ABUAD Journal of Engineering Research and Development (AJERD)
ISSN (online): 2645-2685; ISSN (print): 2756-6811

Volume 7, Issue 1, 35-40

Development and Performance Evaluation of a Heart Disease Prediction Model Using Convolutional Neural Network

Adebimpe ESAN[1], Juwon AKINGBADE[1], Adetunji OMONIJO[2], Adedayo SOBOWALE[1], Tomilayo ADEBIYI[1]

[1]*Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti State, Nigeria*
adebimpe.esan@fuoye.edu.ng/juwonemmanuel22@gmail.com/sobowaleadedayo@gmail.com/
tomilayo.oguntuyi@fuoye.edu.ng

[2]*Department of Family Medicine, Federal Teaching Hospital, Ido-Ekiti, Nigeria*
sundayomonijo97@gmail.com

Corresponding Author: juwonemmanuel22@gmail.com, +2349032869229
Date Submitted: 04/01/2024
Date Accepted: 22/03/2024
Date Published: 01/04/2024

Abstract: *Heart disease is a leading cause of mortality globally and its prevalence is increasing year after year. Recent statistics from the World Health Organization show that about 17.9 million individuals are embattled with heart diseases annually and people under the age of 70 account for one-third of these deaths. Hence, there is need to intensify research on early heart disease prediction and artificial intelligence-based heart disease prediction systems. Previous heart disease prediction systems using machine learning techniques are unable to manage large amount of data, resulting in poor prediction accuracy. Hence, this research employs Convolutional Neural Networks, a deep learning approach for prediction of heart diseases. The dataset for training and testing the model was obtained from a government owned hospital in Nigeria and Kaggle. The resulting system was evaluated using precision, recall, f1-score and accuracy metrics. The results obtained are: 0.94, 0.95, 0.95 and 0.95 for precision, recall, f1-score and accuracy respectively. This show that the CNN-based model responded very well to the prediction of heart diseases for both negative and positive classes. The results obtained were also compared to some selected machine-learning models like Random Forest, Naïve Bayes, KNN and Logistic Regression and results show that the developed model achieved a significant improvement over the methods considered. Therefore, convolutional neural network is more suitable for heart disease prediction than some state-of-the-art machine-learning models. The contribution to knowledge of this research is the use of Afrocentric dataset for heart disease prediction. Future research should consider increasing the data size for model training to achieve improved accuracy.*

Keywords: *Model Accuracy, Dataset, Convolutional Neural Network, Heart Diseases, Prediction, Machine Learning.*

1. INTRODUCTION

Heart disease is a prevalent disease that affects millions of people worldwide and early diagnosis of the disease is crucial in preventing fatalities. According to the World Health Organization, cardiovascular diseases are the leading cause of death worldwide, with 17.9 million deaths annually. Factors such as high cholesterol, obesity, and hypertension increase the risk of heart disease, making early diagnosis difficult. However, with the advancement of technology, researchers have access to a vast amount of patient records and research data, which can be used to detect heart disease at an early stage. Machine learning and artificial intelligence have a significant role in the medical field and have proven to be powerful tools in the early diagnosis of heart disease, and research in this field is rapidly growing. Various machine learning techniques and deep learning models can be used to diagnose illnesses and predict outcomes. These methods make it easy to conduct comprehensive genetic data analysis, forecast knowledge pandemics, and study medical records more thoroughly. Researchers have used machine learning algorithms such as Decision trees [1], Support vector machines [2], and Naive bayes to predict heart disease [3]. However, one of the common problems with machine learning is the high dimensionality of data, which can lead to overfitting and require a large amount of memory. To overcome this issue, researchers use techniques such as feature engineering and feature selection to improve classification and predictions [4]. In recent years, deep neural networks techniques have also been used in the field of medical prognosis [5]. These techniques provide more accurate results and are more robust to noise than traditional Machine Learning models. It is capable of handling large amounts of data and solving complex problems in an easy way [5]. The aim of this study is to apply convolutional neural network for heart disease prediction and the objectives are to design and implement a convolutional neural model for heart diseases prediction and evaluate the performance of the model.

