Journal of Namibian Studies, 35 S1 (2023): 1412-1427 ISSN: 2197-5523 (online)

DNFH: Data Normalization And Feature Selection Using Hybrid Methods For Heart Disease Prediction

[1]Dr. G.T.Prabavathi , [2]M.Sindhu*

[1]Associate Professor of Computer Science
Gobi Arts & Science College
Gobichettipalayam,
Erode Dt., Tamilnadu
gtpraba@gmail.com
*Corresponding Author
[2]Research Scholar Department of Computer Science Gobi Arts & Science College
Gobichettipalayam. sindhuitsri@gmail.com

Abstract

Death from heart disease has consistently ranked high among the leading causes of mortality in the globe. Predicting the likelihood of getting heart disease based on certain features is crucial because of the high cost of identifying heart disease. Feature extraction is an important step in any classification system as it reduces dimensionality and enhances accuracy. This paper proposes a technique called Data Normalization and Feature Selection Using Hybrid Methods (DNFH) for predicting the likelihood of getting heart disease based on selected features. This technique uses Weighted Transform K-Means Clustering (WTKMC) for data normalization; Elastic Net (EN), Random Forest Classifier (RFC), Binary BAT (BBA), and Weighted Binary BAT Algorithm (WBBAT) for ensemble feature selection. Experiments were conducted using the Cleveland dataset and classification accuracy was compared with Support Vector Machine (SVM) and Logistic Regression (LR) models. The results showed significant improvement in accuracy using WTKMC for preprocessing and ensemble feature selection techniques.

Keywords: Data Normalization, Hybrid Feature Selection, WTKMC, WBBAT, BBA, EN, RFC.

1. INTRODUCTION

Heart attack, also known as a myocardial infarction (MI), occurs when an artery in the heart either ruptures or becomes blocked [12]. Modifiable risk factors include smoking, hypertension, high cholesterol, obesity, an unhealthy diet, diabetes, depression, and stress, while non-modifiable risk factors include age, gender, genetic variables, race, and ethnicity. One of the most important and difficult medical topics is the automated prediction of heart illness [10]. According to the World Health Organization, around 18 million people die each year as a result of cardiovascular disease. Because stroke, heart attack, and high blood pressure are among the major causes of death in the United States [34], the government spends $1 billion every day on heart disease medications.

Data on cardiovascular disease are massive. With the aid of preprocessing, the most important risk variables for a heart disease diagnosis may be isolated from the massive information. In Machine Learning [ML], feature selection is a crucial step in the data preparation pipeline used to find the most discriminating data and lower the dataset's dimensionality. It is often crucial in machine learning tasks [5-10] to find features that are simple to extract, invariant to geometric and affine transformations, noise resistant, and good in identifying patterns from a wide range of categories. The particulars of every given scenario should be carefully considered when deciding which characteristics to emphasize [11].

Clustering unlabeled data into equal-variance subsets is a common task in unsupervised machine learning; K-means is one such technique that can do this. It is a popular approach owing to its ease of use and performance on large datasets. This paper aims to highlight the use of the Weighted K-means Clustering (WTKMC) approach for preprocessing the heart disease data set. The L1 and L2 penalties from the lasso and ridge processes for fitting linear and logistic regression models, respectively, are linearly integrated in the elastic net regularized regression method. RF is a popular ML technique because it can be used to both classification and regression issues by combining the results of many decision trees. Based on the bat's echolocation behavior, BAT is a bio-inspired algorithm that uses variable pulse rates of emission and loudness to achieve global optimization

[22]. For a binary version of the bat algorithm (BA), see the Binary BAT (BBA) algorithm [19]. BBA has been shown to outperform other binary heuristic approaches. While the method's velocity update procedures are consistent with BA, this approach encounters premature convergence problems in certain cases.

