	2022	Prediction of heart disease is done with the help of ANN along with optimization techniques to enhance the accuracy for the prediction.	cardiovascular disease dataset	Genetic Algorithm and Neural Network	73.43 %

Abdeldjouad et al. [19]	2020	A hybrid approach is used to predict the disease.	UCI repository	Multi-Objective Evolutionary Fuzzy Classifier (MOEFC), and other Fuzzy based methods	--
Motarwar et al. [20]	2020	Cognitive approaches were used a ML framework of five algorithms is implemented on the dataset.	Cleaveland Heart Disease Dataste	RF, NB, SVM, Hoeffding Decision Tree, and Logistic Model Tree (LMT)	95.08 %
Maji et al. [21]	2019	A hybrid model and other algorithms is implemented on the dataset and validated the prediction using tenfold validation test.	UCI repository	DT, ANN and Hybrid Decision Tree model (HDT).	78.14 %
Krishnan et al. [22]	2019	Heart disease prediction is done using supervised algorithms and best two are selected among them.	UCI repository	DT and NB	91.00 %
Jothi et al. [23]	2021	Applied two ML algorithms and compared them with each other.	UCI repository	DT and K-NN	81.00 %
Ramalingam et al. [24]	2018	Surveyed on five ML algorithms and found the best working way of each algorithm.		SVM, K-NN, NB, DT, RF	--
Garg et al. [25]	2021	ML algorithms are implemented on the dataset to predict the heart disease and concluded that ML algorithms predicts accurately and helps in the prediction of diseases.	Kaggle	K-NN and RF	86.86 %

II. Proposed Methodology

Multiple approaches for cardiovascular disease prediction are mentioned in this section including ANN, SVM, DT, ensemble learning and K-NN.

A. Data collection

Data obtained for forecasting heart disease was taken from [26]. The dataset is a labelled dataset consisting of 14 columns which include 13 independent features and 1 target variable. The split up in the dataset is as follows: training set accounts for 75% of data and rest i.e., 25% accounts for the test data. The dataset was checked for the missing value and the outlier, but they were not found. Workflow diagram for basic ML algorithms is given in Fig.1.

III. Results and Discussions

The analysis was performed on the publicly available dataset [26] consisting of 14 features (13 are independent features and one is labelled feature). The data was sliced into two: training set accounting for 75 % and the rest i.e., 25 % accounting for the test set. On the training set multiple models were trained and tested afterwards. The confusion matrices and ROC curves for various applied models are shown in Fig. 2.

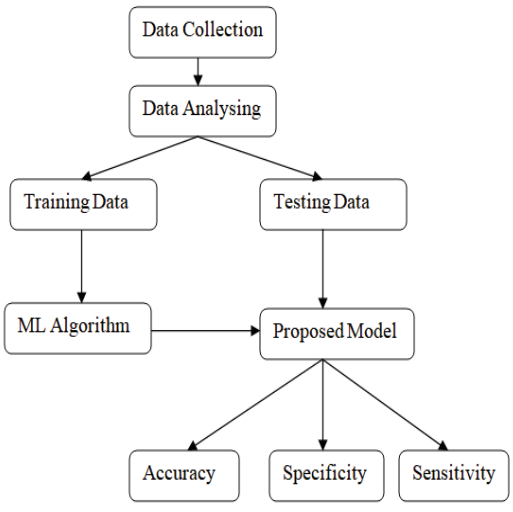
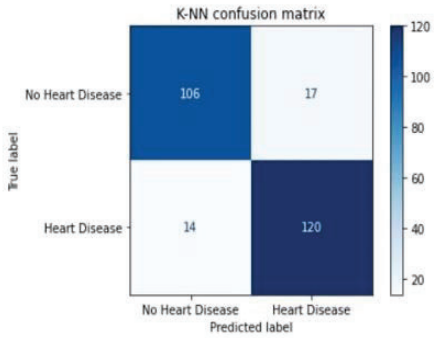
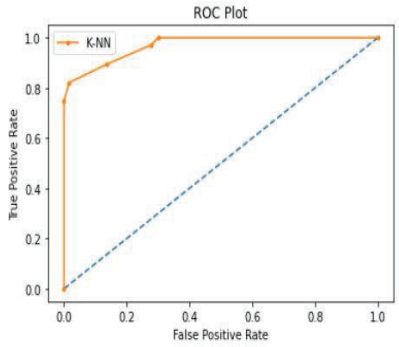


