IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IMPROVING HEART DISEASE CLASSIFICATION WITH OPTIMIZATION AND CNN MODEL

[1]Dr. Manish Shrivastava, [2]Mr. Pravin M. Tambe

[1]Professor, [2]Research Scholar

[12]Dept. of Computer Science & Engineering, VGU, Jaipur, India,

Abstract: Heart disease is a major global health concern, accounting for a significant number of deaths annually. Early detection and accurate prediction of heart disease are crucial for effective management and treatment. Machine learning techniques, particularly classification methods, have shown promise in analyzing clinical data for heart disease detection. However, existing approaches have limitations in terms of data preprocessing, feature selection, and model optimization. In this paper, we propose a novel approach that combines hybrid brave-hunting optimization with a support vector machine (SVM)-coupled deep convolutional neural network (CNN) model for heart disease prediction. The hybrid optimization method enhances feature selection and model performance, while the deep CNN model leverages the power of neural networks for capturing complex relationships in the data. Our experimental results demonstrate the effectiveness of the proposed approach in accurately predicting heart disease. This research contributes to the development of a cost-effective and efficient decision support system for heart disease detection, potentially aiding in early intervention and improved patient outcomes.

***Index Terms* - Heart disease prediction, Machine learning techniques, Deep Convolutional Neural Network (CNN)**

I. INTRODUCTION

The heart is the most important part of the human body which is responsible for pumping oxygen-rich blood to other body parts through a network of arteries and veins. Any type of disorder that affects our heart is heart disease [9] [5]. Heart disease is a grave disease that influences the heart's functionality and gives rise to complications such as infection of the coronary artery and diminished blood vessel function [10] [6]. Heart disease patients do not feel sick until the very last stage of the disease, and then it is too late because the damages have become irretrievable [8]. The term heart disease also referred to as cardiac disease, incorporates various conditions, the symptoms of which include high blood pressure, arrhythmia, stroke, and heart attack [4]. From the recent statistics reported by World Health Organization (WHO), about 20.5 million people die every year due to heart disease, which is approximately 31.5% of all deaths globally. It is also estimated that the number of annual deaths will rise to 24.2 million by 2030 [8]. Among the total deaths, one-third occur with persons below the age of 70 [11]. Sex, smoking, age, family history, poor diet, cholesterol, physical inactivity, high blood pressure, weightiness, and alcohol use are the key risk influences for heart disease [1]. One of the most challenging medical data is data related to heart diseases which has drawn many researchers' attention. Multiple machine-learning methods were examined for the prediction of heart diseases [2].

Machine learning techniques can be used to design a decision support system to detect heart disease through clinical data easily and cost-effective manner [1]. One of the powerful machine-learning techniques for prediction is classification [12]. Classification is a supervised machine learning method that is effective at identifying the disease when trained using appropriate data [8]. A complete model consisting of a modified differential evolution (DE) method, fuzzy analytical hierarchies process (AHP), and a feed-forward neural network (FNN) [3]. Firstly predictor attributes were selected, and then data cleaning was performed to deal with incomplete and inaccurate parts of the data [5].To choose the most important features, the modified differential evolution method was adopted. Furthermore, a reduced set of attributes was fed into an optimized model for a fuzzy AHP with FNN to predict heart disease [13] [4]. A k-means method with particle swamp was developed for detecting hazard factors in coronary heart disease treatment (CAD) [2]. The extracted data are classified using multilayer perceptron (MLP), multinomial logistic regression (MLR), and algorithms of phase rule, as well as C4.5. It was claimed that the results demonstrated the appropriate accuracy [1]. But existing works stated above are not sufficient to be implemented for heart disease prediction because of the limitations [7].

One of the most critical and challenging issues in modern medicine is accurately predicting the onset of heart disease [8]. ML-based analyses frequently look for nonlinear relationships among tens or hundreds of thousands of different variables [6]. For these methods to be most effective, a massive proportion of data for training is necessary. When the information is plentiful, the labels are complicated, time-consuming, or expensive to obtain [14] [3]. During pre-processing, most researchers replaced the missing values, either by using the mean value or the majority mark of that attribute, to make sure the dataset was comprehensive [7]. The missing valued instances were removed which affected the accuracy of pre-processing. Feature selection is a challenging task due to the large exploration space. It grows exponentially according to the number of features available in the dataset [8].In the Modified Deep Convolutional Neural Network (MDCNN) with cuttlefish optimization, the outcomes corroborated that the suggested structure obtained high-level results as it diminished the overhead of the access time. The structure has the highest key generation time when the verification time and transfer time are considered [6]. Moreover, there are open issues, including discretizing the numeric values of features, categorization, and binning levels using advanced metaheuristic algorithms for fine-tuning the predictive models' parameters and using enhancement classification algorithms rather than the label ranking classifier [3].

