Journal of Industrial Information Technology and Application *ISSN(Online): 2586-0852*
Vol.7 No.2

A Novel Weighted Optimization Algorithm to Classify the Heart Disease Using Machine Learning

P. Suganya[1)] and C.P. Sumathi[2,*)]
[1)]Bharathiar University, Coimbatore-641046, Tamil Nādu, India.
[2)]Dept. of Computer Sci, S.D.N.B Vaishnav College for Women, Chrome pet, Chenna, India.

Abstract. Heart disease is difficult to detect due to several risk factors, including high blood pressure, cholesterol, and an abnormal pulse rate. Accurate and timely identification of human heart disease can be very helpful in preventing heart failure in its early stages and will improve the patient's survival. Manual approaches for the identification of heart disease are biased and prone to inter examiner variability. Therefore, detecting heart disease early by utilizing the affluence of high-resolution intensive care records has become a challenging problem. That is why many researchers are trying their best to design a predictive model that can save many lives using data mining. Even though, some Machine Learning (ML) based models are also available, which can reduce the mortality rate, but accuracy is not up to date. According to the recommended study, using a Modified Weighted Empirical Score Optimization (MWESO) with Logistic Regression (LR) algorithm this research identified and predicted human heart disease. Machine learning (ML) algorithms like K-Nearest Neighborhoods (KNN), Support Vector Machine (SVM), Logistic Regression (LR) and Naïve Bayes (NB) have been applied to the heart disease dataset to predict the disease. At first, the LR model was trained. After training, sum of two features decision was combined using a weighted sum optimization. The weights have been assigned to each attribute's decision probability hence that each attribute's effect varies in the summation of weighted empirical score that gave the optimized prediction from the final decision score. The datasets were acquired from the heart diseases repositories from Kaggle. The comparative study has proven that the proposed MWESO algorithm with LR is the most suitable model due to its superior prediction capability to other Machine Learning with an accuracy of 90.7% on heart ailments dataset.

Keywords. Heart disease, Prediction, Modified Weighted Empirical Score Optimization, Machine Learning, Classification

---

[*]Corresponding author: *drcpsumathi@gmail.com*
Received: Apr 19, 2023; Accepted: May 21, 2023; Published: Jun 30, 2023

1. Introduction

A disease in the human body is an unnatural medical condition. It affects negatively the human body organism's functional state. It is generally associated with few signs of illness in the patient body. According to the World Health Organization (WHO), in the last 15 years, an estimated 17 million people die each year from cardiovascular disease, particularly heart attacks and strokes. Heart disease refers to a series of conditions that include the heart, vessels, muscles, valves, or internal electrical pathways responsible for muscle contraction. According to the Centers for Disease Control and Prevention (CDC), heart disease is one of the leading causes of death in India, the UK, the US, Canada, and Australia. Manually, detecting heart disease needs doing several tests. By analyzing the result of tests, it can be assured whether the patient got heart disease or not [1]. It is time consuming and costly to predict heart disease in this conventional way. Cardio Vascular Diseases (CVDs) are a leading cause of clinical (i.e., death and disability), health, and economic burden globally, accounting for approximately 31% (17.9 million) of total deaths each year, one in four deaths in the USA occurs as a result of heart disease [2]. Heart disease is common among both men and women in most countries around the world. Therefore, people should consider heart disease risk factors. Although it plays a genetic role, some lifestyle factors significantly affect heart disease. The known risk factors for heart disease; radiation therapy for age, gender, family history, smoking, some chemotherapy drugs and cancer, malnutrition, high blood pressure, high blood cholesterol levels, diabetes, obesity, physical mobility, stress, and poor hygiene [3]. These are the various risk factors in which the patient's exposure towards developing a Cardio Vascular Diseases (CVD). Researchers are attempting to develop an effective technique for the timely identification of heart diseases as existing heart disease diagnosis methods are ineffective in early detection for various reasons, including accuracy and computational time [4]. When advanced technology and healthcare experts are unavailable, diagnosing and controlling heart disease is incredibly challenging [5]. Many people's lives can be saved with a good, solid diagnosis and treatment [6]. A physician's evaluation of the patient's medical history, physical examination report, and analysis of concerning symptoms are used to diagnose heart disease. However, the findings of this method of diagnosis are sufficient in detecting heart disease patients. Furthermore, it is both costly and computationally challenging to examine [7]. Thus, we build a non-invasive prediction system to handle these issues using ML classifiers. Heart diseases are efficiently diagnosed using an expert decision system relying on ML classifiers and artificial fuzzy logic. As a consequence, the death ratio declines [8, 9]. Numerous researchers used the Cleveland heart disease dataset. In diagnosing clinical data, significant improvement has been observed using ML. Several diseases have been predicted using ML like diabetes, heart disease, breast cancer. ML is closely linked to both statistics and decision-making. For training and testing, the predictive models of ML require appropriate data. The world data is growing day by day, and hospitals are slowly adopting big data systems. Applying data analysis in the medical sector is giving excellent benefits. It improves the result and reduces cost. Effective implementation of ML-based optimization improves physicians' work and increases the productivity of the healthcare service. They can be used for several purposes, such as forecasting the amount of product sold, forecasting covid cases in the upcoming month, the probability of rainfall occurring in a particular area, selling airline tickets, etc. [10, 11]. As the medical sector has a large dataset, these existing data will help the researchers diagnose the disease early by systematically analyzing data [12]. Therefore, ML based optimization process saves much time and therefore improves the efficiency of the diagnosis. This is one of the reasons we have developed empirical optimization models to classify heart disease efficiently.