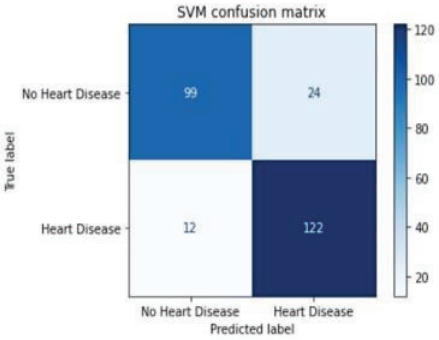
Fig. 1 Flowchart for ML algorithms



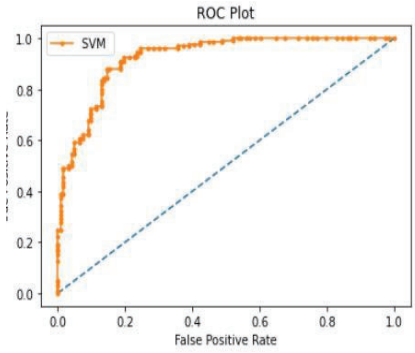
(*a*) K-NN confusion matrix



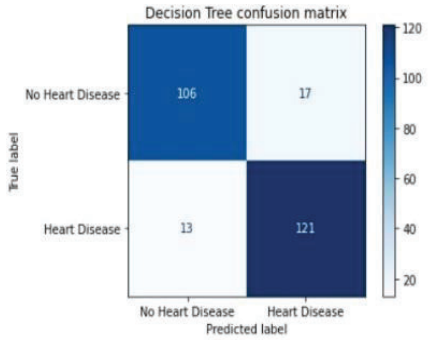
(*b*) K-NN ROC curve



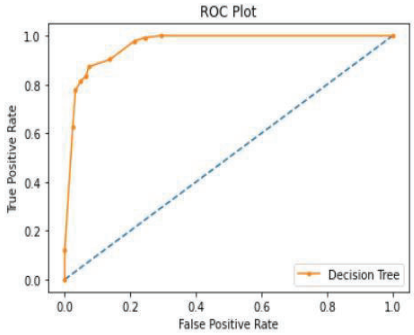
(*c*) SVM confusion matrix



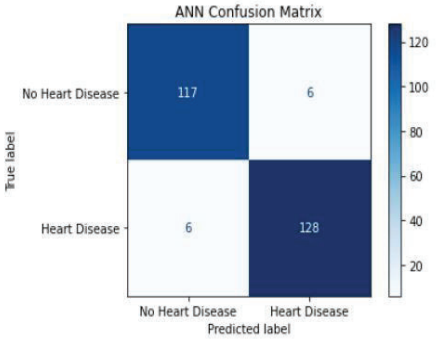
(*d*) SVM ROC curve



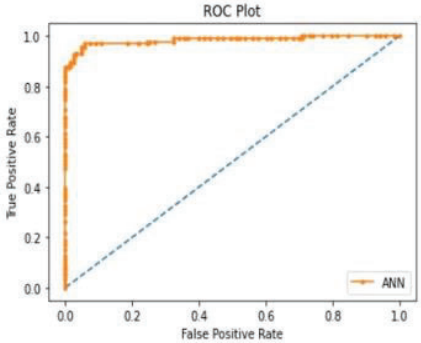
(*e*) Decision tree confusion matrix



(*f*) Decision Tree ROC curve



(*g*) ANN confusion matrix



(*h*) ANN ROC curve

Fig. 2 Confusion Matrices and ROC curves for different applied AI models.

The confusion matrix for the first four models can be seen above; from the test set the patients that are having heart disease and was predicted correctly are 120, 122, 121 and 128 by K-NN, SVM, Decision tree and ANN respectively. Similarly, the patients that are not having heart disease but were predicted as having heart disease are 17, 24, 17 and 6 by K-NN, SVM, Decision tree and ANN respectively. For the models K-NN, SVM, Decision tree and ANN the patients that are not having heart disease and was predicted correctly by the models are 106, 99, 106 and 117 respectively and the patients that are having heart disease but was predicted incorrectly as not having heart disease are 14, 12, 13 and 6 respectively.