The use of machine learning and deep learning algorithms to predict diseases has been the subject of numerous studies. These studies achieved some precision in prediction of diseases by utilizing a variety of optimization methods and machine learning models. Youness and Mohamed (2018) [6] worked on prediction and classification of heart diseases using machine learning algorithms optimized by particle swarm optimization and ant colony optimization but there is a need to optimize the performance of machine learning algorithms using advanced optimization techniques such as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) to have a better accuracy result. A deep neural network model was deployed with Talos optimization by Sumit and Mahesh (2020) [7] and results show that the proposed model using Talos optimization achieved better accuracy compared to other optimization techniques. Rohit *et al*. (2021) [8] used the combination of machine learning and deep learning algorithm to make predictions on heart diseases. Research show that machine learning performed better than deep learning algorithms because the dataset used was not large enough. The mean-based splitting technique, classification, and regression tree were used by Mienye *et al*. (2020) [9] to predict heart diseases and the study show that a significantly high accuracy was achieved on the two datasets used. Polaraju and Prasad (2017) [10] used Multiple Linear Regression Model to predict the chance of heart disease, research show that the accuracy of the Regression algorithm was better than other algorithms used in the study. Pandita *et al*. (2021) [11] proposed a prediction model using some machine learning algorithms and KNN was found to give better results. Akella and Akella (2021) [12] employed Artificial Neural networks for heart disease prediction and an accuracy of 93.03% was obtained. The research revealed that despite achieving high recall and accuracy, increased dataset can improve the model's performance. The study conducted by Miao *et al*. (2018) [13] utilized an improved random survival forest (iRSF) to develop a comprehensive risk model for predicting heart failure mortality and the approach achieved a high accuracy in discriminating between survivors and non-survivors and research revealed that iRSF is a useful approach in predicting heart failure mortality. To improve the accuracy of heart disease prediction systems, more complex and hybrid models were used by Ravindhar *et al.* (2019) [14]. The research employed Logistic Regression, Naive Bayes, Fuzzy KNN, K-Means Clustering and back propagation Neural-Network approaches. Results show that back propagation Neural Network approach achieved the highest accuracy score of 98.2%. Gavhane *et al.* (2018) [15] proposed Heart Disease Prediction System using Multi-layer perceptron algorithm, the resulting system was found efficient and accurate but research revealed that the model's performance can be improved by incorporating more relevant features. M5P, Random Tree, Reduced Error Pruning, and Random Forest ensemble method were employed by Yadav and Pal (2020) [16] for classification and Pearson Correlation, Recursive Features Elimination, and Lasso Regularization were used by the research for feature selection. Results show that the Lasso Regularization technique has the highest performance. To further improve the performance of heart disease prediction system, Ghosh *et al*. (2021) [17] combined five different datasets to obtain a larger and more reliable dataset. Feature selection was done using Relief and LASSO techniques and bagging and boosting were hybridized. Research revealed that the Relief feature selection algorithm, when used with RFBM machine learning algorithm, gives the highest accuracy of 99.05%. However, the limitation of these previous research is the use of foreign dataset for heart disease prediction, hence, this research employed indigenous dataset for heart disease prediction to make it suitable for use in Nigeria.

2. MATERIALS AND METHODOLOGY

This research developed a convolutional neural network (CNN) model for heart disease prediction. The dataset used in training the model was obtained from a government owned hospital in Nigeria and Kaggle. The dataset has fourteen attributes which represents the clinical features of each patients and the dataset contains 1024 data with 14 features. Pre-processing was carried out on the dataset to remove the noise and irrelevant data as well as fill missing values. The dataset was splitted into 80% for training and 20% for testing and thirteen of the attributes was used for training the convolutional neural network model while the fourteenth attribute was taken as an output parameter for the model to predict the output value which varies between 0 and 1 where 0 represents the absence of disease while 1 represent the presence of disease. Confusion matrix was used to evaluate the developed model using: precision, accuracy, recall and f1-score metrics. The architecture of the developed model is shown in Figure 1 and the 14 attributes used in the dataset is shown in Table 1.

2.1 Data Pre-processing

The acquired dataset was pre-processed using the following techniques:

i. Data Cleaning: Raw dataset collected from Nigerian hospitals were cleaned to remove noise or outliers that are of no relevance for the developed system.
ii. Removal of Stopwords and Punctuations: Stopwords words like punctuation marks, special characters, and emoji in the dataset were removed to reduce the data size and improve the system's performance accuracy and reduce computational time.
iii. Tokenization: Tokenization was carried out on the dataset to break the unstructured textual data into discretised elements.
iv. Vectorization: Textual data was converted into vectors using one hot representation to enable models understand the input data.