The fundamental goal of this study is to improve the accuracy level of heart disease prediction by using the WTKMC method for data preprocessing and employing RFC, EN, BBA, and the suggested WBBAT Algorithms for Ensemble Feature Selection. Here is how the rest of the paper is structured. In Section 2, we'll look at some of the research that's been done on the history of hybrid data normalization and feature selection algorithms from a number of different authors. In Section 3, the hybrid architecture is shown, and in Section 4, the research outcomes are summarized. Conclusions and suggestions for further research are presented in Section 5.

2. BACKGROUND STUDY

Prediction of heart disease, a main cause of death globally, is the issue studied here. It is crucial to choose the most relevant variables that might forecast the risk of developing heart disease. In literature, various researchers [20][28][29] have applied a variety of feature selection techniques to enhance the predictable rate of heart disease. Few feature selection algorithms are given in this Section.

In [6], Ganjei and Boostani et al. proposed hybrid feature selection schemes that are hybrid, wrapper, and filter with a good performance by comparing various feature rating and selection criteria methods. They tested it against modern procedures of Sequential Forward Selection (SFS) in the field and selected the best components for the hybrid approach. Jain and Singh et al. [8] introduced a combination of ReliefF Feature Ranking and Principal Component Analysis (ReliefF-PCA) as a method for selecting features to enhance diabetes and breast cancer diagnosis. Mohan et al. [10] aimed to improve early detection and death rates by applying machine learning algorithms to propose a unique categorization of cardiac disease using Hybrid Random Forest with a Linear Model (HRFLM). Nisar and Tariq et al. [12] introduced a novel approach called Hybrid Based Feature Selection (HBFS) to identify important features

and compared it with multiple feature selection methods on various datasets using 10-fold cross-validation.

Saha et al. [15] proposed a correlation-sequential forward feature-based hybrid feature selection technique to select relevant features for accurate prediction using KNN, Decision Tree (DT) and RF algorithms. Shahid et al. [17] used a hybrid Particle Swarm Optimization based Extreme Learning Machine (PSO-ELM) approach with feature selection to classify coronary artery disease with high accuracy. Tania and Shill [20] used a hybrid kernel function feature selection strategy to improve Support Vector Machine (SVM) classification performance by selecting a small number of representative features. Prabavathi et al. [13] investigated many machine learning techniques for predicting cardiovascular disease.

The proposed DNFH technique aims to normalize the dataset using WTKMC algorithm and select the most relevant features using hybrid methods RFC, EN, BBA and WBBAT. The research aims to identify which features are crucial in predicting heart disease using data from the Cleveland dataset. The goal is to achieve a significant improvement in feature selection accuracy to improve the categorization model and prediction of heart disease.

3. Data Preprocessing and Feature Selection

This chapter outlines the research objectives and concludes with a summary of the various methods used and their importance for predicting heart disease. The overall flow is represented in Figure 1.

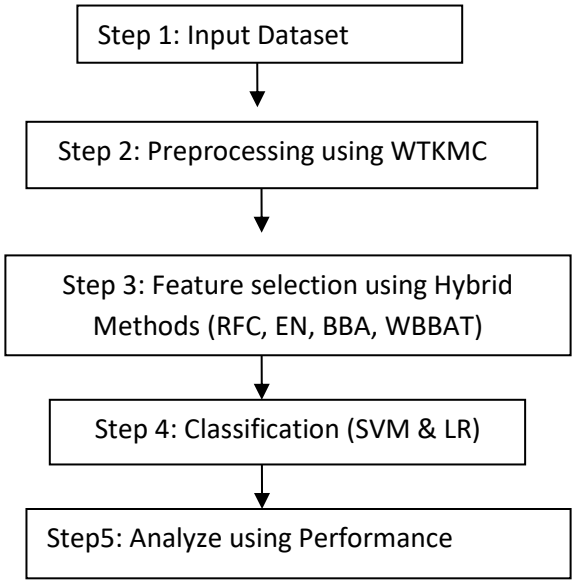

Figure 1: DNFH Process flow

Data is collected from https://www.kaggle.com/datasets/johnsmith88/heart-disease-datasetcontaining 1025 records of patients aged 40-79 years who underwent a medical examination. Patient data comprises demographic information like age and gender as well as clinical data like systolic and diastolic blood pressure, cholesterol, and glucose levels. It also includes a binary target variable indicating the presence or absence of cardiovascular disease in each patient and it is commonly used for predictive modeling tasks.