The primary purpose of this study is to classify patients with heart disease using medical records. The classification model in general can predict the severity stage of the patients with heart disease. This research work has used different ML algorithms to classify heart disease. First, the data was pre-processed. Then the LR algorithm was trained with the dataset. The weighted sum rule was applied to the decision score provided by the trained algorithm. The summation of weighted empirical optimization model used the new score to classify the disease. So, in this work, three weighted score fusion models were generated, which provided an improved performance compared to the previous separate ML algorithms. A ML optimization-based predictive system can reduce physicians' pressure, and a weighted score optimization approach will help diagnose disease more efficiently.

The organization of the article is as follows: Section 2 describes related work based on prediction of disease using ML, Section 3 describes data description, pre-processing and proposed methodology based on weighted empirical optimization classifier, Section 4 describes results and discussion based on evaluation metrics and conclusion in section5.

2.Literature Review

Nowadays, many researchers around the world are focusing on ML algorithms in the health field to forecast different diseases. And the use of ML in the medical sector has given a notable change in the performance of treatment. This section discusses the various diseases such as diabetes, cardiovascular and kidney, which have been diagnosed using ML algorithms.

Comparative analysis of various diseases has been done using machine learning algorithms [12, 13, 14, 15, 16, 17]. In [2], a comparative analysis was developed using machine learning algorithms. Both cardiovascular and diabetes disease datasets were used for classification. Different machine learning algorithms like XGBoost, random forest, and weighted ensemble models were used to predict disease. The essential features that contribute most to the dataset were identified. At last, the performance parameters had shown to observe the results. Ensemble models showed slightly better percentage of accuracy than other models. The researchers also proposed to apply the model in a real-world scenario to check the risk of the disease occurring. They used ten-fold cross-validation, and its accuracy was slightly low. So, to increase the accuracy, we have used a fusion of two algorithms rather than a single algorithm.

Multiple diseases such as cardiovascular, Chronic Kidney Disease (CKD), and diabetes were identified in [18, 19, 20, 21]. Support vector, decision tree, and random forest algorithms were used for classification with a standardized decision support model in [18]. A Chi-square method was employed to select the best features. They implemented SVM (linear, polynomial, and radial) procedures using extracted features. The performance was evaluated with the help of accuracy, specificity, miss classification rate classification parameters, and other parameters. Improved SVM-Radial gave the best accuracy of all the algorithms. We got a better accuracy using weighted score fusion.