2.2 Design of the CNN-based Model for Heart Disease Prediction

The structure of convolutional neural network used for prediction of heart disease consists of a few pairs of convolution and max-pooling layers. The Pooling layer follows the convolutional layer and it was applied along frequency domain. The max-pooling layer makes the architecture more tolerable to minor differences in positions of the object's parts and leads to faster convergence. Finally, the fully connected layers combine inputs from all the positions into a 1-D feature vector and the Softmax layer is the final output layer that performed multi-class classification which takes as input a number of scores

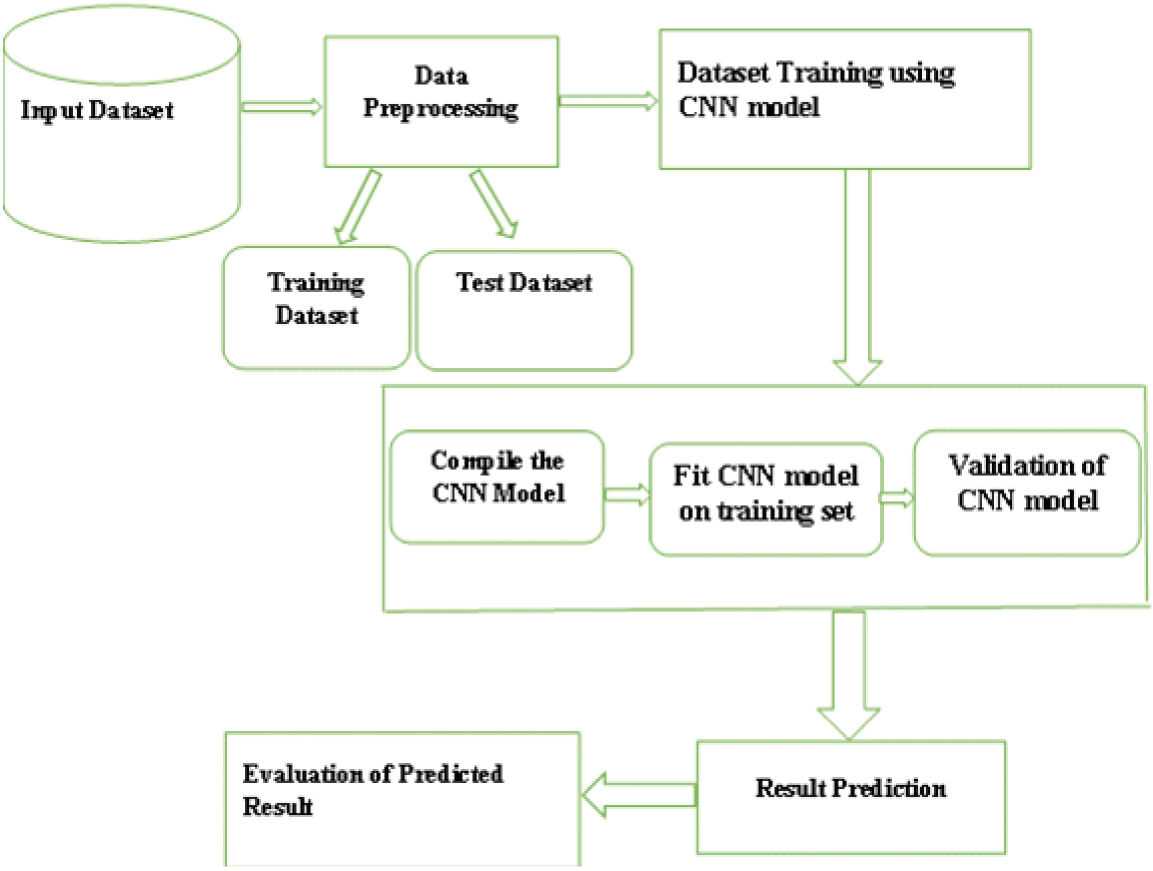

Figure 1: Architecture of the CNN based heart disease predictor

Table 1: Description of 14 attributes obtained from the dataset

S/N	Clinical Features	Description
0	Age	Age
1	Sex	Gender
2	Cp	Chest Pain type
3	Trestbps	Resting blood pressure
4	Chol	Serum Cholesterol
5	Fbs	Fasting blood sugar
6	Restecg	Resting electrocardiographic results
7	Thalach	Maximum Heart rate achieved
8	Exang	Exercise-induced angima
9	Oldpeak	ST depression induced by exercise
10	Slope	The slope of the peak exercise ST
11	Ca	Number of major vessels (0-3)
12	Thal	3=normal; 6=fixed defect;
13	Target	Finding for the existence of heart disease status

values (), and squashes them into values in the range between 0, and 1. The convolution process can be expressed in equation 1, in which *Xij* is the *ith* and *jth* matrix in the row and column direction, respectively. From the equation, K is the convolution kernel, and γ is the output matrix.

