



Automated Cardiovascular Disease Diagnosis Using Honey Badger Optimization with Modified Deep Learning Model

Shoba V¹, Mustafa Khan GZ², Senthil Kumar S K³, Tilak Kumar S⁴,

Taheer Khan⁵

¹ Assistant Professor/CSE, Sri Sairam College of Engineering, Bangalore – 562106

^{2, 3, 4, 5} Final Year/CSE, Sri Sairam College of Engineering, Bangalore -562106

ABSTRACT

The leading cause of death among people around the world is cardiovascular disease (CVD). In order to prevent patients from other damages, precise diagnostics of CVD on time is a crucial factor. Researcher workers are inspired to apply machine learning (ML) for accurate and quick diagnosis of CVD. ML algorithm extracts patterns and hidden relationships in the medical dataset for detecting or predicting disease development. But the prediction of CVD is a challenging task. The increasing size of healthcare dataset has made it a complex task for practitioners to make disease predictions and understand the feature relations. This study develops an Automated Cardiovascular Disease Diagnosis using Honey Badger Optimization with Modified Deep Learning (ACVD-HBOMDL) Model. The major aim of the ACVD-HBOMDL technique lies in the classification of CVD using feature selection (FS) and hyperparameter tuning strategies. Initially, the ACVD-HBOMDL technique applies min-max scaler to preprocess the medical data. To elect an optimal subset of features, the HBO algorithm is used in this work. For CVD classification, deep learning modified neural network (DLMNN) classifier is used and its hyperparameters can be optimally chosen by Bayesian optimization. The experimental results of the ACVD-HBOMDL technique can be tested on benchmark medical dataset and the obtained results demonstrate the significant outcomes of the ACVD-HBOMDL technique over other existing techniques.

Keywords: Cardiovascular disease, computer aided diagnosis, deep learning, feature selection, parameter tuning, and honey badger algorithm.

1 Introduction

Cardiovascular disease (CVD) is the most prominent disease which affects the heart and blood vessels. The loss of human lives is mainly related to a lack of earlier disease diagnosis, and a pre-emptive prediction of CVD risk will considerably mitigate the situation. Usually, medical information is presented in the form of electronic medical records gathered from the patient. The widespread usage of data in healthcare is to build a decision support system that uses patient data together with artificial intelligence and domain knowledge. This could give data to assist healthcare professionals and detect errors or critical situations and alert the medical professional consequently. Artificial intelligence (AI) is more commonly applied in prediction to resolve these problems, amongst other deep learning (DL) and machine learning (ML) being



the majority. This prediction model analyses a massive quantity of healthcare information to define whether the patient has the disease and attain more precise prediction outcomes when compared to manual diagnoses.

Therefore, this Study develops an Automated Cardiovascular Disease Diagnosis using Honey Badger Optimization with Modified Deep Learning (ACVD-HBOMDL) Model. The ACVD-HBOMDL technique applies min-max scaler to preprocess the medical data.

2 Recent Works

Nagavelli examine the distinct ML methodologies dependent upon heart disease (HD) recognition. Primarily, Naïve Bayes (NB) with a weight system has been utilized to predict HD. Secondly, based on the feature of time domain, data model, and frequency domains, are automatic and analyze ischemic HD localized or recognition. Finally, the cardiac arrest automatic detection system utilizing an enhanced SVM dependent upon the duality optimized method is also investigated. In the design concept of soft sensors to measure cardiovascular risk of patients in real time is recorded. Blood oxygenation, electrocardiogram (ECG), body temperature, and data attained in patient interviews can be handled for extracting distinguishing features. The projected OANN comprises a group of 2 important procedures such as distance based misclassified instance removal (DBMIR) and teaching–learning-based optimization (TLBO) technique for ANN named (TLBO-ANN). The projected OANN technique drives on 2 stages online and offline predictions. In an effectual infrastructure employed to HD was generated employing the UCI Repository database and the health service sensors for predicting the people who suffer in HD. Additionally, the classifier systems can be utilized for classifying the patient data to detection of HD. During this training stage, the classification is trained to utilize the data in benchmark database. During this testing stage, the current patient information for identifying disease was utilized for identifying the occurrence of diseases. In the experiment, a benchmark database was tested utilizing a group of classifications such as logistic regression (LR), J48, SVM, and MLP. Atallah and Al-Mousa introduce a majority voting ensemble system which is capable of predicting the feasible occurrence of HD from humans. The predictive was dependent upon easy affordable medicinal tests led in some local hospitals. Besides, the purpose of this scheme is for providing further confidence and accuracy to Doctor Analysis as the method was trained utilizing actual data of well and ill patients. This method classifies the patient as dependent upon the majority vote of many ML techniques for providing further correct results as having only one method.