Preprocessing is the process of preparing raw data for analysis or modeling. It involves various steps like cleaning the data by handling missing values, outliers, or incorrect data. Preprocessing transforms the data by scaling, normalizing, and standardizing to ensure that all the features are on the same scale. Next section highlights the proposed WTKMC algorithm for preprocessing the data which applies the concept of the Transform K-Means algorithm.

3.1. Weighted Transform K-Means Clustering

Clustering can be used as a preprocessing step before applying the machine learning model. It aids in the discovery of dataset subsets that may be utilized to fine-tune the classification model's performance. Using clustering for preprocessing can

extract certain attributes that are more common among the dataset which can be used to build a more accurate predictive model. To ensure that each dataset only belongs to one group with comparable qualities, K-Means, an unsupervised technique, separates the unlabeled dataset into k clusters. The method computes new cluster centroids based on newly assigned data points, and repeats this process until all data points have been assigned to clusters. K-means is useful for many different types of data analysis.

A framework that learns a basis (transform) to operate (analyze) on data and provide the relevant coefficients is called a transform learning framework. [14]. this algorithm is more suitable for solving denoising and reconstruction problems. Anurag G et.al, [3] implemented a transform-learning-based clustering system. The transform learning framework incorporates the k-means clustering loss. Using transform learning for transforming the input data before feeding it into machine learning model greatly enhances the result accuracy. The proposed WTKMC is a clustering algorithm that incorporates the concept of Transform Learning to enhance the traditional K-Means Clustering algorithm. The algorithm aims to find optimal cluster centroids and assign data points to clusters based on their similarity in each feature space.

Assuming n data points is represented as

$$x_i = \left(F_{(i,1)}, F_{(i,2)}, \dots, F_{(i,m)}\right), \quad 1 \le i \le m, \text{ ----- (1)}$$

where each data item has a feature vector in feature space F_i, with i ranging from 1 to m. The goal is to divide the data set into manageable chunks $\{x_i\}m$, where i=1 into k disjoint clusters $\{\pi_u\}k$, u=1. Given a partitioning $\{\pi_u\}k$, u=1, the generalized centroid for each partition u may be expressed as

$$C_u = \left(C_{(u,1)}, C_{(u,2)}, \dots, C_{(u,m)}\right), \text{ ------ (2)}$$

where, for $1 \le l \le m$, the l-th component $C_{(u,l)}$ is in Fl.

C_u is defined as follows:

$$C_u = \operatorname{argmin}\left(\sum_{x \in \pi_u} D^{\alpha}\left(\tilde{x}_X\right)\right) \text{ ------ (3)}$$

The generalized centroid is the point in the distance metric (D) that is empirically averaged to be the nearest to each data item in the cluster u. So, it is essential to find optimal solutions to the

m-problem set below, each of which is a convex optimization problem:

$$C_{(u,l)} = \operatorname{argmin}\left(\sum_{x\in\pi_u} D_l(F_l, F_l)\right),$$

$$1 \le l \le m \text{ ------- (4)}$$

There exists a closed-form solution for the two relevant feature spaces.

Euclidean case:

$$C_{(u,l)} = \frac{1}{\sum_{x\in\pi_u} 1}\sum_{x\in\pi_u} F_l,$$

$\text{where} X = (F_1, F_2, \ldots, F_m)$. ----- (5)

To create a weighted distortion measure D between x and x, assume that there exist m valid distortion m measures $D_l \sum_{l=1}^{m}$. between the corresponding m component feature vectors of x and x (some of which may be produced from the same base distortion).