Kibria and Matin [22] discussed some fusion models to diagnose CVDs along with its severity. The proposed approach has been experimented with different test training ratios for binary and multiclass classification problems, and for both of them, the fusion models performed well. The highest accuracy for multiclass classification was found as 75%, and it was 95% for binary.Enaset.al [23] research is to provide a comparative analysis of different machine learning models to reach the most supporting decision for diagnosing heart disease with better accuracy as compared to existing models. The comparative study has proven that the XGB is the most suitable model due to its superior prediction capability to other models with an accuracy of 91.6% and 100% on two different heart ailments datasets, respectively. Salhi et al [24] described data analytics to detect and predict disease's patients. Starting with a pre-processing phase where selected the most relevant features by the correlation matrix, then applied three data analytics techniques (neural networks, SVM and KNN) on data sets of different sizes, in order to study the accuracy and stability of each of them. The neural networks are easier to configure and obtain accuracy up to 93%.

Yekkala et al. [25] used Particle Swarm Optimization (PSO) in conjunction with particle methods (Random Forest, AdaBoost, and Bagged Tree) to more accurately predict the results. The Heart Stalog dataset has 270 samples and 14 attributes, taken from the UCI database [5]. The data has already been processed, and PSO is used as a feature selection method to delete unnecessary and missing data. Powerful features continued, and the AdaBoost, Bagging, and Random Forest. The two factor's importance to full features. Eventually, the performance of each algorithm is measured. As a result, Bagged Tree performed 100%, Random Forest 90.37%, and AdaBoost 88.89%. According to test results, Yekkal et al. [25] proved that using Bagging Trees on PSO will improve learning accuracy in predicting heart disease. Amin et al. [26] show a heart disease prediction model using a genetic algorithm, neural network, Naïve Bayes, Bagging Trees, Decision Tree, Core Density, and SVM. Learning is faster, more stable, and accurate compared to back-propagation. Collected risk factors data of 50 patients and the hybrid model resulted in 96% training accuracy and 89% test accuracy. Amin and his colleagues then developed the system using the hybrid fuzzy and k-nearest neighbour approach to predict heart diseases; in another system, using the neural network community was used with an accuracy of 89.01% in the diagnosis of heart disease.

Almazroiet.al [27] research contributes to the body of literature by selecting a standard well defined, and well-curated dataset as well as a set of standard benchmark algorithms to independently verify their performance based on a set of different performance evaluation metrics. From our experimental evaluation, it was observed that decision tree is the best performing algorithm in comparison to logistic regression, support vector machines, and artificial neural networks. Decision trees achieved 14% better accuracy than the average performance of the remaining techniques. Khourdifi and Bahaj [28] opined that for CVD prediction machine learning algorithms can provide good results in comparison to other techniques as they can model complex problems with non-linearity.

The authors explored the concept of selective features selections which imply that not all features are important to predict the outcome. Further, the authors proposed to use particle swarm optimization and ant colony optimization techniques in conjunction with neural networks, random forest and support vector machines. From the literature survey, this research concluded that a variety of techniques are used in the literature for heart disease, however, for survival prediction there is a gap to experimentally evaluate the ML algorithms on a standard data set. This research fills this gap by considering ML based empirical optimization algorithm and evaluate them on a heart disease dataset repository from Kaggle.

3. Research Methodology

The workflow of the system has been implemented in different stages including Pre-processing of the dataset, proposed NWEO algorithm, classification and performance evaluation as depicted in figure 1.

Heart disease is diagnosed with the help of Kaggle datasets. Moreover, it is divided into training and testing set.

3.1 Tools Used

The Pandas tool is an open-source python package providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data. It is used to conduct this study, which is written in python or C. Tools for writing and reading data between in-memory data structures and various formats such as Microsoft Excel, Comma-Separated Values (CSV) format. Matplotlib is a comprehensive library for creating animated, interactive, and static visualizations in Python used for ML. In ML, it is useful to understand the vast amount of data through different visualization.

