Prediction of Cardiovascular Disease using DeepLearning Algorithm

S.Irin Sherly[1,*], J. Sandhiya[2], S. Priyanga[2], M. A. Sharon Victoriya[2], K. Sorna Ajantha[2]

[1]Department of Information Technology, Panimalar Engineering College, Chennai, India
[2]Department of Information Technology, Panimalar Institute of Technology, Chennai, India

Emails:irinkutty@gmail.com; sandhiyajayavelu23@gmail.com; priyankas30102002@gmail.com; victoriyasharon4@gmail.com ; ajantha6aju@gmail.com

Abstract

The leading cause of death, which affects millions of individuals globally is the cardiovascular disease. Heart problems are a major issue in health care, particularly in the field of cardiology. Due to a number of risk factors, including diabetes, high blood pressure, high cholesterol, an irregular pulse rate, obesity, and smoking, cardiac illness is difficult to detect. Due to these limitations, researchers are now using Data Mining and Deep Learning Algorithms to predict heart related disorders. The Cardio Vascular Disease (CVD) is as complicated as it sounds if left untreated. So, the early prediction of this could save millions of people from silent attacks, myocardial infarction etc. Many machine learning algorithms like Naïve Bayes, K-Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA) are used for cardiovascular disease prediction using text datasets and their efficiencies tend to differ. Generally, convolutional neural network (CNN) algorithm is mostly used for prediction using images. But our concept is to switch over this and predict heart disease using the CNN algorithm for Cleveland dataset which consists of numerical. In this dataset we consider 14 attributes and used K Nearest Neighbor and CNN algorithm. In terms of accuracy, CNN beats KNN, proving that deep learning algorithms may support decision-making and prediction-making based on vast volumes of data supplied by the healthcare sector.

Keywords: Myocardial infarction; Convolutional neural network; Cleveland dataset; Prediction

1. Introduction

Over 18 million people worldwide lose their lives to heart disease each year, according to a World Health Organization report. It is expected that this count reaches to 75 million by 2030. In world population, Americans are getting more heart related diseases when compared to other countries. India is also having highest death rate for cardiovascular diseases [1]. Most of the heart related diseases happen due to nicotine use, smoking, family history, obesity, cholesterol and unhealthy diet. People with cardiovascular diseases are more likely to get heart attacks. With the ever-rising heart disease rates, we need to find a system that can predict the disorder in an early stage without requiring much of the overpriced scans. It is necessary for the health professionals to diagnose the disease at initial stages to reduce the death rate. As a result, substantial amount of research need to be carried out by the researchers to help aid the medical physicians. To predict heart disease, a number of machine learning algorithms have already been applied. Some of the most popular machine learning methods for heart disease prediction are Support Vector Machine, Logistic Regression, Nave Bayes, Decision Tree, KNN, and Artificial Neural Networks. Other than machine learning techniques, feature selection algorithms must be employed to extract features from the data. To overcome this disadvantage, deep learning algorithms are used which can extract features automatically without human intervention. As a result, we propose building a system that can predict the vulnerability of a cardiac issue using a convolutional neural network using data from the UCI machine learning library.The effectiveness of the model is evaluated using the Cleveland heart disease data set. This system aims in building predictive analytics for the people of age group over 25-79. On this dataset, a comparison was conducted utilising machine learning and deep learning algorithms, with CNN proving to be more efficient in terms of

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

performance parameters. The suggested approach appears to be effective in predicting cardiovascular disease when compared to existing machine learning techniques.