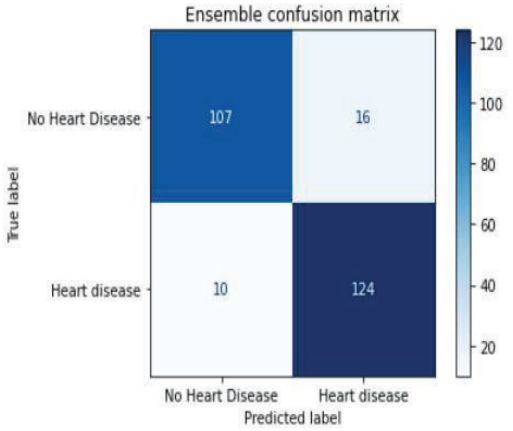


Fig. 3 Confusion matrix for Ensemble Model

Confusion matrix for ensemble classifier can be seen above. Voting classifier is used for ensemble model and the type of voting that is used is hard voting. In hard voting the output is the class having the highest majority votes from each of the classifier. There are 124 patients that are having heart disease and was predicted correctly by the model whereas there are 16 patients that were not having heart disease but was predicted incorrectly as having heart disease by the model. Similarly, there are 107 patients that are not having heart disease and were predicted correctly by the model and 10 patients that are having heart disease but were predicted incorrectly as not having heart disease by the model.

Fig. 3 consist of confusion matrix for ensemble classifier. Ensemble model is implemented with voting classifier. This model consisted of K-NN, Decision tree, SVM.

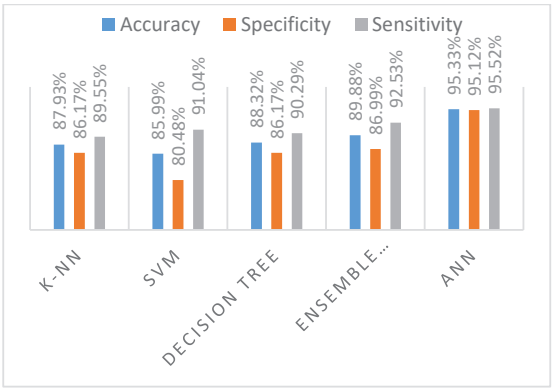


Fig. 4 Performance comparison of different models.

Fig. 4 shows the comparative analysis of the accuracy, sensitivity and specificity performed by different machine learning models (k-NN, SVM, Decision Tree, ANN and Ensemble classifier) on the given dataset.

IV. Conclusion

The objective of the work is to find the optimal method for correct prediction of heart failure cases. Various AI methods can be utilized by the experts to co-relate the diagnosis. The result of the model depends on the accuracy of algorithms. If the accuracy of the models is low, the ouput generated may be wrong or less accurate. Results can be improved by using more data along with multiple feature. In this study 14 main attributes were selected mentioned and five different techniques i.e. SVM, K-NN, ANN, ensemble model, and decision is applied, and it is found that the accuracy of the ANN was the highest i.e., 95.33%.

References

[1] Rajdhan, A., Agarwal, A., Sai, M., Ravi, D. and Ghuli, P., 2020. Heart disease prediction using machine learning. International Journal of Research and Technology, 9(04), pp.659-662.

[2] Shah, D., Patel, S. and Bharti, S.K., 2020. Heart disease prediction using machine learning techniques. SN Computer Science, 1(6), pp.1-6.

[3] Singh, A. and Kumar, R., 2020, February. Heart disease prediction using machine learning algorithms. In 2020 international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE.

[4] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y.R. and Suraj, R.S., 2021, January. Heart disease prediction using hybrid machine learning model. In 2021 6th International Conference on Inventive Computation Technologies (ICICT) (pp. 1329-1333). IEEE.

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

[6] Katarya, R. and Srinivas, P., 2020, July. Predicting heart disease at early stages using machine learning: A survey. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 302-305). IEEE.

[7] Ramprakash, P., Sarumathi, R., Mowriya, R. and Nithyavishnupriya, S., 2020, February. Heart disease prediction using deep neural network. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 666-670). IEEE.

[8] Sharma, S. and Parmar, M., 2020. Heart diseases prediction using deep learning neural network model. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(3), pp.2244-2248.

