International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 1640-1648
https://doi.org/10.70917/ijcisim-2026-4725
Article

Predictive Analytics in Cardiology: Using Machine Learning to Identify Heart Failure Early

M. Madhuri[1,2], A. Suresh Babu[3]

[1]Department of Computer science and Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu, Andhra Pradesh, India,
[2]Department of Computer science and Engineering-Allied, MVSR Engineering College, Hyderabad, Telangana,
[3]Professor, Department of Computer science and Engineering, Jawaharlal Nehru Technological University Anantapur, Ananthapuramu, Andhra Pradesh, India.
Corresponding Author: mmadhuri_cse@mvsrec.edu.in,
madhuri.chetty@gmail.com, sureshalladi.cse@jntua.ac.in

Abstract: The creation of an algorithm using machine learning and natural language processing methods is the main goal of this study. Predicting cardiac disease is a tough problem in medical data analysis, given that it is now one of the leading causes of mortality globally. The varity of types such as health records, electronic health records, network monitoring (body), and patients diagnosing data conditions by projecting medical sensors and wearable technology onto the human body are used. Machine learning (ML) has demonstrated its value in aiding decision-making and predicting outcomes from the extensive datasets provided by the healthcare industry. The study presents the innovative MLP-EBMDA (Multi-Layer Perceptron for Enhanced Brownian Motion-based Dragonfly Algorithm) for heart disease prediction and uses an efficient unsupervised method for feature selection. The process begins with obtaining input from the dataset, followed by preprocessing and the proposed feature selection technique, which adeptly chooses relevant features. Subsequently, The MLP-EBMDA is a novel technique is used to classify cardiac disease, making early-stage heart disease prediction easier. The suggested method successfully classifies cardiac disease as normal or abnormal with an excellent impressive accuracy rate of 96.4%.
Keywords: NA

1. Introduction

According to WHO evaluations, cardiac failures are more day by day and it has become a crucial challenges in health systems. However, forecasting its development remains a significant difficulty because to the multiple risk factors involved, such as high cholesterol and diabetes, high blood pressure, atypical pulse rate, and numerous other conditions. By using vast amounts of data produced by the healthcare sector to inform choices, ML (Machine Learning) approaches have proven effective in the prediction of cardiac disease. The approach used for predicting heart disease is displayed using several feature combinations and many conventional classification methods. Additionally, the proposed prediction model of heart disease that controls the training data distribution and finds and removes outliers, and predicts heart disease using XGBoost, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), as well SMOTE-ENN (Synthetic Minority Over-Sampling Technique-Edited Nearest Neighbor). A comparative study between the suggested and current systems has been done. In terms of accuracy, it was discovered that the suggested model outperformed the conventional models. It is still necessary to compare a number of data samples with the model's parameters and the large medical datasets.

Furthermore, this study presented a novel approach that uses machine learning techniques to identify significant traits, improving the prediction accuracy of cardiovascular illness. As a result, the suggested model performed accurately and effectively. In the near future, a range of machine learning approaches may be employed to improve disease prediction. The characteristics for the prediction or expectation of cardiac disease have been chosen using an optimized unsupervised feature selection approach. Additionally, using a variety of currently available methodologies, all supervised and unsupervised methods have been investigated for heart disease prediction .

A few machine learning and data mining techniques, such as decision trees, logistic regression, fuzzy logic, ANNs (Artificial Neural Networks), KNNs (K-Nearest Neighbors), and SVMs (Support Vector Machines), have been used to predict the cardiac disease. It provides a thorough summary of the current techniques as well as intuitive thoughts pertaining to the systems in place. Based on the dragonfly method, multilayer perceptrons integrated with improved Brownian motion have been used for classification. To achieve the effectiveness of forecasting heart disease, its performance is finally examined . This work uses machine learning techniques to provide an efficient strategy for predicting cardiac disease. Heart disease prediction is frequently predicated on the intricate relationship between pathological and clinical data. Because of this challenge, academics and medical professionals are becoming increasingly interested in developing precise and efficient methods for predicting cardiac disease. This aids in the early diagnosis of heart disease and predicts the illness's course. Therefore, the accuracy rate of cardiac disease predicting as normal one is improved by employing efficient choice of features and categorization algorithms based on the recommended approaches.

