# DISCOVERING INSIGHTS INTO HEART HEALTH: A SURVEY OF DATA MINING AND MACHINE LEARNING METHODS

[1] **Hemant Sudarshan, UG Scholar, Department of Computer Science and Technology, Dayananda Sagar University, Bengaluru.**

**hemanthsudarshan2002@gmail.com, 8985837147**

[2]**M.Chithambarathanu, Assistant Professor, Department of Computer Science and Technology, Dayananda Sagar University,Bengaluru.**

**chithambaramthanu@gmail.com,**

**9500420634**

**Abstract: In this modern era, technology is growing at a fast pace in all areas, which helps us to live life more comfortably. This tremendous technological advancement has changed the lifestyle of humans in both positive and negative ways. However, changes in daily routines can lead to many fatal diseases, and one of the most life-threatening diseases is heart disease. Heart disease can be caused due to many aspects such as stress, high blood pressure, pollution, a habit of consuming fast food items, lack of exercise, and others. In the present generation, different types of heart diseases like heart attack, vessel blockage, cardiac arrest, and strokes can be identified irrespective of age. Nowadays, the prediction of heart disease is a big issue in the medical sector, as different types of heart diseases can be diagnosed by conducting various tests. Data Mining Techniques (DMT) can be used to predict heart diseases, which can help doctors identify the disease at an earlier stage and provide appropriate treatment to the patients.**

**Keywords: Heart Disease, Data Mining, Multi-verse Optimizer, Multi- Layer Perceptron.**

## I. Introduction

The healthcare industry produces enormous quantities of information to predict diseases in advance. To address this issue, the medical industry is adopting DMT techniques for extracting useful information to predict heart diseases [1]. These techniques are used to solve complex real-world problems adopted in every sector and extract hidden patterns from vast medical data available in hospitals. Heart disease is a significant issue for the current generation due to the difficulty in identifying it at an early stage [2]. Identifying heart disease requires considering several parameters, and each type of heart disease has different parameters compared to other types [3]. Every method should adopt different attributes for predicting heart disease based on the results to identify the type of disease. DMT techniques should help the medical practitioner to diagnose disease in a fast and accurate way [4]. The cause of heart disease may vary from person to person depending on their regular activities and behavior. These diseases are not specific to a particular age group; rather, they are seen in every age group [5]. To identify the exact type of heart disease, one must first select the different types of attributes that help to predict the disease in an accurate manner. Many techniques used in the existing system have ignored the most important parameters like age, stress, type of work, previous health condition, etc [6]. The existing medical data contains a huge amount of data with enormous attributes, and there is an issue in finding the exact attributes for identifying heart disease. It takes more time to cleanse and categorize the data before predicting heart diseases [7].

## II. Related Work

The prediction of heart-related diseases has become a major issue in the healthcare industry. Numerous methods in DMT have been developed to predict heart diseases at an earlier stage, which helps medical practitioners facilitate more precise and efficient decision-making. Various methods have been compared, and the best method that provides the highest accuracy has been identified [8]. A simple cost-cutting method was proposed for identifying heart diseases and enhancing the Decision Support System (DSS) to make decisions. The authors used the Naive Bayesian (NB) data classification technique, which uses the Bayesian theorem for diagnosing coronary diseases [9].

This technique requires critical Clinical variables like attributes or parameters and post-symptoms to design the model [10]. Data were taken from the UC-Irvine (UCI) data repository, which contains attributes such as sex, age, hypertension, obesity, BMI, height, weight, and test reports. This model is also known as the independent feature model, which is mainly used for supervised learning. The authors proved that the NB model is effective and accurate when used with the Advanced Encryption Standard (AES) algorithm [11]. A new model was proposed in an online community for maintaining the heart disease dataset taken from the Kaggle data repository. Initially, the dataset contained 12 attributes and 800 records, but the model selected 11 out of those attributes [12]. The data processing was done using the Sci-Kit-learn library of Python, which contains mechanisms used to process the algorithms. The Ischemic Heart Disease technique was proposed to detect heart diseases and ischemic strokes, which are generally caused without any prior warnings [13].This model uses five classification techniques, namely Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbour (KNN), NB, and Support Vector Machine (SVM). These classification techniques were compared, and finally, SVM was proven to be the most suitable A comparative analysis was conducted on this technique in contrast to other methods[14]. A risk prediction model was used with structured data using a Convolution Neural Network (CNN) to develop the uni-model disease risk prediction algorithm. This model adopted the dataset from the UCI repository and was designed to predict whether the patient suffers from heart disease or not [15]. The database contained 12 attributes like age, blood sugar, bloo1d pressure, sex, resting result of cardiograph, age, height, weight, etc. The CNN-based Uni-model Disease Risk Prediction (CNN UDRP) method uses a deep learning algorithm to predict heart disease [16]. In this paper, heart disease prediction was done using NB and KNN algorithms. These two algorithms were evaluated, and it was proved that Naive Bayes is better than the KNN algorithm [17].