$D^{\alpha}(x, x) = \sum_{l=1}^{m} \alpha_l D_l(F_l, F_l)$, ------- (6)

where l = (1, 2,.., m) and the feature weights $\sum_{l=1}^{m}$ is positive and add up to 1. For a given x, the weighted distortion D is convex in x because it is a convex combination of distortion measurements that are themselves convex.

3.2 Feature Selection

Overfitting and a loss of model accuracy may be avoided by selecting characteristics that are both relevant and unique. The model may be made more understandable and interpretable if the most crucial elements contributing to the prediction of the target variable are isolated. In this work, after preprocessing the dataset, feature selection algorithms EN, RFC, BBA and WBBAT were applied.

Random Forest Classifier

Random Forest is a classifier that uses an average of many decision trees trained on independent subsets of a dataset to increase predicted accuracy [27]. Using a random selection of features and samples from the training data, RFC generates several decision trees. The final prediction is reached by averaging the results of many decision trees, each of which has been trained on a unique subset of the data. When compared with other classifier algorithms [31][32][33], RFC takes less

training for handling large datasets bring the predictive outcome with high accuracy.

Elastic net

In linear regression models, the elastic net is a regulated regression approach because it combines the lasso and ridge model penalties. The key benefit of the EN is that it retains the efficiency of the ridge penalty and the quality of the feature selection imposed by the lasso penalty [4][28]. In machine learning, Elastic Net is often used for feature selection, regression, and classification tasks, particularly when the number of features exceeds the number of samples. As a regularization approach for high-dimensional data, it has found useful applications in areas as diverse as finance, healthcare, and bioinformatics.

Binary BAT

The Bat Algorithm (BA), or the BAT algorithm, is a metaheuristic optimization technique for solving hard optimization problems. It's based on how bats use echolocation to find food and navigate around obstacles by making ultrasonic noises and listening to the echoes. Function optimization, feature selection, clustering, and image processing are just few of the optimization challenges where it has been put to use. This algorithm is known for its simplicity, efficiency, and ability to escape from local optima. Optimization tasks like dimensionality reduction and feature selection need binary search spaces because of their discrete nature. A binary representation of the answer is necessary for these issues. SeyedaliMirjalili et al. [19] proposed BBA which is a binary version of BAT that can significantly outperform few algorithms on majority of the benchmark functions. By flipping their coordinates from ''0'' to ''1'' and vice versa, the BBA's synthetic bats are able to successfully navigate and hunt in binary search areas.

Weighted Binary BAT

The proposed Weighted Binary Bat (WBBAT) algorithm is an extension of the BBA algorithm that addresses the issues of slow convergence and limited exploration in feature selection. Weights are used in the WBBAT algorithm to refine the search and increase the likelihood that important characteristics will be chosen. In WBBAT, each feature is assigned a weight that reflects

its importance or relevance to the problem being solved. The feature weight can be determined based on domain knowledge, prior information, or through an adaptive mechanism during the optimization process.

The algorithm follows a similar structure to the BBA but with modifications in the position update equation. A binary vector, where each element represents the presence or absence of a feature, is used to represent the location of each bat. The weighting mechanism is applied during the position update to bias the search toward more important features. The bat searches the vertices of a hypercube, which depicts the search space in feature selection as an n-dimensional Boolean lattice. Considering that looking to accept or reject a certain feature, it is represented as the bat's position as a binary vector. In this research, sigmoid-based, Binary Bat Algorithm that exclusively accepts binary input for decision making is used.

$$S\left(u_i^j\right) = \frac{1}{1+e^{-u_i^j}} \text{ ------ (7)}$$

Therefore, Equation (7) can be replaced by (8):

$$x_j^i = \begin{cases} 1 & \text{if } S\left(u_i^j\right) > \sigma \\ 0 & \text{otherwise} \end{cases} \text{ ------(8)}$$

where $u(0,1)$ is provides binary values for the coordinates of each bat in the Boolean lattice, where 1 indicates if the feature is present and 0 indicates that it is not.