[9] Yadav, D.C. and Pal, S., 2020. Prediction of heart disease using feature selection and random forest ensemble method. International Journal of Pharmaceutical Research, 12(4), pp.56-66.

[10] Waigi, D., Choudhary, D.S., Fulzele, D.P. and Mishra, D., 2020. Predicting the risk of heart disease using advanced machine learning approach. Eur. J. Mol. Clin. Med, 7(7), pp.1638-1645.

[11] Princy, R.J.P., Parthasarathy, S., Jose, P.S.H., Lakshminarayanan, A.R. and Jeganathan, S., 2020, May. Prediction of cardiac disease using supervised machine learning algorithms. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 570-575). IEEE.

[12] Terrada, O., Cherradi, B., Hamida, S., Raihani, A., Moujahid, H. and Bouattane, O., 2020, September. Prediction of patients with heart disease using artificial neural network and adaptive boosting techniques. In 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet) (pp. 1-6). IEEE.

[13] Mohan, S., Thirumalai, C. and Srivastava, G., 2019. Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, pp.81542-81554.

[14] Repaka, A.N., Ravikanti, S.D. and Franklin, R.G., 2019, April. Design and implementing heart disease prediction using naives Bayesian. In 2019 3rd International conference on trends in electronics and informatics (ICOEI) (pp. 292-297). IEEE.

[15] Tarawneh, M. and Embarak, O., 2019, February. Hybrid approach for heart disease prediction using data mining techniques. In International Conference on Emerging Internetworking, Data & Web Technologies (pp. 447-454). Springer, Cham.

[16] Khan, M.A., Abbas, S., Atta, A., Ditta, A., Alquhayz, H., Khan, M.F. and Naqvi, R.A., 2020. Intelligent cloud based heart disease prediction system empowered with supervised machine learning.

[17] Jindal, H., Agrawal, S., Khera, R., Jain, R. and Nagrath, P., 2021. Heart disease prediction using machine learning algorithms. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing.

[18] Arroyo, J.C.T. and Delima, A.J.P., 2022. An Optimized Neural Network Using Genetic Algorithm for Cardiovascular Disease Prediction. Journal of Advances in Information Technology Vol, 13(1).

[19] Abdeldjouad, F.Z., Brahami, M. and Matta, N., 2020, June. A hybrid approach for heart disease diagnosis and prediction using machine learning techniques. In International conference on smart homes and health telematics (pp. 299-306). Springer, Cham.

[20] Motarwar, P., Duraphe, A., Suganya, G. and Premalatha, M., 2020, February. Cognitive approach for heart disease prediction using machine learning. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-5). IEEE.

[21] Maji, S. and Arora, S., 2019. Decision tree algorithms for prediction of heart disease. In Information and communication technology for competitive strategies (pp. 447-454). Springer, Singapore.

[22] Krishnan, S. and Geetha, S., 2019, April. Prediction of Heart Disease Using Machine Learning Algorithms. In 2019 1st international conference on innovations in information and communication technology (ICIICT) (pp. 1-5). IEEE.

[23] Jothi, K.A., Subburam, S., Umadevi, V. and Hemavathy, K., 2021. Heart disease prediction system using machine learning. Materials Today: Proceedings.

[24] Ramalingam, V.V., Dandapath, A. and Raja, M.K., 2018. Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), pp.684-687.

[25] Garg, A., Sharma, B. and Khan, R., 2021. Heart disease prediction using machine learning techniques. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012046). IOP Publishing.

[26] Available at: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset. Last access: 28 Oct, 2022.

[27] Suyal, P., Pant, J., Dwivedi, A., & Lohani, M. C. (2016, September). Performance evaluation of rough set based classification models to intrusion detection system. In 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Fall) (pp. 1-6). IEEE.

[28] Bhatt, C., Kumar, I., Vijayakumar, V., Singh, K. U., & Kumar, A. (2021). The state of the art of deep learning models in medical science and their challenges. Multimedia Systems, 27(4), 599-613.

[29] Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018, December). Convolutional neural network (CNN) for image detection and recognition. In 2018 first international conference on secure cyber computing and communication (ICSCCC) (pp. 278-282). IEEE.

[30] Malakouti, S. M. (2023). Heart disease classification based on ECG using machine learning models. Biomedical Signal Processing and Control, 84, 104796.

[31] Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. Algorithms, 16(2), 88.