II. Literature review

The review of the various existing methods is listed below with their merits and demerits.

Table 2.1: Literature Review

Sr.No	Author	Methods	Merits	Demerits
1	JamshidPirgaziet *al* [1]	Ensemble machine learning classifier	The representative and maturity of the algorithm were high	Low accuracy level
2	Ch. Anwar ul Hassan *et al*. [2]	Ensemble learning-based convex optimization	The accuracy of the performance is high	Time complexity is high
3	Ibrahim M. El-Hasnonyet *al*. [3]	Ensemble learning-based hyperparameter optimization	High F1 score rate in the learned model	Data loss occurred discrediting the numeric values of features
4	Mohammad Ayoub Khan and Fahad Algarni [4]	Modified salp swarm optimization (MSSO) and an adaptive neuro-fuzzy inference system (ANFIS)	Feature selection achieved the highest fitness values for all iterations	Low effectiveness in optimization technique
5	Pooja Raniet *al*. [5]	SMOTE-based Genetic Algorithm (GA)	Missing data were balanced in the data set based on SMOTE	The severity of heart disease cannot be diagnosed with this system
6	Mohammad Ayoub Khan [6]	Modified Deep Convolutional Neural Network (MDCNN)	Normal and abnormal heart functioning is	sometimes it gets trapped in a local optimum

		with cuttlefish optimization	diagnosed by using the MDCNN	
7	Prerna Sharmaet *al* [7]	Modified Artificial Plant Optimization (MAPO) algorithm	MAPO reduces the dimensionality to the most significant information with comparable accuracies	High computational time
8	Karna Vishnu Vardhana Reddyet *al* [8]	Instance-based learner (IBk) classifier-based sequential minimal optimization (SMO)	A Low-level MAE rate was produced by SMO	Low classifier's predictive performance

III. Challenges

The challenges that arise during the various pieces of research are considered as follows,

- The crow search algorithm has shown its capability to provide the optimal solution. However, this search strategy does not guarantee convergence due to the poor exploration of the search space. The search strategy of the crow algorithm poses major challenges when faced with extremely multimodal formulations in heart disease prediction [4].
- In ML, a learning model has led to good results on the training data but fails when applied to never-before-seen data, especially in settings where there are hundreds or thousands of covariates. The model is overfitting the data it was trained on [3].
- CNN network is not initially trained; thus, pre-trained network weights aid to solve more issues concerning feature extraction or configuration. Very deep networks are complex to be trained in. More complex models require more time for training using hundreds of systems with expensive CPUs [2].
- During pre-processing, most researchers replaced the missing values, either by using the mean value or the majority mark of that attribute, to make sure the dataset was comprehensive [8].
- The missing valued instances were removed which affected the accuracy of pre-processing. Data imputation is a challenging task due to the large exploration space. It grows exponentially according to the number of data available in the dataset [8].

IV. PROPOSED METHODOLOGY

The main aim of the research is to design and develop a heart disease-related model based on sensor nodes using a optimized SVM coupled deep CNN classifier. The sensor nodes collected the data of the patient from the base station. The main stages involved in the research are; Pre-processing, and classification. Initially, the input of the patient's data will be collected from sensor nodes. Then it undergoes the process of the pre-processing stage, in pre-processing stage proceeds, where the quality of the data will be assessed and the irrelevant data indication will be removed followed by the removal of the missing data. Further, the pre-processed data was passed through the proposed optimized SVM coupled deep CNN classifier, it categorizes and labels groups of vectors within a data then its hyperparameters are well-tuned by hybridizing brave-hunting based optimization which utilizes the characteristics of both Lion Optimization[15]and Coyote Optimization[16].In the last instance, the designed and developed sensor-based model shows if any heart-related diseases were present in the patient data in terms of normal or diseased using test data by the proposed optimization-based SVM coupled deep CNN model. The performance of the criterion is utilizing the design and developed model accuracy, precision, recall, and F measure. The implementation will be done in the python tool. The proposed model block diagram is illustrated in figure 1.