$$\boldsymbol{\gamma}(\mathbf{i},\mathbf{j}) = \mathbf{X}(\mathbf{i},\mathbf{j}) \ * \ \mathbf{K} \tag{1}$$

The structure principle of the fully connected layer was expressed using Equation 2

$$\boldsymbol{o} = \boldsymbol{f}\left(\boldsymbol{W}_{\boldsymbol{f}}^{\boldsymbol{T}} \ \boldsymbol{X} \ + \boldsymbol{b}\right) \tag{2}$$

In Equation 2, $\boldsymbol{o}$ is the vector composed of output values, $\boldsymbol{x}$ is the vector composed of input values, $\boldsymbol{W}_{\boldsymbol{f}}^{\boldsymbol{T}}$ is the vector composed of weight values, $\boldsymbol{b}$ is the vector composed of threshold values, and f is the activation function. The sequence diagram which shows the activities involved in the design of the convolutional neural network model of the developed heart disease prediction system is shown in Figure 2.

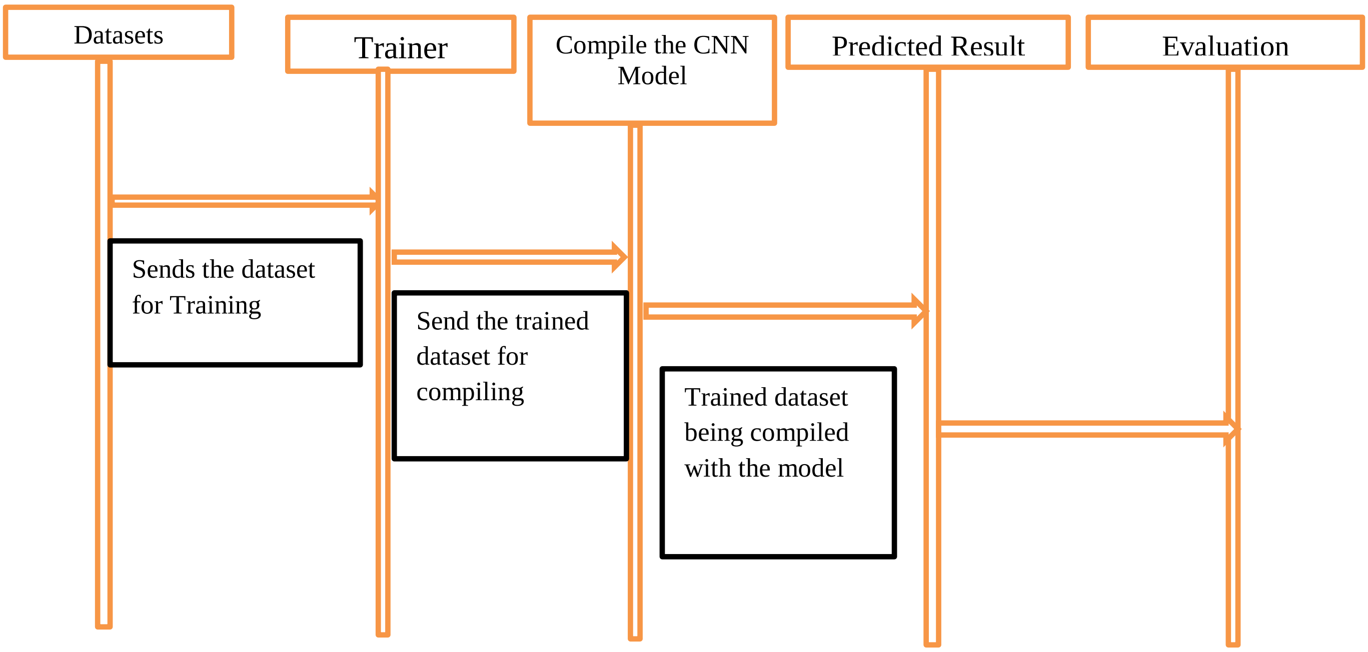

Figure 2: Sequence diagram of the convolutional neural network model

3. RESULTS AND DISCUSSION

Results from the evaluation of the developed model show that the model correctly predicted 126 instances as positive (True Positive) and incorrectly predicted 6 datasets as positive (False Positive). 118 datasets were correctly predicted as Negative (True Negative) while 7 datasets were misclassified as Negative (False Negative). The classification report was calculated from the values of the True Positive, False Negative, False Positive and True negative. 0 represents the Negative class and 1 represents the Positive class. Table 2 shows the precision, recall, f1-score and support scores of each class label. The macro-averaged F1 score was computed using the arithmetic mean of all the per-class F1 scores while the weighted-average F1 score was calculated by taking the mean of all per-class F1 scores when considering each class's support**.** The accuracy of the developed CNN-based model is 95%, the precision is 94% and f1-score is 95%. This result shows a great improvement to previous machine learning model using support vector machine [2].