3.2 Dataset Description

Misleading datasets are used in this research. They are taken from the Kaggle repository. The first dataset contains a total of 303 cases, 138 of which are healthy people and 165 have heart disease [29]. The datasets have been selected with 76 attributes and pre-processed to produce 14 only for reducing the redundant variables. Four attributes are used to indicate common symptoms of the patient, and the remaining attributes are used to indicate ECG values. The attributes for datasets are shown in detail in Table 1.

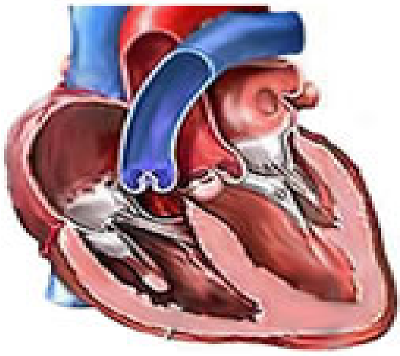

Table.1 Attribute information of dataset

S. No	Attribute used	Attribute details
1.	Sex	The patient's gender specifies in binary form. Male=1, Female=0
2.	Age	Age of the patient where ranges from 29 to 77 years
3.	Resting Blood Pressure (RBP)	Patient RBP ranges from 120 to 154
4.	Fasting blood sugar (FBS)	Patient FBS, higher than 120 mg/dl. True=1, False=0
5.	Maximum Heart Rate (MHR)	Patient MHR range from 71 to 202
6.	Serum cholesterol (Chol.)	Patient chol range from 120 to 154 mg/dl
7.	Exercise induced angina (Exang.)	Exang represent in binary 1-yes, 0-no
8.	Chest pain type (CP)	CP ranges from 1 to 4, where 1 represents typical angina, 2 represents atypical angina, 3 represents Nonanginal pain and 4 represents No pain
9.	Resting electrocardiographic results (RECG)	Patient RECG ranges from 0 to 2 where 0 represents normal, 1 represents ST-T wave, 2 represents probable or definite left ventricular hypertrophy.

10.	Old peak (OP)	ST depression relative to rest ranges from 0 to 6.2
11.	Thallium scan (Thal.)	Thal represent patient's heart scan 3,6,7. 3 represents Normal, 6 represents fixed defect and 7 represents reversible defect.
12.	Number of coloured by fluoroscopy (CA)	CA ranges as 0.3 related to darkness of the color
13.	Slope of peak exercise ST segment (Slope)	Slope for peak range 1 to 3 where 1 represents upsloping, 2 represents flat and 3 represents down sloping
14.	Target (TRT)	Diagnosis of heart disease, 0 represents absence of heart disease, 1 represents presents of heart disease.

Correlation is used to determine the relationship between two continuous, quantitative variables. The determination of relevant features is performed using the correlation technique. The correlation matrix is computed to detect the relationship between attributes of the dataset. This can improve the ML cancelled weakly correlated attributes. The correlation matrix for the two datasets is plotted in Figure 2 to well understand the correlation between the attributes. It is depicted in different colors, the dark color represents that the attribute are strongly correlated with another and light color performs a weakly correlated with another. Correlation values are in range from (−0.4 to +1.0). Positive correlation increases or decreases the column attributes together. Negative correlation performs that one attribute will increase and another one decreases or vice versa.

3.4 Data Pre-Processing

Pre-processing data means the changes which are made on data before it is fed as an input to the algorithm. Data obtained from many sources is described as raw data, not suitable for analysis. In order to obtain better results, it is necessary to remove outliers, noise, and irregularities from the data, known as data cleaning as described below [30].

Data Cleaning: The data that needs to be analysed using algorithms of machine learning may be noisy, inconsistent and incomplete. It also deals with the missing values for attributes of interest as it changes the proper average value for the attribute. Likewise, invalid attribute values are cleared and filled manually with its mean value. Data is cleaned up by manipulating missing values, smoothing out noisy data and removing outliers [31]. The dataset utilized to predict heart disease based on 14 variables is shown in Figure 2.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Figure.2 recommended dataset based on randomizing the rows