Machine learning involves three steps: the feature selection process, data pre-processing, and classification. The data preprocessing is the process of turning comprehensible data into something that can be understood. When using unsupervised learning for feature selection, the smallest set of features that best defines the cluster, based on the given criteria, is identified. The artificial neural network (ANN) is a feed-forward kind of process, the MLP allows for the categorization of objects into multiple labels or likelihood-based prediction. One optimization method that makes advantage of meta-heuristics is the dragonfly algorithm. This study uses Brownian motion since it's an isotropic method that doesn't depend on direction. This makes it likely to find food sources inside the lower portion of the population. The dragon fly method's randomization stage is improved and its discovery capabilities are strengthened by the application of the enhanced Brownian algorithm. The algorithm performs better and can withstand interruptions with the aid of this hybridization technique. The proposed approach combines MLP and EBMDA to improve efficacy and predict heart illness, improved unsupervised technique to increase the accuracy of heart disease prognosis, along with MLP as well as EBMDA hybridization.

Heart Disease

One of the body's components that is essential to every other element, including the brain, is the heart, which pumps or circulates blood to every area of the body. Death may result if the heart stops pumping blood to the brain because separate neuronal systems in the brain die, which stops all other body parts' nerves or tissue from functioning. so that in the end, the heart determines life. People can live healthy lives when their hearts are operating correctly. It is difficult and more involved to predict cardiovascular disease than to obtain an automatic diagnosis of illness. Due to the fact that healthcare facilities retain vast amounts of complicated, difficult-to-analyse data.

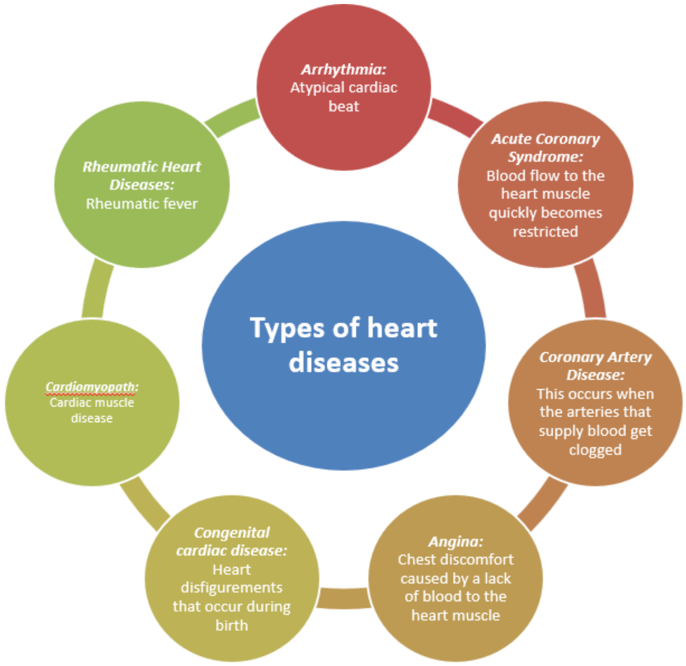

Fig.1: Types of heart disease .

Several studies have been undertaken to study about the health prediction systems by using ML algorithms and data mining techniques. by numerous authors in order to attain a precise automatic diagnosis of diseases as well as improve decision-making in medical centers.

In order to predict cardiac disease , this work aimed to develop two distinct Adaptive Neuro Fuzzy and Inference System based classification models. It has been demonstrated that the classifiers acquired the abilities required for dataset classification. Testing, accuracy, and training results have been evaluated to see how well the suggested approach performs for categorization. Therefore, the suggested model may be further enhanced by using many derivatives in the Jacobian matrix calculation to speed up the convergence of the result.