Table 2: Result from classification of the developed CNN model

Class Label	**Precision**	**Recall**	**f1-score**	**Support**
0	0.95	0.95	0.95	132
1	0.95	0.94	0.95	125

The macro average for the precision, recall, f1-score and support are 0.95, 0.95, 0.95 and 257 respectively and weighted average of 0.95, 0.95, 0.95 and 257 respectively as shown in Table 3. This shows that the developed model responded to the prediction of both negative and positive classes very well.

Table 3: Results from the Macro Average and Weighted Average

Accuracy				
Macro avg	0.95	0.95	0.95	257
Weighted avg	0.95	0.95	0.95	257

Figure 3 shows the graph of the accuracy of the model during training and testing. From the diagrams, the accuracy increases rapidly in the first two epochs, indicating that the network is learning fast. Afterward, the line flattens indicating that not too many epochs are required to train the model further.

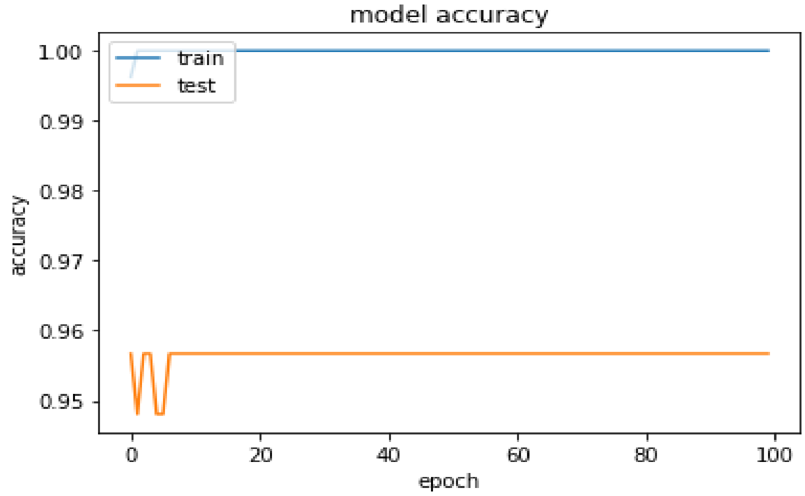

Figure 3: The graph of model accuracy during training and testing

Figure 4 depicts the graph of the model loss during training and testing. From the graph, the loss on the training set decreases rapidly for the first two epochs. For the test set, the loss does not decrease at the same rate as the training set, but remains almost flat for multiple epochs. This means the model is generalizing well to unseen data and this is a great improvement to heart disease prediction.

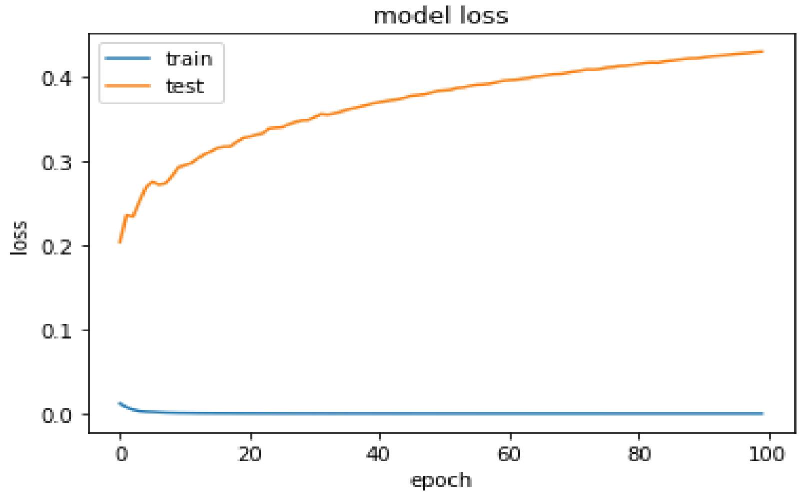

Figure 4: The graph of loss function during training and testing

The accuracy of the developed system was compared to some selected machine learning models like Random Forest, KNN, Naïve Bayes and Logistic Regression and an accuracy of 0.85, 0.83, 0.84 and 0.86 were achieved respectively as shown in Table 4. The results show that the developed CNN-based model performed better than the selected machine learning models. This result is in line with the research by [18] where two selected machine learning models; Random forest, and Naïve Bayes were employed for heart disease prediction and the accuracies achieved are almost similar to those obtained in this research.