3.5 Proposed MWEO algorithm

The proposed algorithm specifies decision probability of (D_P) for each feature of heart dataset to predict the test data. From this decision score, the prediction was made for the test data. Different weights have been assigned to each feature D_P so that each features effect varies in the summation of weights. If one of the features have a higher rate for the right decision, we assigned a bigger weight to that and a comparatively smaller weight to the other features. So, if one of the features has a weight of .65, then the other will have (1-.65) =.35. Here used a loop to check which weights provided the best accuracy for the summation of weighted empirical score model and selected them. The sum of the weights used in the empirical model should be 1 for scaling. Then selected the weights that have the best result for the summation of empirical optimized model. The equation for the weighted sum is as follows equation 1

$$D_W = \sum_{p=1}^{N} W_p * D_p \tag{1}$$

Where N is the number of the features used for summation of empirical optimized model, then have used two features for every combination of empirical optimized model, so N=13. D_W is the weighted sum, which is the new decision score of the weighted empirical score optimization model. Based on this score, the final decision was given. W_p represents the weighted probability that has been assigned to feature with the decision score, and D_p is the decision probability score for any individual attributes.

The process of building a weighted empirical score optimization model is illustrated below. The thirteen features were used to create an empirical optimized model. A weighted score level optimization model was developed by merging those as a single weighted empirical feature score, and it worked better than the individual features scores. The algorithm for weighted empirical score optimization model is given below:

Algorithm for MWESO algorithm

Input: W_p represents weights with the decision score, D_p represents individual attributes for decision score after training, N represents number of separated features

Output: D_W represents new decision score of the weighted empirical score optimization model

Step 1 W_1 = 1

Step 2 for p = 0 to 20 do

Step 3 D_p = 0

Step 4 $\sum_{p=1}^{N} W_p$

Step 5 W_1 = W_1 -0.05

Step 6 W_2 = 1 - W_1

Step 7 for p = 0 to N do

Step 8 Select the weights (W1, W2) that gave the highest decision score.

Step 9 Final weighted empirical score optimization using selected weights.

After selecting weights, the weighted empirical optimization rule was applied at the last step in the algorithm. A weight was assigned to each individual features decision score in a weighted sum. To select the weights, we used a loop for using various values of weights and the weights that have the highest decision score were selected. The individual result was combined, but the outcome of the 13 features in LR algorithm is not taken equally. Rather than weights were assigned that decided the effect of any features in the weighted empirical score optimization model. Train the above algorithm with LR model and this step is implemented at the decision level.

Logistic Regression is a specific form of Generalized Linear Model (GLM), which is frequently abbreviated as a GLM. It is like linear regression, but it predicts true or false. Instead of adjusting a line to the data, LR provides a s-shaped logistic function. The output probability for any given problem is found from the curve, and it is generally used for binary classification. Logistic regression can work with both continuous and discrete data. In linear regression, the line is fit by least squares, whereas logistic regression uses maximum likelihood. The likelihood is calculated for different curves, and the curve with maximum likelihood is selected for classification. Logistic regression determines the probability of the binary response.

The line for linear regression is selected in such a way that the distance between the summation of all points and lines should be minimum. The equation for the plane is written as:

$$y = w^T x + b \quad (2)$$

The trouble with linear regression is that we need to adjust the best fit line whenever new data comes. The exact coefficient is needed to find that will understand which the best match line for that model is, and then those particular coefficients or slopes will be adjusted to achieve the optimal plane. If x is the data point, then w^T x is the distance between a particular data and plane, considering b=0. Distance between data point and the particular plane is: $\sum_{p=1}^{N} W^T x_p$

Classes are classified according to the equation: $y_p W^T x_p \geq 0$ which are correctly classified, $y_p W^T x_p \leq 0$ which are incorrectly classified.

To get the best fit line, cost function $\sum_{p=1}^{N} y_p W^T x_p$ should be maximum. Here, W^T is the coefficient which needs to update until maximum summation is found. The impact of an outlier has been avoided by adding the sigmoid function in the equation. Then the equation becomes:

Max $\sum_{p=1}^{N} f\ (y_p\ W^T x_p)$

The sigmoid function is denoted as f in this equation. In LR, we tuned the parameter C, the regularization strength. For our proposed MWEO model, we used C = 1.