The effectiveness of many classifiers, such as KStar, J48, SMO, Bayes Net, and Multilayer Perceptrons, for predicting heart illness was assessed in a comparative research by using the WEKA workbench. With accuracy rates of 89% and 87%, respectively, the Sequential Minimal Optimization (SMO) and Bayes Net methods proved to be the best models during k-fold cross-validation. The authors pointed out that present accuracy levels are still below ideal for clinical deployment, despite their relative effectiveness in comparison to the other studied designs. such that if accuracy performance is further enhanced, a better diagnostic conclusion may be made.

An increased chance of rescuing a specific patient can result from early identification of cardiac disease. The Dragon Fly algorithm for specific attributes is being introduced in this work with an emphasis on fitness. The characteristics of a certain feature that has been chosen are detected using deep learning. Compared to particle swarm optimization, this algorithm's accuracy rate was 6.7% higher, and it outperformed Gray wolf optimization by 6.96% .

This study The primary focus of this study is the use of higher-order spectra to describe coronary artery disease using ECG data. The essay uses Decision Trees and K-Nearest Neighbors. These algorithms yielded accuracy results of 98.17% and 98.99%, respectively. The algorithms employed in the study produce a superior accuracy performance outcome for characterizing coronary artery disease.

The author in this study explained about unstructured patient data of the heart failure detection system is important of this work. This research works makes use of Naïve Bayes, SVM, Random Forest Decision Tree, Neural Network, and Logistic Regression. For every method, the corresponding accuracy values are 80%, 84.8%, 83.8%, 86.6%, 86.6%, and 87.7%. Of all the algorithms, Naïve Bayes has the higher efficiency rate of prediction.

2. Proposed Work

The cardiac failures dataset is used as input in the proposed work. Next, pre-processing is carried out, and then the optimal unsupervised approach is used to choose features. Human heart disease was classified using the MLP-EBMDA based on the characteristics that were chosen. Lastly, an analysis of the suggested methodology's performance in terms of several factors has been conducted.

a. MLP- EBMDA approach.

The MLP-EBMDA was used to forecast human heart disease depending on the features that were chosen.

MLPs may identify an unknown pattern using other known patterns that have similar distinctive characteristics, a process called as generalization. This implies that because noisy or partial inputs resemble pure and full inputs, they will be categorized. The Multi-Layer Perceptron (MLP) can identify nonlinear functions since it comprises one or more hidden layers along with to an input and an output layer.

The algorithm, Dragon fly is a swarm intelligence-based metaheuristic approach that draws inspiration from the natural static and dynamic behaviors of dragonflies. The discovery and exploitation phases are the two primary phases of optimization. Dragonflies were used to simulate these two stages as they either statically or dynamically searched for food or avoided the adversary.

The isotropic approach of the Brownian motion method—which is direction-independent—increases the capacity for discovery. However, the time-based random form and restricted size of the steps prevent the search space from growing, guaranteeing motion continuity. For this reason, the Dragon Fly algorithm is being used.

After doing study, we came to the conclusion that our algorithm can operate effectively and can produce superior outcomes when compared to other algorithms.

b) The Dragon Fly algorithm

In order to get the best outcome, dragonflies often operate according to the norm of separating from their current neighbor, organizing toward wholesome food, veering away from the adversary, being coherent toward their neighbor, and being desirous of the food. In order to achieve the global optima in the dragonfly algorithm (DA), exploitation and exploration have been accomplished through the synthesis of the static as well as dynamic behaviors of the dragonfly. This is the explanation of the DA mathematical model:

The dragonfly of size N's population matrix is given by

$$Y_j^E = (y_1, y_2, \cdots y_N$$

J is 1, 2, 3, in this case. The jth dragonflies' locations within the E search field are shown by N, YE j, and N is the number of search agents. It is possible to calculate many components for each N neighbor person, including cohesion, alignment, and segregation. It calculates the segregation parameter.

$$K_j = -\sum_{i=1}^{\bar{N}} Y - Y_i$$

The following equation represents each member's cohesiveness inside the swarms individually.