3.3 Classification using SVM and LR

The SVM is a widely used machine learning technique for both classification and regression. When there are more characteristics than samples, or when the data is not easily separated linearly, SVMs shine. In comparison to other classification techniques, support vector machines (SVMs) offer various benefits, such as their capacity to deal with high-dimensional data, their resistance to noise and outliers, and their flexibility in dealing with non-linear decision boundaries. LR is widely used in machine learning because it is straightforward to understand and can accommodate both linear and non-linear limits on the decisions that may be made. The model predicts the likelihood of a binary outcome given a set of inputs. In this work, preprocessing with and without WTKMC is compared. For

feature selection RFC, EN, BBA, WBBAT were applied and then SVM and LR performance metrics were used for evaluating the results.

4. RESULTS AND DISCUSSION

Accuracy refers to a model's propensity to properly identify whether or not a given patient has cardiac disease. Predictions of cardiac disease need to be as accurate as possible to avoid both needless medical procedures and delays in diagnosis and treatment. Recall is crucial since incorrect negative prognoses might have catastrophic effects on health. The F1 score is a measure of the model's overall effectiveness in predicting cardiovascular illness, and it combines accuracy and recall.

In this Section results obtained from the classifiers are presented and their performance in terms of afore mentioned metrics are discussed. Finally, the implication of the results and potential future research directions are presented. Figure 2 shows 14 different attribute correlations of heart disease data set. Figure 3 shows projection of the prevalence of heart disease in persons of varying ages and the distinct eight features selected in the algorithm is shown in Figure 4.

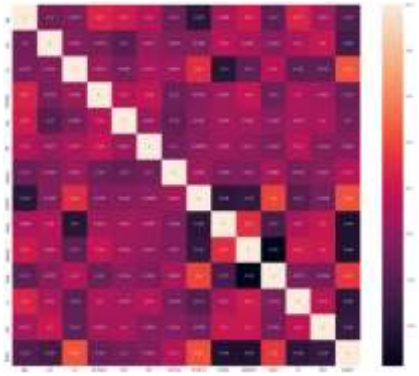

Figure 2: Heat map of 14 attributes

Figure 3: Age Categorization

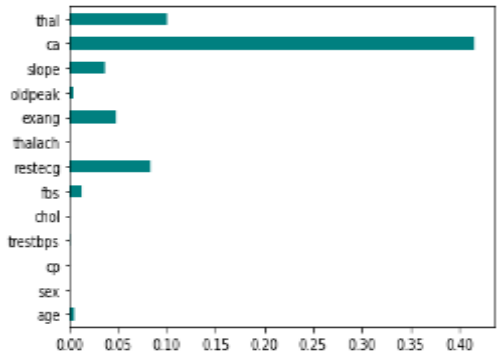

Figure 4: Selected Features

Methods	Feature Selection Accuracy	Total number of Features	Selected Features
RFC	89	14	6
EN	90	14	7
BBA	90.56	14	5
WBBAT	91	14	6
Ensemble Feature selection	96.50	14	8

Table 1: Feature Selection Accuracy

Table 1 shows the feature selection accuracy, total number of features (sex, cp, trestbps, chol, target, thalach, thal, ca, slope, oldpeak, exang, restecg, fbs, age) and the selected features (thal, ca, slope, oldpeak, exang, restecg, fbs, age) for the five different