Machine learning hybrid classifier model algorithms, such as Random Forest (RF) classifier algorithm and K-means, have been proposed for predicting heart problems. In this technique, the dataset is evaluated using machine learning algorithms like NB classifier and J48 tree classifier. The results of these algorithms are compared, and results are generated using the RF classifier. The robustness of this method is tested using the confusion matrix [18]. A prediction model has been developed that is capable of detecting heart diseases based on the factors extracted from the (European RealTime Integrated) ERIC laboratory, which contains 209 test cases. This model compares three classifier algorithms and proves that RF Classifier is the best when compared to the other two models such as NB and K-means. The main idea of this model is to devise a predictive model for the detection of cardiovascular heart diseases using techniques [19]. A Multi-Verse Optimizer (MVO) model based on the MultiLayer Perceptron (MLP) was proposed for the early detection of Coronary Artery Disease (CAD) [20]. This model dataset has 54 features. This MVO is a powerful bio-inspired algorithm used for training MLP and CAD datasets. Many traditional algorithms like NB, Bagging Tree (BT), RF, SVM classification, Regression Trees (RT), J48 tree, and Random tree are compared with the MVO Classifier [21]. To evaluate the performance, two metrics that were used are Accuracy (ACC) and Area under the Curve (AUC). This model is evaluated and proved that the MVO algorithm is more effective when compared with other traditional methods [22]. A study was conducted to compare and investigate the accuracy of different DMT classification schemes and Machine Learning (ML) techniques used for predicting heart disease. For this investigation, the Cleveland dataset consisting of 303 instances was adopted as the main database for testing and training [23]. To increase the volume of data, 10 Fold-Cross Validation was used.

ISBN: 978-81-945891-0-5

Several classifier algorithms like Single Conjunctive Rule Learner (SCRL), DT, SVM, KNN, Radial Basis Function (RBF), and NB were used for analysis and comparison. Bagging, boosting, and stacking were also applied to the given dataset [24]. The results show that SVM using ML techniques provides the highest accuracy compared to all other methods [25].He proposed model in this paper is helpful in predicting heart diseases and other diseases using DMT. It focuses on the existing DM classifier techniques used in medical data analysis [26]. The study mainly concentrated on LR, NB, and Artificial Neural Networks (ANN) for predicting heart disease and concluded that Neural Networks is the best method for disease prediction [27].Different evaluation methods used today for predicting heart disease were collected from the available large set of clinical data. A novel prototype was proposed to extract substantial knowledge from past clinical data with the help of a hybrid genetic algorithm, which would be helpful for doctors to make appropriate decisions at the right time for providing treatment. Earlier studies compared NB, DT, and KNN algorithms and showed that the decision tree is the best algorithm for predicting heart diseases [28-29]. A scalable framework was proposed to predict heart diseases based on certain attributes using large medical data. The main focus of the paper is to predict heart disease using a minimal or small number of attributes [30]. The proposed prediction model uses RF on Apache Spark, which can be used on ever-increasing clinical data and to make appropriate decisions. By using this approach, the accuracy is increased to 98%. This model is compared against the NB classifier approach with RF and proves that RF is better [31].

### III. Methods and Results

In the present study, a comparison of various models for the purpose of prediction of heart diseases like KNN, DT, Nb, LR, Rf, NN, J48, MVO-MLP, and the results shown in Table 1.

| S.NO | Technique | Accuracy |
|---|---|---|
| 1. | Naïve Bayesian with AES | 89.77 % |
| 2. | Support Vector Machine | 97.91 % |
| 3. | Navie Bayes using CNN UDPRP | 82 % |
| 4. | Random Forest | 92 % |
| 5. | MVD-MLP | 92.39 % |
| 6. | Support Vector Machine, MLP | 84.15 % |
| 7. | Neural Networks | 100 % |
| 8. | Decision tree | 52 % |
| 9. | Random Forest using Spark | 98 % |

This paper compared different DMT classifier algorithms used for predicting heart-related diseases. Different models were proposed using classifier techniques for heart disease prediction, and these models used UCI, Kaggle, and Sci-Kit data set repositories for construction mechanisms. Each algorithm was tested with the above datasets, and the results were analyzed with different existing techniques. It is clear to infer that NN and DT are the best algorithms for heart disease prediction.

In the future, this work can be extended by incorporating more advanced machine learning techniques and exploring the potential of hybrid models. Moreover, the proposed models can be tested with more diverse and larger datasets to increase their reliability and generalizability. Finally, the models can be further improved by incorporating feature selection and extraction techniques to enhance their accuracy and reduce their complexity.

ISBN: 978-81-945891-0-5

## V. References


**1.** Repaka, A.N., Ravikanti, S.D. and Franklin, R.G.: 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 (2019).

**2.** Maji, S., and Arora, S.: Decision tree algorithms for prediction of heart disease. In Information and Communication Technology for Competitive Strategies, pp. 447-454, Springer, Singapore (2019).