$$B_j = \frac{\sum_{i=1}{}^{N} Y_i}{\bar{N}} - Y$$

The following is how the alignment is shown in swarms:

$$X_j = \frac{\sum_{i=1}^{\bar{N}} Q_i}{\bar{N}}$$

Here, the labels "j" and "i" refer to both the current individual and the neighboring individual. Qi represents the speed of the person who is next to you. From Eqs.above the attractiveness toward the food supply Hj and the divergence from foes are calculated

$$H_j = Y^{+} - Y$$

$$I_j = Y^{-} + Y$$

Here, the positions of the opponent and food are shown by Y− and Y+. Y is the dragonfly's location as of right now. Equation above provides the equation for the location of the dragonfly with respect to the step vector for a population where there is at least one neighbor dragonfly.

$$\Delta Y_{a+1} = \left(kK_j + xX_j + bB_j + hH_j + iI_j\right) + w\Delta Y_a$$

The factors in this case are cohesiveness, food desire factor, adversary deviation factor, alignment, segregation, and x, k, h, i and b stand for the weight of inertia, and a for the number of iterations. The location of the dragonfly is provided by equation

$$Y_{a+1} = Y_a + \Delta Y_{a+1}$$

The levy flying function, which is provided by equation, is used to update the dragonfly's location.

$$Y_{a+1} = Y_a + \mathrm{lev}(c)Y_a$$

As a result, the dragonfly's fitness is determined using a streamlined position vector and is maintained until a certain need is satisfied.

MLPs are a type of feed-forward neural network (FNN) in which the neurons that process information are arranged in many parallel layers and the input is sent via NNs in a single direction. The output layer is the final layer, and the input layer is the first of numerous parallel layers. the layers in between are referred to as hidden layers, an FNN is referred to as an MLP when it just has one hidden layer.

3. Results and Discussions

This section discusses the outcomes of the suggested method for classifying cardiac disease as normal or abnormal. Additionally, a comparison analysis is carried out.

1) Data description

In this investigation, the heart disease dataset is considered. One section and one subsection make up the dataset. For training and model building, the first section is used. To find out how accurate the model is, the second section is used for testing. UCI Cleveland Information Set. 100 data and 8 characteristics make up the dataset. The collection is labeled.

2) Performance metrics

Among the performance metrics utilized to evaluate the proposed system are accuracy, precision, recall, and F1-score.

a) Accuracy

The following equation can be used to express accuracy, which is the prediction ratio that was successfully attained for the model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

In this situation, TP stands for True Positive, which is abnormal and exactly predicted, & TN for True Negative, which is expected and normal. False Negatives (FN) are predicted wrongly but are abnormal. False Positives (FP) are predicted incorrectly yet are normal.

b) Precision

It can be expressed as the percentage of the tuple, which is found using the following formula, that precisely predicts the abnormal patient has heart disease.

$$\text{Precision} = \frac{TP}{TP+FP}$$

c) Recall

A person's likelihood of having cardiac illness is determined by the ratio of tuples that correctly predicts abnormal persons. It is provided by:

$$\text{Recall} = \frac{TP}{TP+FN}$$

d) F1-score

The representation of harmonic mean of precision and recall is F1-score. This formula is used to calculate it.

$$F1 - \text{Score} = \frac{2^{\star}\ \text{precision}\ ^{\star}\ \text{recall}}{\text{precision} + \text{recall}}$$

Three traditional studies are compared with the suggested approach in order to evaluate the effectiveness of the newly implemented system. As a result, an analysis of the study [25] has been considered. In this case, previous research several machine learning models, such as Random Tree, Radial Basis Function Network (RBF Network), Random Forest, Naïve Bayes Tree (NB Tree), J48, DBMIR–TLBO–ANN, and TLBO–ANN, were taken into consideration for a comparative analysis.

Table: 1 A comparison between the suggested and conventional research works [25].