3 Proposed Work Explanation

In this study, we have designed a novel ACVD-HBOMDL technique for accurate CVD diagnosis process using FS and hyperparameter tuning processes. The ACVD-HBOMDL technique encompasses data scaling, HBO based feature subset selection, DLMNN classification, and BO based parameter optimization. Fig. 1 represents the workflow of ACVD-HBOMDL algorithm.



3.1 Data Scaling Process

At the initial stage, the ACVD-HBOMDL technique applies min-max scaler to preprocess the medical data. Min-Max normalization performs a linear conversion on the original data. $MinA$ and $MaxA$ denote the minimal and maximal values of the attribute A . Min-Max normalization map a value of A to v 'in the range by calculating new_MinA , new_MaxA . The threshold value for Min-Max ranges from zero to one $V' = \frac{V - MinA}{MaxA - MinA} (new_MaxA - new_MinA) + new_MinA$

3.2 Algorithmic Design of HBO Based Fs Technique

In this work, the HBO-FS technique is applied to electing the features. In HBO, the honey badger's (HB) dynamic search performance with honey-search and digging strategies was separated from the exploitation and exploration stages. The honeyguide bird realizes hives that could not obtain honey or consequently, HB defines its meals by digging and finding mouse nest, and it can be mathematically expressed as follows: The presented technique begins by initializing the amount of HBs based on the population count (N_s) and the succeeding location:

$$Y_j = lb_j + r1 \times ub_j - lb_j$$

Y_j represent the HB location, lb_j and ub_j shows the lower and upper boundaries of every location from the searching space, and $r1$ indicates the arbitrary integer amongst $[0, 1]$. The intensity (Int) has specified that it would be proportionate to concentrate, prey strength, and the length amongst the j th HB, along with the prey. The prey moves faster if the smell strength is different and stronger. The succeeding formula is exploited to compute the intensity:

$$Int_j = r2 \times SS \quad 4\pi d^2 j SS = (Y_j - Y_{j+1})^2$$

$$dj = Y_{prey} - Y_j$$

3.3 CVD Classification Using Optimal DLMNN Model

For CVD classification, the DLMNN classifier is used in this study. All the inputs are offered to distinct nodes prevalent from the input area of the classifications. The weighted signify arbitrarily assigned value and are joined to all the inputs. The succeeding one is Hidden Layer (HL). The node in this HL labelled as hidden node executes the purpose of summing the input product values and the weight vector of each input node that is related to it. In DL-NN, the weighted value can be optimised utilizing the KHOA and DLMNN. Random weighted value provides additional backpropagation (BP) technique to accomplish the resultant in optimizer approach. The activation function has been implemented and this layer outcome was transported to consecutive layers. These weights are maximum effect on the classification outcome. The algorithmic stages in the DLMNN technique are as follows.

Step1: Allot the score value of chosen features and their matching weighted as: $E_i = \{E_1, E_2, E_3 \dots E_n\}$, $W_i = \{W_1, W_2, W_3 \dots W_n\}$,

Step2: Increase the input with weight vector which can be arbitrarily selected and afterwards



it should be completely summed up as

$$R = \sum_{i=1}^n E_i W_i.$$

whereas: W_i —Weight values, R -Summed value, E_i —Input entropy value.

Step3: Calculate the activation function (AF) by means of

$$AF_i = f \sum_{i=1}^n E_i W_i !$$

Step4: Estimate the outcome of HLs as

$$Y_i = A_i + \sum G_i W_i.$$

whereas: G_i —Values which are altered by the application of AFs, A_i —Bias value, W_i Weighted amongst the input as well as HLs.

Step5: Re-execute the above 3 steps to all the layers of DLMNN. Eventually, evaluation of the resulting unit with add up each input data weighted for attaining the resultant layer neuron value,

$$O_{ui} = A_i + \sum P_i W_i.$$

Step6: Contrast the network effect with the aim value. The variance between these 2 values procedures the error data. The value has been mathematically expressed as: $E_r = T_{ai} - O_{ui}$.