**3.** Mehta, D. B., and Varnagar, N. C.: Newfangled Approach for Early Detection and Prevention of Ischemic Heart Disease using Data Mining. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI,. pp. 1158-1162 IEEE (2019).

**4.** Kannan, R., and Vasanthi, V.: Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease. In Soft Computing and Medical Bioinformatics, pp. 63-72 Springer, Singapore (2019).

**5.** Hao, Y., Usama, M., Yang, J., Hossain, M. S., and Ghoneim, A.: Recurrent convolutional neural network based multimodal disease risk prediction. Future Generation Computer Systems, 92, 76-83, (2019).

**6.** Ksiazek, W., Abdar, M., Acharya, U. R., and Pławiak, P.: A novel machine learning approach for early detection of hepatocellular carcinoma patients. Cognitive Systems Research, 54, pp.116-127 (2019).

**7.** Jalali, S. M. J., Karimi, M., Khosravi, A., and Nahavandi, S.: An efficient neuroevolution approach for heart disease detection. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC),pp. 3771-3776 IEEE (2019).

**8.** Bindushree, D. C., and Rani, V. U.: A review on using various DM techniques for evaluation of performance and analysis of heart disease prediction. In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), pp. 686-690 IEEE (2017).

**9.** Sowmiya, C., and Sumitra, P..:Analytical study of heart disease diagnosis using classification techniques. In 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), pp. 1-5 IEEE (2017).

**10.** Shanmugasundaram, G., Selvam, V. M., Saravanan, R., and Balaji, S.: An Investigation of Heart Disease Prediction Techniques. In 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA), pp. 1-6 IEEE (2018).

**11.** Raju, C., Philipsy, E., Chacko, S., Suresh, L. P., and Rajan, S. D.: A Survey on Predicting Heart Disease using Data Mining Techniques. In 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), pp. 253-255 IEEE (2018).

**12.** Altayeva, A., Zharas, S., and Im Cho, Y.: Medical decision making diagnosis system integrating k-means and Naïve Bayes algorithms. In 2016 16th International Conference on Control, Automation and Systems (ICCAS), pp. 1087-1092 IEEE (2016).

**13.** Pahwa, K., and Kumar, R.: Prediction of heart disease using hybrid technique for selecting features. In 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), pp. 500-504 IEEE (2017).

**14.** Shaji, S. P.: Predictionand Diagnosis of Heart Disease Patients using Data Mining Technique. In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0848-0852 IEEE (2019).

**15.** Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., and Lamgunde, A.: Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) pp. 1-8 IEEE (2017).

**16.** Babu, S., Vivek, E. M., Famina, K. P., Fida, K., Aswathi, P., Shanid, M., and Hena, M.: Heart disease diagnosis using data mining technique in 2017. International conference of Electronics, Communication and Aerospace Technology (ICECA), 1, pp. 750-753 IEEE (2017).

**17.** Chauhan, A., Jain, A., Sharma, P., and Deep, V.: Heart disease prediction using evolutionary rule learning. In 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT), pp. 1-4 IEEE (2018).

**18.** Sonet, K. M. H., Rahman, M. M., Mazumder, P., Reza, A., and Rahman, R. M.: Analyzing patterns of numerously occurring heart diseases using association rule mining. In 2017 Twelfth International Conference on Digital Information Management (ICDIM), pp. 38-45 IEEE (2017).

**19.** Geetha, M. C. S., Shanthi, I. E., and Sehnaz, N. S.: Analyzing the suitability of relevant classification techniques on medical data set for better prediction. In 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 665-670 IEEE (2017).

**20.** Hasan, S. M. M., Mamun, M. A., Uddin, M. P., and Hossain, M. A.: Comparative Analysis of Classification Approaches for Heart Disease Prediction. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), pp. 1-4 IEEE (2018).

**21.** Mohan, S., Thirumalai, C., and Srivastava, G.:Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554 (2019).

22. Suvarna, C., Sali, A., and Salmani, S.: Efficient heart disease prediction system using optimization technique. In 2017 International Conference on Computing Methodologies and Communication (ICCMC), pp. 374-379 IEEE (2017).

23. Mahboob, T., Irfan, R., and Ghaffar, B.:Evaluating ensemble prediction of coronary heart disease using receiver operating characteristics. In 2017 Internet Technologies and Applications (ITA), pp. 110-115 IEEE (2017).

24. Rathnayakc, B. S. S., and Ganegoda, G. U.: Heart diseases prediction with data mining and neural network techniques. In 2018 3rd International Conference for Convergence in Technology (I2CT),pp. 1-6 IEEE (2018). 25.Aldallal, A., and Al-Moosa, A. A. A.: Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP), pp. 150-154 IEEE (2018).

26. Mane, T. U.: Smart heart disease prediction system using Improved K-means and ID3 on big data. In 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), pp. 239-245 IEEE (2017).

ISBN: 978-81-945891-0-5