Classifierss	**Sensitivityy**	**Specificityy**	**Accuracyy**	**F- scoree**	**Kappaa**
OANN (Dbmritlbo-Ann)	95.43	94.33	93.51	91.84	89.86
TLBO-ANN	96.21	81.71	88.65	92.81	83.69
J48	76.94	77.03	77.55	81.59	66.64
Random tree	84.41	81.11	82.94	85.63	65.38
RBF Networkk	75.19	71.56	73.49	74.66	76.81
NB Treee	70.24	83.94	83.21	77.97	61.06
Random forestt	79.98	79.44	81.18	81.93	69.95
Proposed Method (MLP EBMDA)	96.93	96.48	96.31	96.46	96.76

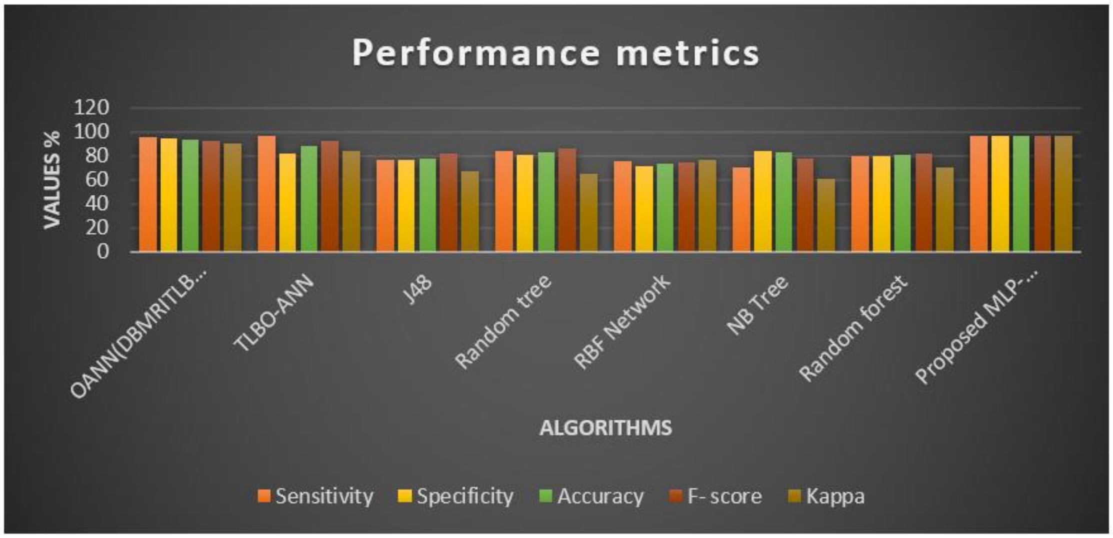

Fig.2 Comparison of the recommended and traditional studies

Table 1 displays the comparative analysis findings. The classic OANN (DBMRITLBO-ANN) demonstrated a sensitivity of 95.43%, whereas TLBO-ANN demonstrated a sensitivity of 96.212%, J48 shown a sensitivity of 76.94%, and Random Tree demonstrated a sensitivity of 84.41%. Comparably, the suggested system had a greater sensitivity rate of 96.92% than the RBF Network, NB Tree, and RF, which all displayed sensitivity rates of 75.19%, 70.24%, and 79.98%, respectively. However, the key indicator of a method's effectiveness is its accuracy. Comparing the suggested system to the usual research, it demonstrates a higher accuracy rate of 96.63%. This is illustrated visually in Fig. 2 and demonstrates the effectiveness of the suggested strategy.

A strategy for predicting heart disease based on test results is shown in figure 1. The Fbeta-measure assesses precision, accuracy, recall, and F1-score using the beta parameter selection. Since it is more important to correctly classify as many positive samples as possible than to maximize the number of correct classifications, the F2-score is more suitable for some applications. We are concentrating on F1 Score since we need to improve prediction rate in order to accurately classify as many positive samples as possible.

For the 84 test results, the TP and TN were 56 & 6, respectively. Similarly, F1 and F2 provide values of 10 and 12, respectively. There is an accuracy of 0.81. By using 66 testing data, the suggested model yields result similar to the following: precision of 87.98, F1-score of 87.56, recall of 87.87, accuracy of 84.35, and TP and FP of 44 and 11. Similarly, using 45 testing data, the suggested model yields result for TP and FN of 33 and 5, TN and FN of 2 and 5, precession (96.95) recall (96.54), accuracy (96.47), and F1-score (96.68). In a similar manner, Table 1 analyzes and tabulates the remaining data.