4. Result and Discussion

In this research, proposed MWESO with LR, KNN, LR, SVM and NB classifier algorithms are applied to the heart diseases datasets acquired from Kaggle repository respectively. The dataset used in this research is splitting into 75% and 25%, which 75% of original data is considered as training dataset and 25% as testing dataset. Training dataset is used to train a model and testing dataset to check the performance of the trained model. A notable improvement was found in the result of using weighted empirical optimization models. The key reason behind the proposed model's improvement is that if there is a miss classification in first weight score, there is a probability that another weight score may classify that particular data correctly. So, after summation of weighted empirical score, there is a chance that we get the correct result for that specific case. This concept assists in giving the weighted empirical optimization model a better efficiency.

The model's performance can be interpreted from the value of these parameters. The first True Negative (TN) classifier predicted "no heart disease" and identified patients who are not affected by sepsis. The second False Positive (FP) classifier predicted individuals who are not affected by "heart disease". The third False Negative (FN) classifier accurately recognized patients with heart disease while predicting "no heart disease". The fourth Ture Positive (TP) is a classifier that predicted "heart disease" and identified those who had it. To illustrate the classifiers' performance, a confusion matrix has been used for the proposed MWESO classification model with LR, KNN, RF, SVM and NB classifier. It summarizes the results of the predictions of a model. The matrix is given in table 2.

Table. 2 Confusion Matrix for all the classification model.

Machine Learning Models	TP	TN	FP	FN
Proposed MWESO with LR	36	33	3	4
KNN	33	28	8	7
NB	28	34	6	8
LR	32	28	2	14
SVM	29	30	8	9

4.1 Performance Parameter

Classification model performance is measured with the term of accuracy, precision, recall, sensitivity and specificity.

Accuracy is the measure of the percentage of correctly classified objects.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} * 100$$

Precision is also referred to as the false-positive rate. From precision, we get the number of correctly classified observations as positive to the total classified positive observations.

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

Recall is often referred to as a truly positive rate. It is the ratio of total positive assumptions and the total amount of positive class attributes.

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

Sensitivity: It determines how much of a classifier to identify positive labels.

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity: It is assessed what proportion of patients to identify negative labels.

$$\text{Specificity} = \frac{TP}{FP+FN}$$

Table.3 Performance metrics based on proposed model with various classification models

Model	Accuracy	Precision	Recall	Sensitivity	Specificity	Kappa
Proposed MWESO with LR classifier	0.907	0.9231	0.9000	0.9000	0.9167	0.9565
KNN classifier	0.8026	0.8049	0.8250	0.8250	0.7778	0.7104
NB classifier	0.8158	0.8235	0.7778	0.7778	0.8500	0.7124
LR classifier	0.7895	0.9412	0.6957	0.6957	0.9333	0.7035
SVM classifier	0.7763	0.7838	0.7632	0.7632	0.7895	0.7026