Step7: At the moment, the resulting unit weights the aim value. Consequently, the comparative error was defined. In accordance with that error, a value δ_i can be measured and it should be exploited to distribute the error at resulting back to every unit on the network: $\delta_i = E_r * f'(O_{ui})$.

Step8: The weighted correction was measured by exploiting the BP technique. This relation was offered as: $\tau_{vci} = \beta \delta_j E_j$.

Finally, the hyperparameters can be optimally chosen by BO algorithm.

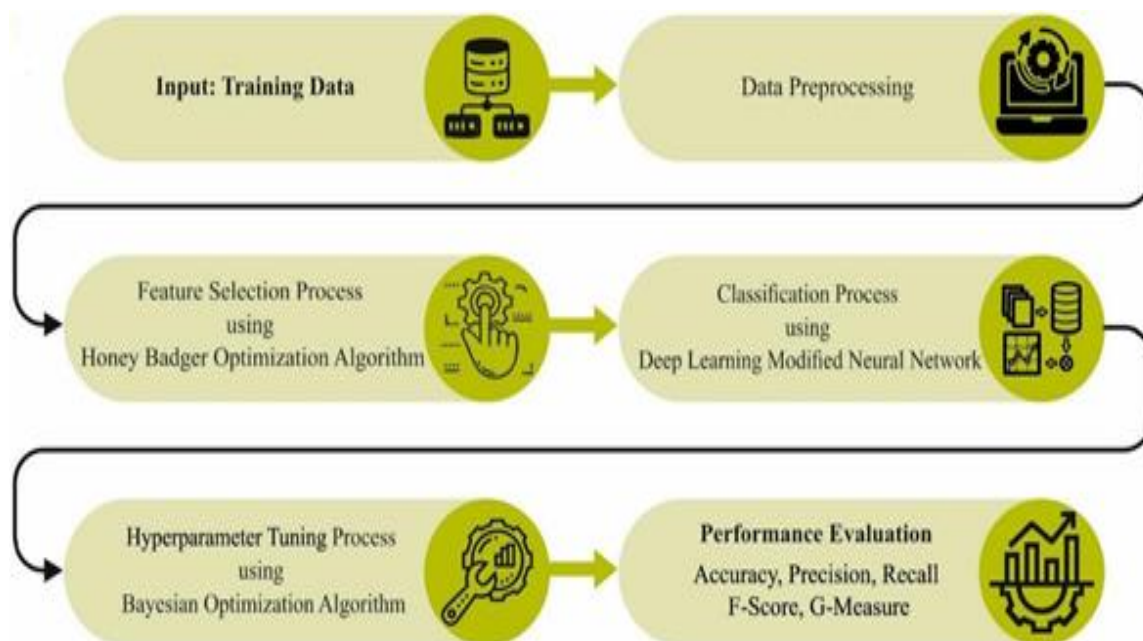


Figure 1: Proposed block diagram

3.4 Mathematical Expressions and Symbols

Bayes Theorem computes the conditional probability of event A assumed another event B, $P(A|B)$ as follows $P(A|B) = P(B|A) * P(A) / P(B)$.

An upgraded version of Bayes theorem is expressed as: Posterior = Likelihood * Prio.

The possibility function was determined as the probability of observing the data provided the OF is $P(F|S) = P(S|F) * P(S)$.

The acquisition task was executed in this case utilizing the Expected Improvement technique: $C(x) = E \max F(x) - P(x^+), 0]$

4 Results and Discussion

The CVD classification output of the ACVD-HBOMDL technique can be tested using the dataset from Kaggle repository. The dataset is taken from three other research datasets used in different research papers such as Cleveland, Hungarian, Switzerland, Long Beach VA, and Statlog dataset. The HBO algorithm has chosen the following features: chest pain type, resting bp s, cholesterol, resting ecg, and max heart rate.