Table: 2 Performance analysis of proposed studies

Testing Sequence	**Sensitivity**	**Precision**	**Accuracy**	**F1-score**	**Recall**
Test 1	82.22	82.67	81.57	84.67	85.35
Test 2	80.91	80.49	83.89	83.78	86.65
Test 3	87.23	87.98	84.35	87.56	87.87
Test 4	**96.95**	**96.45**	**96.47**	**96.68**	**96.54**
Test 5	85.43	85.76	79.89	84.79	82.35

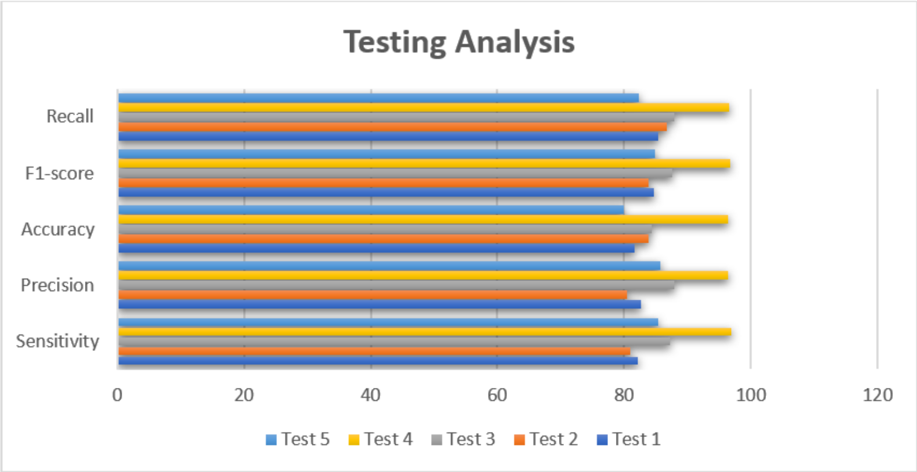

Fig.3 Performance Analysis between the testing experiments

The accuracy rate of the suggested model was 96.47% for 100 testing data, 81.57% for 82 percent of the data, 83.89% for 45 data, 84.35% for 62 data, and 79.89% for 85 percent of the data. E3 experimental data can yield highly accurate results. Fig 3 displays the variability in exact outcomes for every trial.

In the T4 trial, the suggested MLP-EBMDA system outperforms traditional methods in heart disease prediction. On 84 testing samples, the model's accuracy was 96.54%; when trained on 83%, 87%, 85%, & 84% of the dataset, its accuracy was 85.35%, 86.65%, 87.87%, as well as 82.35%, respectively. High accuracy results may be achieved using T4 experimental data.

The Precision rate of the suggested model was 96.45% for 100 testing data, 82.67% for 82 percent of the data, 80.49% for 45 data, 87.98% for 62 data, and 85.76% for 85 percent of the data. T4 Experimental data can yield highly accurate results.

The suggested model's accuracy rate sensitivity was 96.95% for 100 testing data, 82.22% for 66 data, 80.91% for 45 data, 87.23% for 62 data, and 85.43% for 82 percent sensitivity. The differences in recollection outcomes for every experiment.

All of the experimental setups had different F1-scores. Using 84 testing samples, the suggested model earned an F1-score of 96.68%, whereas scores of 84.67%, 83.78%, 87.56%, and 84.79% were obtained when 83%, 87%, 85%, and 84% of the data were utilized for testing, respectively.

4. CONCLUSION

This study assesses machine learning's potential for health data mining. Early prognosis is essential since heart disease is a major worldwide concern and can have catastrophic consequences if treatment is delayed. This importance is underscored by reports on death rates from the World Health Organization. The outcome of this work is on using machine learning to detect heart disease early on. Specifically, it offers a Multilayer Perceptron connected with Enhanced Brownian Motion via the Dragonfly algorithm (MLP-EBMDA). This hybridization methodology is used in the classification process, and an optimal unsupervised method is used to pick features to improve heart disease prediction. The suggested system is evaluated with key characteristics such as precision, F1-score, recall, and accuracy rate. Based on these criteria and the chosen testing results, a comparative analysis with current approaches helps assess the effectiveness of heart disease prediction. The accuracy rate of the suggested system is 96.47% approx., Sensitivity rate 96.95% approx., Precision rate 96.45% approx., Recall rate 96.54% approx., F1-score rate 96.68 approx., in distinguishing between normal and abnormal conditions.