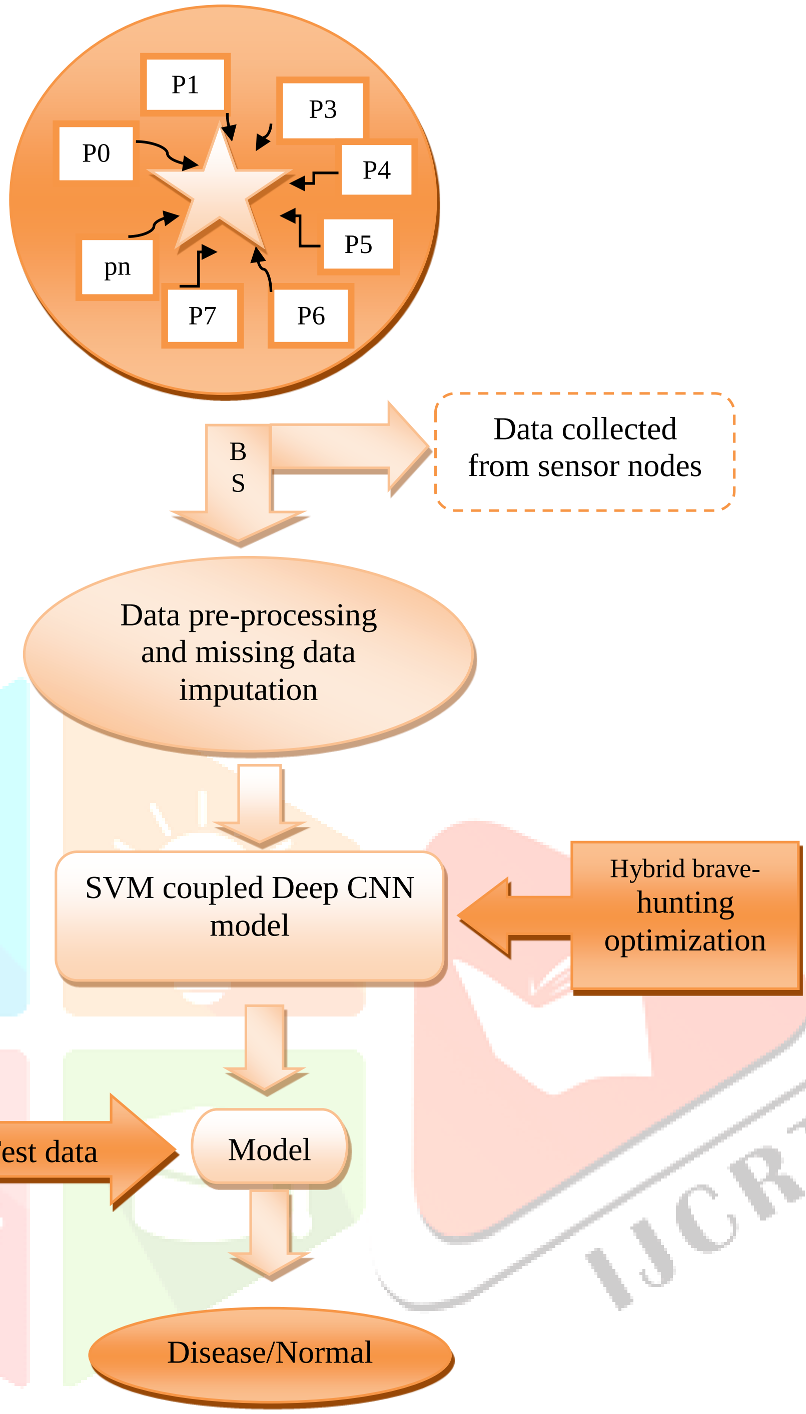

Figure 4.1: Block Diagram Representation of Developed Heart Disease-related Model

V. OBJECTIVES

The major objectives involved in this research are as follows,

- To develop the method for the prediction of the heart disease sensor model.
- The proposed method performance is enhanced by developing the Hybrid Brave-Hunting optimization, which well-tunes the hyper parameters present in the classifier.
- The developed method's performance is compared with the various existing reviewed methods.