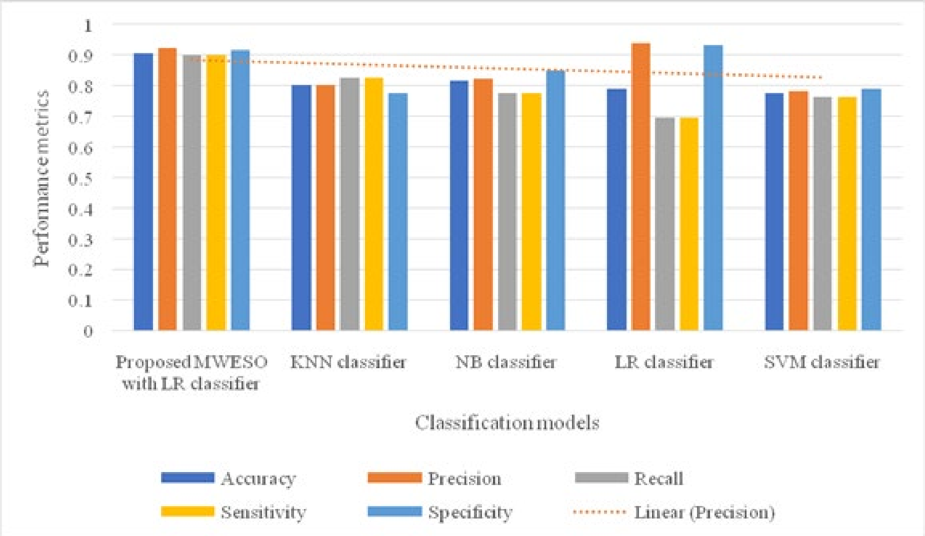

Figure .3 Graphical representation of performance evaluation based on various models

Table 3 and figure.3 shows the performance of proposed MWESO classification model with LR, KNN, LR, SVM and NB classifier. After the process of building a weighted empirical score optimization model based on merging a single weighted empirical feature score which result as performance parameters increased (LR) for proposed MWESO with LR classifier. It has the highest accuracy of 90.7% among the other classifier models.

5. Conclusion

A significant percentage of the world's population is struggling with heart disease. To identify patients with heart disease, empirical optimization approach is necessary. This paper has focused on implementing a weighted score prediction model to identify patients with and without heart disease using LR algorithms.The proposed MWESO with LR classification classifies only the presence and absence of heart disease compared with various ML classifier.This was performed on heart disease datasets from Kaggle containing 13 features but different in number of recorded instances. The weighted empirical optimization models reduced the risk factors and improved the output in terms of accuracy and other factors. The research workhas a good improvement in the merging of weighted score which makes the model more reliable. After training LR model, the decision from each features effect varies in the summation of weightswas taken and the scores were merged to form a new score. This new score is the decision score of our optimization model. Based on this score, the output will be predicted. The accuracy obtained for proposed MWESO with LR model has achieved the best results with an accuracy of 90.7%based on empirical decision score.

Reference

[1] Md. Jubier Ali, Badhan Chandra Das, Suman Saha, Partha Chakraborty, A Comparative Study of Machine Learning Algorithms to Detect Cardiovascular Disease with Feature Selection Method, Aug 2022.

[2] Felman, A. (2018). Everything you need to know about heart disease. Medical News Today, https://www.medicalnewstoday.com/articles/237191#types, accessed date: 05/02/2021.

[3] Thomas J, Princy RT. March. "Human heart disease prediction system using data mining techniques,". In: 2016 international conference on circuit, power and computing technologies (ICCPCT). IEEE; 2016. p. 1–5.

[4] L. A. Allen, L. W. Stevenson, K. L. Grady et al., "Decision making in advanced heart failure: a scientific statement from the American heart association," Circulation, vol. 125, no. 15, pp. 1928–1952, 2012.

[5] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," Journal of Intelligent Learning Systems and Applications, vol. 5, no. 3, Article ID 35396, 2013.

[6] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150–154, 2011.

[7] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average

Parkinson's disease symptom severity," Journal of 'e Royal Society Interface, vol. 8, no. 59, pp. 842–855, 2011.

[8] S. I. Ansarullah and P. Kumar, "A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method," International Journal of Recent Technology and Engineering, vol. 7, no. 6S, pp. 1009–1015, 2019.

[9] S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, "Fuzzy logic-based decision support system for component security evaluation," 'e International Arab Journal of Information Technology, vol. 15, no. 2, pp. 224–231, 2018.

[10] Hafsa Binte Kibria, Oishi Jyoti, and Abdul Matin. Forecasting the spread of the third wave of covid-19 pandemic using time series analysis in bangladesh. Informatics in Medicine Unlocked, page 100815, 2021.

[11] Sam Cramer, Michael Kampouridis, Alex A Freitas, and Antonis K Alexandridis. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Systems with Applications, 85:169–181, 2017.

[12] Karthikeyan Harimoorthy and Menakadevi Thangavelu. Multi-disease prediction model using improved svmradial bias technique in healthcare monitoring system. Journal of Ambient Intelligence and Humanized Computing, pages 1–9, 2020.

[13] An Dinh, Stacey Miertschin, Amber Young, and Somya D Mohanty. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC medical informatics and decision making, 19(1):211, 2019.