REFERENCES

1. N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System,” IEEE Access, vol. 8, pp. 133034–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.
2. A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, “A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms,” Mobile Information Systems, vol. 2018, pp. 1–21, Dec. 2018, doi: 10.1155/2018/3860146.
3. M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, “Identification of Heart Failure by Using Unstructured Data of Cardiac Patients,” in 2016 45th International Conference on Parallel Processing Workshops (ICPPW), IEEE, Aug. 2016, pp. 426–431. doi: 10.1109/ICPPW.2016.66.
4. A. K. Dwivedi, “Analysis of computational intelligence techniques for diabetes mellitus prediction,” Neural Comput Appl, vol. 30, no. 12, pp. 3837–3845, Dec. 2018, doi: 10.1007/s00521-017-2969-9.
5. M. Sultana, A. Haider, and M. S. Uddin, “Analysis of data mining techniques for heart disease prediction,” in 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), IEEE, Sep. 2016, pp. 1–5. doi: 10.1109/CEEICT.2016.7873142.
6. C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques,” Algorithms, vol. 16, no. 2, p. 88, Feb. 2023, doi: 10.3390/a16020088.
7. V. Kirubha and S. M. Priya, “Survey on Data Mining Algorithms in Disease Prediction,” International Journal of Computer Trends and Technology, vol. 38, no. 3, pp. 124–128, Aug. 2016, doi: 10.14445/22312803/IJCTT-V38P122.
8. E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, “Using recurrent neural network models for early detection of heart failure onset,” Journal of the American Medical Informatics Association, vol. 24, no. 2, pp. 361–370, Mar. 2017, doi: 10.1093/jamia/ocw112.
9. K. Srinivas and G. Raghavendra Rao, “Survey on Prediction of Heart Morbidity Using Data Mining Techniques,” International Journal of Data Mining & Knowledge Management Process, vol. 1, no. 3, pp. 14–34, May 2011, doi: 10.5121/ijdkp.2011.1302.
10. “Heart Disease Prediction with Data Mining Clustering Algorithms,” International Journal of Computing, Communication and Instrumentation Engineering, vol. 4, no. 1, Jan. 2017, doi: 10.15242/IJCCIE.DIR1116009.
11. Kumar, C. A., Vijay, K., Shiva Chandra, V. et al., “Experimental–AI approach (hybrid) for evaluating the strength characteristics of eco-efficient concrete incorporating coal bottom ash,” Sci Rep, vol. 16, p. 25227, 2026, doi: 10.1038/s41598-026-55732-1.
12. S. Mohan, C. Thirumalai, and G. Srivastava, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,” IEEE Access, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
13. A. Javeed, S. U. Khan, L. Ali, S. Ali, Y. Imrana, and A. Rahman, “Machine Learning-Based Automated Diagnostic Systems Developed for Heart Failure Prediction Using Different Types of Data Modalities: A Systematic Review and Future Directions,” Comput Math Methods Med, vol. 2022, pp. 1–30, Feb. 2022, doi: 10.1155/2022/9288452.
14. M. Nilashi et al., “Coronary Heart Disease Diagnosis Through Self-Organizing Map and Fuzzy Support Vector Machine with Incremental Updates,” International Journal of Fuzzy Systems, vol. 22, no. 4, pp. 1376–1388, Jun. 2020, doi: 10.1007/s40815-020-00828-7.
15. A. Al Ahdal et al., “Monitoring Cardiovascular Problems in Heart Patients Using Machine Learning,” J Healthc Eng, vol. 2023, pp. 1–15, Feb. 2023, doi: 10.1155/2023/9738123.