feature selection techniques. RF classifier selected six characteristics (thalach, exang, old peak, restecg, fbs, age) from a total of 14 with 89% accuracy. EN achieved 90% feature selection accuracy and used 7 features (ca, slope, oldpeak, exang, restecg, fbs, age). With a feature selection accuracy of 90.56%, the BBA approach did marginally better. It found 5 significant features (exang, restecg, fbs, age, trestbps) indicating its ability to capture the dataset's most essential qualities. WBBAT chose 6 features (exang, restecg, fbs, age, oldpeak, ca) with feature selection accuracy of 91%. When compared to the RFC and BBA, WBBAT achieved little higher accuracy finding the most relevant traits. Compared to the above approaches, the proposed Ensemble Feature selection method showed accuracy of 96.50% with significant improvement compared to RFC, EN, BBA, WBBAT. It selected eight (thal, ca, slope, oldpeak, exang, restecg, fbs, age) features out of a total of fourteen, demonstrating its capacity to successfully identify a higher number of significant features. Hence, the findings illustrate the effectiveness of ensemble feature selection approach.

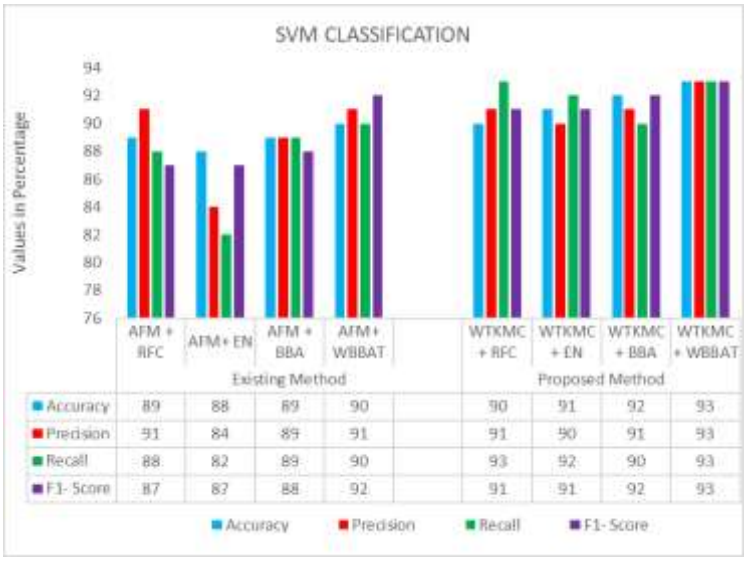

Figure 5: SVM Classification Results

Figure 6: LR Classification Results

Figures 5 and 6 compare the accuracy, precision, recall, and F1-Score of the SVM and LR classification algorithms, respectively. With the suggested WTKMC strategy for preprocessing and the defined feature selection techniques, the findings show that SVM achieves a higher accuracy level of 93% across all measures. Logistic Regression followed closely behind, with an accuracy of 92% and precision of 93%. However, its recall and F-measure were slightly lower at 91% and 92%, respectively for the combination of WTKMC with WBBAT feature selection methods. These findings emphasize that WTKMC and WBBAT achieved better and reliable classification results.

5. CONCLUSION

Long-term, human lives might be saved via improved early diagnosis of heart diseases if medical professionals had a better grasp on how to handle raw healthcare data pertaining to cardiac information. This research proposes a technique for predicting the likelihood of getting heart disease based on specific features. In this work, Weighted Transform K-Means Clustering is applied for data preprocessing and Ensemble feature selection is used with hybrid methods to select the most relevant features. Using the hybrid methods, classification accuracy of 96.50% is achieved. This work achieved high accuracy with WTKMC preprocessing technique and WBBAT feature selection. Classification accuracy using SVM and LR models show significant improvement using WTKMC preprocessing. The technique is useful for identifying the critical features. In future, more

machine learning techniques can be incorporated to improve the accuracy of heart disease prediction.