VI. CONCLUSION AND FUTURE SCOPE

In conclusion, heart disease is a significant health concern with a high mortality rate worldwide. Machine learning techniques, particularly classification models, have been explored for heart disease prediction using clinical data. However, existing approaches have limitations in terms of feature selection, missing value handling, and the need for large labeled datasets. Future research should focus on addressing these challenges and improving the accuracy and efficiency of heart disease prediction models. Possible avenues for future work include exploring advanced metaheuristic algorithms for feature selection and parameter tuning, developing enhancement classification algorithms, and investigating novel approaches such as deep convolutional neural networks with optimized architectures. Additionally, efforts should be made to collect comprehensive and diverse datasets to train and validate these models effectively. By overcoming these challenges and advancing the field of heart disease prediction, we can contribute to earlier detection, timely interventions, and improved patient outcomes.

REFERENCES

[1] Dr. Manish Shrivastava and Mr. Pravin M. Tambe, “Develop a Sensor Based Model Using Machine Learning for Heart Related Disease,” IJIRCCE, Volume 11, Special Issue 2, pp.510-512, 2023.

[2] Pirgazi, J., GhanbariSorkhi, A. and IranpourMobarkeh, M., “An Accurate Heart Disease Prognosis Using Machine Intelligence and IoMT,” Wireless Communications and Mobile Computing, 2022.

[3] Hassan, C.A.U., Iqbal, J., Irfan, R., Hussain, S., Algarni, A.D., Bukhari, S.S.H., Alturki, N. and Ullah, S.S., “Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers,” Sensors, vol. 22, no.19, p.7227, 2022.

[4] El-Hasnony, I.M., Elzeki, O.M., Alshehri, A. and Salem, H., “Multi-label active learning-based machine learning model for heart disease prediction,” Sensors, vol.22, no.3, pp.1184, 2022.

[5] Khan, M.A., and Algarni, F., “A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS,” IEEE Access, vol.8, pp.122259-122269, 2020.

[6] Rani, P., Kumar, R., Ahmed, N.M. and Jain, A., “A decision support system for heart disease prediction based upon machine learning,” Journal of Reliable Intelligent Environments, vol. 7, no.3, pp.263-275, 2021.

[7] Khan, M.A., “An IoT framework for heart disease prediction based on MDCNN classifier,” IEEE Access, vol.8, pp.34717-34727, 2020.

[8] Sharma, P., Choudhary, K., Gupta, K., Chawla, R., Gupta, D. and Sharma, A., “Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning,” Artificial intelligence in medicine, vol. 102, pp. 101752, 2020.

[9] Reddy, K.V.V., Elamvazuthi, I., Aziz, A.A., Paramasivam, S., Chua, H.N. and Pranavanand, S., “Heart disease risk prediction using machine learning classifiers with attribute evaluators,” Applied Sciences, vol.11, no.18, pp.8352, 2021.

[10] Kumar R, Rani P “Comparative analysis of decision support system for heart disease,” Adv Math Sci J, vol. 9, no.6, pp.3349–3356, 2020.

[11] Ali, F., El-Sappagh, S., Islam, S.R., Kwak, D., Ali, A., Imran, M. and Kwak, K.S., “A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion,” Information Fusion, vol. 63, pp.208-222, 2020.

[12] Szczepańska, E., Białek-Dratwa, A., Janota, B. and Kowalski, O., “Dietary Therapy in Prevention of Cardiovascular Disease (CVD),” A Review of the Latest Approaches to Nutrition in CVD, Nutrients, vol. 14, no.13, pp.2649, 2022.

[13] Tomov, N.S. and Tomov, S., “On deep neural networks for detecting heart disease,” arXiv preprint arXiv, pp.1808.07168, 2018.

[14] Dutta, A., Batabyal, T., Basu, M. and Acton, S.T., “An efficient convolutional neural network for coronary heart disease prediction,” Expert Systems with Applications, vol.159, pp.113408, 2020.

[15] Ponikowski P, Anker SD, AlHabib KF, Cowie MR, Force TL, Hu S, Jaarsma T, Krum H, Rastogi V, Rohde LE, Samal UC, Shimokawa H, Siswanto BB, Sliwa K, Filippatos G “Heart failure: preventing disease and death worldwide,” ESC Heart Fail vol.1, no.1, pp.4–25, 2014.

[16] Pierezan, J. and Coelho, L.D.S., “Coyote optimization algorithm: a new metaheuristic for global optimization problems,” In 2018 IEEE congress on evolutionary computation (CEC), pp. 1-8, 2018, July.

[17] Ali, E.S., Abd Elazim, S.M. and Abdelaziz, A.Y., “Ant Lion Optimization Algorithm for optimal location and sizing of renewable distributed generations,” Renewable Energy, vol.101, pp.1311-1324, 2017.