Table 4: Comparison of some machine learning algorithms for prediction of heart diseases

Algorithm	**Accuracy**
Random Forest	0.85
KNN	0.83
Naïve Bayes	0.84
Logistic regression	0.86
Developed model (CNN)	95%

4. CONCLUSION AND RECOMMENDATION

This research developed a heart disease prediction system using Convolutional Neural Network (CNN) approach. The dataset for training the model was obtained from a Nigerian hospital and Kaggle. The developed system was evaluated using confusion matrix with an accuracy of 95%. The results obtained were compared to some selected machine-learning models and it was discovered that the developed CNN model performed better than the machine learning models used with an accuracy of up to 0.1. This shows that CNN outperforms conventional machine learning models for heart diseases prediction. However, future research can increase the data size for model training to achieve improved accuracy.

REFERENCES

[1] Reddy, V. K., Raju, K.P., Kumar, M. J, Sujatha, C.H., & Prakash, P. R. (2016). Prediction of Heart Diseases Using Decision Tree Approach. International Journal of Advanced Research in Computer Science and Software Engineering, 6(3), 530-532.

[2] Yamala, S. (2020). Prediction of Heart Diseases Using Support Vector Machine. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)***,** 8 (2), 126-135.

[3] Uma, N.D. (2018) Prediction System for Heart Disease using Naïve Bayes and particle swarm Optimization. Biomedical Research, 29 (12), 2646-2649.

[4] Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, 117, 47-58.

[5] Zhu, W., Xie, L., Han, J., & Guo, X. (2020). The Application of Deep Learning in Cancer Prognosis Prediction. *Cancers*, *12*(3), 603.

[6] Youness, K. & Mohamed, B. (2018). Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization. *International Journal of Intelligent Engineering and Systems*, *12(1), 242-252.*

[7] Sumit, S. & Mahesh, P. (2020). Heart Diseases Prediction using Deep Learning Neural Network Model. *International Journal of Innovative Technology and Exploring Engineering (IJITEE*), 9 (3), 2244-2248.

[8] Rohit, B., Aditya, K., Mohammad, S., Gaurav, D., Sagar, P., & Parneet, S. (2021). Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Comput Intell Neurosci, 2021, 1-11. doi: 10.1155/2021/8387680.

[9] Mienye, I. D., Sun, Y., & Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked*, 20(1), 1-5.

[10] Polaraju K. & Durga Prasad, D. (2017). Prediction of Heart Disease using Multiple Linear Regression Model. International Journal of Engineering Development and Research, 5(4), 1419-1425.

[11] Pandita, A., Vashisht, S., Tyagi, A. & Yadav, S. (2021). Prediction of Heart Diseases using Machine Learning Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, 9(5), 2422-2429.

[12] Akella, A, & Akella, S. (2021). Machine Learning Algorithms for Predicting Coronary Artery Diseases: Efforts toward an open source Solution. *Future Science OA,* 7(6), 698-706.

[13] Miao, F., Cai, Y. P., Zhang, Y. X., Fan, X. M., & Li, Y. (2018). Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest. *IEEE Access*, 6(1), 7244-7253.

[14] Ravindhar, N. V., Anand, H., Shanmugasundaram, R., & Winster, G. (2019). Intelligent Diagnosis of Cardiac Disease Prediction using Machine Learning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(11), 1417-1421.

[15] Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, K. (2018). Prediction of heart disease using machine learning. Second *International conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1275-1278, DOI:10.1109/ICECA.2018.8474922

[16] Yadav, D. C., & Pal, S. (2020). Prediction of heart disease using feature selection and random forest ensemble method. International Journal of Pharmaceutical Research, 12(4), 56-66.

[17] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, M., Ignatious, E., Shultana, S., & Beeravolu, A. (2021). Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques. *IEEE Access*, 9(1), 19321-19330.

[18] Shah, D., Patel, S.B., & Bharti, S.K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, 1(6). 1-10