VI. REFERENCES

1. Adele Cutler, D. Richard Cutler & John R. Stevens, Random Forests, Ensemble Machine Learning Methods, and Applications, 2011, pp 157-176, doi: 10.1007/978-1-4419-9326-7_5.
2. Alenizi, M. A., & Mansour, H. Y.A, A New Intelligent Hybrid Feature Selection Method, IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), 2018, doi:10.1109/aiccsa.2018.8612812.
3. Anurag Goel., &Angshul Majumdar., Transformed K-means Clustering, Computer Science and Machine Learning, 2021, doi: 10.48550/arXiv.2111.13921.
4. Bahr kadhim Mohammed &Ashwaq Abdul Sada Kadhim., Elastic Net Variable Selection Regularization Method for Regression Discontinuity Designs with Application, International Journal of Agricultural and Sciences, 2021, vol. no. 17, pp. 1423-1430, doi: https://connectjournals.com/03899.2021.17.1423
5. Friedman, L., &Komogortsev, O. V., Assessment of the Effectiveness of 7 Biometric Feature Normalization Techniques, IEEE Transactions on Information Forensics and Security, 2018, doi:10.1109/tifs.2019.2904844.
6. Ganjei, M.A., &Boostani, R., A Fast Hybrid Feature Selection Method, 9th International Conference on Computer and Knowledge Engineering (ICCKE), 2019, doi:10.1109/iccke48569.2019.8964884.
7. Hajizamani, M., Helfroush, M. S., &Kazemi, K., Optimum Feature Selection Using Hybrid Grey Wolf Differential Evolution for Motor Imagery Brain-Computer Interface. 10th International Conference on Computer and Knowledge Engineering (ICCKE), 2020, doi:10.1109/iccke50421.2020.9303629.
8. Jain, D., & Singh, V., Diagnosis of Breast Cancer, and Diabetes using Hybrid Feature Selection Method, 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018, doi:10.1109/pdgc.2018.8745830.
9. Jayed, T., Hasan, M. A. M., &Masrur, T., Predicting N1-and N6-methyladenosine RNA Modifications using Hybrid Feature Selection Approach,5th International Conference on Advances in Electrical Engineering (ICAEE), 2019, doi:10.1109/icaee48663.2019.8975621.
10. Mohan, S., Thirumalai, C., & Srivastava, G., Effective Heart Disease Prediction using Hybrid Machine Learning Techniques, IEEE Access, 2019, vol.no. 7, pp. 81542-81554, doi:10.1109/access.2019.2923707.

11. Ni, C., Liu, W., Gu, Q., Chen, X., & Chen, D., FeSCH: A Feature Selection Method using Clusters of Hybrid-data for Cross-Project Defect Prediction, IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), 2017, doi:10.1109/compsac.2017.127.
12. Nisar, S., & Tariq, M., Intelligent feature selection using hybrid-based feature selection method, Sixth International Conference on Innovative Computing Technology (INTECH), 2016, pp 168-172, doi:10.1109/intech.2016.7845025.
13. Prabavathi G.T. & Sindhu M, A Review on Heart Disease Predictive System Using Machine Learning Algorithms, Shodha Prabha, 2022, vol. no. 47, pp. 37-42.
14. Ravishankar S, B. Wen and Y. Bresler, Online Sparsifying Transform Learning—Part I: Algorithms, In IEEE Journal of Selected Topics in Signal Processing, 2015, vol. 9, no. 4, pp. 625-636, 2015.
15. Saha, P., Patikar, S., &Neogy, S., A Correlation - Sequential Forward Selection Based Feature Selection Method for Healthcare Data Analysis, IEEE International Conference on Computing, Power, and Communication Technologies (GUCON), 2020, doi:10.1109/gucon48875.2020.9231205.
16. Sathya Durga, V., &Jeyaprakash, T., An Effective Data Normalization Strategy for Academic Datasets using Log Values, International Conference on Communication and Electronics Systems (ICCES), 2019, pp. 610-612.
17. Shahid, A. H., Singh, M. P., Roy, B., &Aadarsh, A., Coronary Artery Disease Diagnosis Using Feature Selection Based Hybrid Extreme Learning Machine, 3rd International Conference on Information and Computer Technologies (ICICT), 2020, doi:10.1109/icict50521.2020.00060.
18. Spencer, R., Thabtah, F., Abdelhamid, N., & Thompson, M., Exploring feature selection and classification methods for predicting heart disease, DIGITAL HEALTH, 2020, doi:10.1177/2055207620914777.
19. SeyedaliMirjalili., Seyed Mohammad Mirjalili., & Xin-She Yang, Binary bat algorithm, Neural Computing & Applications, Springer 2013, doi:10.1007/s00521-013-1525-5.
20. Tania, F. A., & Shill, P. C., A Hybrid Kernal Support Vector Machine with Feature Selection for the Diagnosis of Diseases, 4th International Conference on Electrical Information and Communication Technology (EICT), 2019, doi:10.1109/eict48899.2019.9068804.
21. Thejas, G. S., Joshi, S. R., Iyengar, S. S., Sunitha, N. R., & Badrinath, P., Mini-Batch Normalized Mutual Information: A Hybrid Feature Selection Method, IEEE Access, 2019, vol.no. 7, pp 116875–116885, doi:10.1109/access.2019.2936346