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

[15] Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M Bui, Julian MW Quinn, and Mohammad Ali Moni. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Computers in Biology and Medicine, 136:104672, 2021.

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

[17] Manoj Diwakar, Amrendra Tripathi, Kapil Joshi, Minakshi Memoria, Prabhishek Singh, et al. Latest trends on heart disease prediction using machine learning and image fusion. Materials Today: Proceedings, 37:3213–3218, 2021.

[18] Hafsa Binte Kibria, Abdul Matin, Nusrat Jahan, and Sanzida Islam. A comparative study with different machine learning algorithms for diabetes disease prediction. In 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pages 1–8. IEEE, 2021.

[19] Karthikeyan Harimoorthy and Menakadevi Thangavelu. Multi-disease prediction model using improved svmradial bias technique in healthcare monitoring system. Journal of Ambient Intelligence and Humanized Computing, pages 1–9, 2020.

[20] Jing Xiao, Ruifeng Ding, Xiulin Xu, Haochen Guan, Xinhui Feng, Tao Sun, Sibo Zhu, and Zhibin Ye. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. Journal of translational medicine, 17(1):1–13, 2019.

[21] Md Kamrul Hasan, Md Ashraful Alam, Dola Das, Eklas Hossain, and Mahmudul Hasan. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access, 8:76516–76531, 2020.

[22] Sibo Qiao, Shanchen Pang, Gang Luo, Silin Pan, Zengchen Yu, Taotao Chen, and Zhihan Lv. Rlds: An explainable residual learning diagnosis system for fetal congenital heart disease. Future Generation Computer Systems, 128:205–218, 2022.

[23] Hafsa Binte Kibria and Abdul Matin. 2022. The Severity Prediction of The Binary and Multi-Class Cardiovascular Disease - A Machine Learning-Based Fusion Approach. A Preprint - March 10.

[24] Enas M. Abd Allah, Doaa E. El-Matary, Esraa M. Eid, Adly S. Tag El Dien.2022.Performance Comparison of Various Machine Learning Approaches to Identify the Best One in Predicting Heart Disease, Journal of Computer and Communications, vol-10, pp:1-18 https://www.scirp.org/journal/jcc ISSN Online: 2327-5227.

[25] Dhai Eddine Salhi, Abdelkamel Tari, and M-Tahar Kechadi.2021. Using Machine Learning for Heart Disease Prediction. Research gate. https://www.researchgate.net/publication/349470771.

[26] Yekkala I, Dixit S, Jabbar MA. August. Prediction of heart disease using ensemble learning and Particle Swarm Optimization. In: 2017 international conference on smart technologies for smart nation (SmartTechCon). IEEE; 2017. p. 691–8.

[27] Amin SU, Agarwal K, Beg R. April. Genetic neural network-based data mining in the prediction of heart disease using risk factors. In: 2013 IEEE conference on information & communication technologies. IEEE; 2013. p. 1227–31

[28] Abdulwahab Ali Almazroi.2021. Survival prediction among heart patients using machine learning techniques. Mathematical Biosciences and Engineering, Vol-19, issues-1, pp: 134–145.

[29] Y. Khourdifi, M. Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst., 12 (2019), 242–252. doi: 10.22266/ijies2019.0228.24.

[30] Heart Disease Dataset. https://www.kaggle.com/johnsmith88/heart-disease-dataset

[31] Naik, A. and Naik, N. (2021) A Generalized Model for Cardiovascular Disease Classification Using Machine Learning Techniques. In: Chiplunkar, N.N. and Fukao, T., Eds., Advances in Artificial Intelligence and Data Engineering, Advances in Intelligent Systems and Computing, Vol. 1133, Springer, Berlin, 15-26. https://doi.org/10.1007/978-981-15-3514-7_2

[32] Sathya, K. and Karthiban, R. (2020) Performance Analysis of Heart Disease Classification for Computer Diagnosis System. International Conference on Computer Communication and Informatics (ICCCI 2020), Coimbatore, 22-24 January 2020.