Preparation of Figures and Tables

Class	Accuracy	Precision	Recall	F-Score	G-Measure
Epoch-500					
Normal	93.05	95.08	93.05	94.05	94.06
Disease	95.71	93.92	95.71	94.80	94.81
Average	94.38	94.50	94.38	94.43	94.43
Epoch-1000					
Normal	98.04	97.52	98.04	97.78	97.78
Disease	97.77	98.24	97.77	98.01	98.01
Average	97.91	97.88	97.91	97.89	97.89
Epoch-1500					
Normal	87.34	97.42	87.34	92.11	92.24
Disease	97.93	89.67	97.93	93.62	93.71
Average	92.64	93.54	92.64	92.86	92.98
Epoch-2000					
Normal	94.47	94.98	94.47	94.73	94.73
Disease	95.55	95.09	95.55	95.32	95.32
Average	95.01	95.04	95.01	95.02	95.02
Epoch-2500					
Normal	93.40	95.45	93.40	94.41	94.42
Disease	96.03	94.23	96.03	95.12	95.12
Average	94.72	94.84	94.72	94.77	94.77
Epoch-3000					
Normal	98.93	99.82	98.93	99.37	99.37
Disease	99.84	99.05	99.84	99.45	99.45
Average	99.39	99.44	99.39	99.41	99.41

Figure 2: CVD classifier outcome of ACVD-HBOMDL technique with distinct epochs.

5 Conclusions

In this study, we have designed a novel ACVD-HBOMDL technique for accurate CVD diagnosis process using FS and hyperparameter tuning processes. Initially, the ACVD-HBOMDL technique applies min-max scaler to preprocess the medical data. To elect an optimal subset of features, the HBO algorithm is used in this work. For CVD classification, the DLMNN classifier is used and its hyperparameters can be optimally chosen by Bayesian optimization. The experimental results of the ACVD-HBOMDL technique can be tested on benchmark medical dataset and the obtained results demonstrate the significant outcomes of the ACVD-HBOMDL technique over other existing techniques with improved accuracy of 99.39%. In the future, data clustering and outlier detection processes can be involved to boost the classification performance of the ACVD-HBOMDL technique.



References

1. Wan Aezwani Wan Abu Bakar; Nur Laila Najwa Bt Josdi; Mustafa B. Man; Muhammad Amierusyahmi B. Zuhairi, Year: 2023, "A review: heart disease prediction in machine learning & deep learning." In 2023 19th IEEE international colloquium on signal processing & its applications (CSPA), pp. 150-155.
2. M. Siddhartha, Year: 2023, Heart Disease Dataset (Comprehensive) Statlog + Cleveland + Hungary Dataset. [Online]. Available: <https://www.kaggle.com/datasets/sid321axn/heart-statlog-Cleveland-Hungary-final>.
3. P. K. Shrivastava; M. Sharma; P. Sharma; A. Kumar, Year: 2023 "HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms, " Meas., Sensors, Vol: 10.
4. B. Kolukisa; B. Bakir-Gungor, Year: 2023 "Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis, " Comput. Standards Interfaces, Vol: 84, no: 103706.
5. M. Khaleel Malnajjar; Abu-Naser; S. Samy, Year: 2022 "Heart Sounds Analysis and Classification for Cardiovascular Diseases Diagnosis Using Deep Learning. [Online]. Available: <http://dspace.alazhar.edu.ps/xmlui/handle/123456789/3534> no: 100657.
6. M. A. Qureshi; K. N. Qureshi; G. Jeon; F. Piccialli, Year: 2022, "Deep learningbased ambient assisted living for self-management of cardiovascular conditions, " Neural Comput. Appl., Vol: 34, no: 13, pp. 10449-10467.
7. Ivo Sérgio Guimarães Brites; Lídia Martins da Silva; Jorge Luis Victória Barbosa; Sandro José Rigo; Sérgio Duarte Correia; Valderi Reis Quietinho Leithardt, Year: 2022 "Machine learning and IoT applied to cardiovascular diseases identification through heart sounds: A literature review" in Proc. Int. Conf. Inf. Technol. Syst. Cham, Switzerland: Springer, pp. 356-388.
8. Umarani Nagavelli; Debabrata Samanta; Partha Chakraborty, Year: 2022, "Machine learning technology-based heart diseasedetection models, " J. Healthcare Eng., pp. 1-9.
9. Fatma A Hashim; Essam H. Houssein; Kashif Hussain; Mai S. Mabrouk; Walid Al-Atabany, Year: 2022, "Honey badger algorithm: New metaheuristic algorithm for solving optimization problems, " Math. Comput. Simul. Vol: 192, pp. 84-110.
10. Ghada Atteia; Amel A; Alhussan,; Nagwan Abdel Samee, Year: 2022, "BO-ALLCNN: Bayesianbased optimized CNN for acutelymphoblastic leukemia detection in microscopic blood smear images, " Sensors, Vol: 22, no: 15, pp. 5520.