22. Xin-She Yang.,Bat Algorithm: Literature Review and Applications, Int. J. Bio-Inspired Computation, 2013, vol 5. no.3, 141-149, doi:10.1504/IJBIC.2013.055093.
23. Youguo Li., & Haiyan Wu.,A Clustering Method Based on K-Means Algorithm. Elsevier, Science Direct- Physics Procedia, 2012, vol. no. 25, pp 1104-1109, doi:10.1016/j.phpro.2012.03.206.
24. Zhang, N., Xiong, J., Zhong, J., & Thompson, L. Feature Selection Method Using BPSO-EA with ENN Classifier, Eighth International Conference on Information Science and Technology (ICIST), 2018, doi:10.1109/icist.2018.8426154.
25. Zhou, Z., Wang, Y., & Li, M., Feature Selection Method Based on Hybrid SA-GA and Random Forests, International Conference on Computing and Data Science (CDS), 2020, doi:10.1109/cds49703.2020.00034.
26. Zou, H. and T. Hastie., Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: series B (Statistical Methodology), 2005, vol. no. 67(2), pp 301-320.
27. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and Regression Trees. Wadsworth, New York (1984).
28. Burak Kolukisa., Burcu Bakir-Gungor., Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis, Computer Standards & Interfaces, 2023, vol.no. 84, doi:https://doi.org/10.1016/j.csi.2022.103706.
29. Yvan Saeys., Thomas Abeel., & Yves Van de Peer., Robust Feature Selection Using Ensemble Feature Selection Techniques,Machine Learning and Knowledge Discovery in Databases, European Conference, 2008, vol.no 5212, pp 313–325.
30. Xiao-Yan Gao., Abdelmegeid Amin., Hassan Shaban Hassan., &EmanM.Anwar., Improving the Accuracy for Analyzing Heart Disease Prediction Based on the Ensemble Method, Hindawi, 2021,doi: https://doi.org/10.1155/2021/6663455.
31. Kaushalya Dissanayake.,& Md Gapar Md Johar., Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms, Applied Computational Intelligence and Soft Computing, 2021, doi: https://doi.org/10.1155/2021/5581806
32. Tasnim& Habiba.,Comparative Study on Heart Disease Prediction Using Data Mining Techniques and Feature Selection, second International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST),2021, doi: 10.1109/ICREST51555.2021.9331158.
33. Harshitha.B, Maria Rufina., & Shilpa B.L., Comparison of Different Classification Algorithms for Prediction Heart Disease by Machine Learning Techniques, SN Computer Science, 2022, doi: https://doi.org/10.1007/s42979-022-01512-3
34. World Health Organization. Cardiovascular diseases Available from https://www.who.int/cardiovascular_diseases/en/ (n.d., accessed 9 June 2019)