2. Related Work

Oluwarotimi Williams Samuel et al. [3] have proposed HNCL and adaptive multi-layer networks which has a drawback of domino effect, in which the changes made to one component of the system may affect the operation of another. Sharma et al. [5] used the Modified Artificial Plant Optimization fingertip video dataset to estimate heart rate using an optimal feature selection and other machine learning algorithms (MAPO). Due to backdrop motion and a warped frame of view, the dataset features may present unique issues. Amin et al. [2] concluded that Decision Tree algorithm can be used effectively to predict heart disease, and age, chest pain type, maximum heart rate achieved, and ST depression induced by exercise are important features for predicting heart disease. The study can be useful for developing decision support systems for early diagnosis and treatment of heart disease. The authors collected a dataset of 303 patients with 14 attributes related to heart disease. The authors used various data mining techniques such as Decision Tree, Random Forest, Support Vector Machine (SVM), Naïve Bayes, and Logistic Regression to predict heart disease in the patients. Ali et al. [7] suggested a hybrid diagnostic system that uses a two-statistical model for feature refining and a deep neural network for classification. However, it did not take into account the temporal complexity of the proposed hybrid diagnostic method. The grid search method, which has limitations, is used in ANN and DNN models to estimate the ideal width of each hidden layer. Rajesh et al [16] developed a system using the Algorithm based on Bayes Theorem, which uses risk indicators are included to analyse data sets. Additionally, decision forests in combination with various other methods were utilised to predict cardiac disease. Data maybe over-fitting by using Naive Bayes and decision tree on the same dataset, the results maybe contradictory. The neural network categorization was used to predict the occurrence of cardiac ailments based on a survey of people's risk levels [11]. Due to the increasing number of attributes and combinations from it , an attempt to select the best attribute was done using feature selection methods like genetic algorithm, particle swarm optimization and recursive feature elimination[12]. Additionally, to the general classification a rewarding technique called the associative classification achieves high classification accuracy [15]. A prediction system along with other classifiers used a hybrid technique called Voting that is based on Naive Bayes and Logistic Regression with various combinations of features to predict heart disease [2]. Irin Sherly et al., [31] proposed ECG signal quality improvement using different types of filters. Irin Sherly et al., [32] proposed ensemble based prediction for issues related to heart using gradient boosting algorithm. They used four different datasets from various hospitals with 14 features and shown gradient boosting can perform well in all the four datasets. The existing systems tend to predict the disease only using few attributes of patients with less accuracy. It also generates more false positives which may lead to misdiagnosis and untimely treatment. Detection is not possible at an earlier stage in most of the cases as it turns out to be insignificant for which the patient doesn't consult a medical physician. To overcome this situation, people who suspect their heart's condition is getting worse can undergo our simple prediction analysis which is comparatively cost and time efficient and the results generated have optimal accuracies too.[19-20]

3. Methodology

A. Data Description

The Cleveland dataset contains information regarding heart disease diagnosis. The dataset was composed from the Cleveland Clinical Foundation and it's available at UCI Repository. The heart disease dataset has 303 unique records with 76 different attributes. Only 14 of these qualities are used in our investigation, as shown in Fig.1. This dataset contains six occurrences having missing values. It's recommended to impute these values before applying any other tasks. This dataset contains information regarding heart disease diagnosis and includes continuous, binomial and nominal features.

B. Proposed Method

Initially the 303 individual patient data of the Cleveland dataset is taken. Preprocessing is done as the first step missing values are handled by replacing it with mean or median of the correspoding attribute. The cheat pain type is a categorical value. This variable is encoded as numerical value. Scaling the attributes is done to have a mean of 0 and a standard deviation of 1. In segmentation, data is segmented into features and labels.

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

'Labels' are the final outcome that needs to be predicted and 'features' are attributes which are used to predict the labels. In splitting of data, one part of the data is used as a sample to fit the model and the other is used as a sample to provide evaluation on the final model.

age	gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63	1	1	145	233	1	2	150	0	2.3	3	0	6
67	1	4	160	286	0	2	108	1	1.5	2	3	3
67	1	4	120	229	0	2	129	1	2.6	2	2	7
37	1	3	130	250	0	0	187	0	3.5	3	0	3
41	0	2	130	204	0	2	172	0	1.4	1	0	3
56	1	2	120	236	0	0	178	0	0.8	1	0	3
62	0	4	140	268	0	2	160	0	3.6	3	2	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3
63	1	4	130	254	0	2	147	0	1.4	2	1	7
53	1	4	140	203	1	2	155	1	3.1	3	0	7
57	1	4	140	192	0	0	148	0	0.4	2	0	6
56	0	2	140	294	0	2	153	0	1.3	2	0	3
56	1	3	130	256	1	2	142	1	0.6	2	1	6
44	1	2	120	263	0	0	173	0	0	1	0	7
52	1	3	172	199	1	0	162	0	0.5	1	0	7
57	1	3	150	168	0	0	174	0	1.6	1	0	3
48	1	2	110	229	0	0	168	0	1	3	0	7
54	1	4	140	239	0	0	160	0	1.2	1	0	3
48	0	3	130	275	0	0	139	0	0.2	1	0	3
49	1	2	130	266	0	0	171	0	0.6	1	0	3

Figure 1: Cleveland Heart Disease Dataset Sample

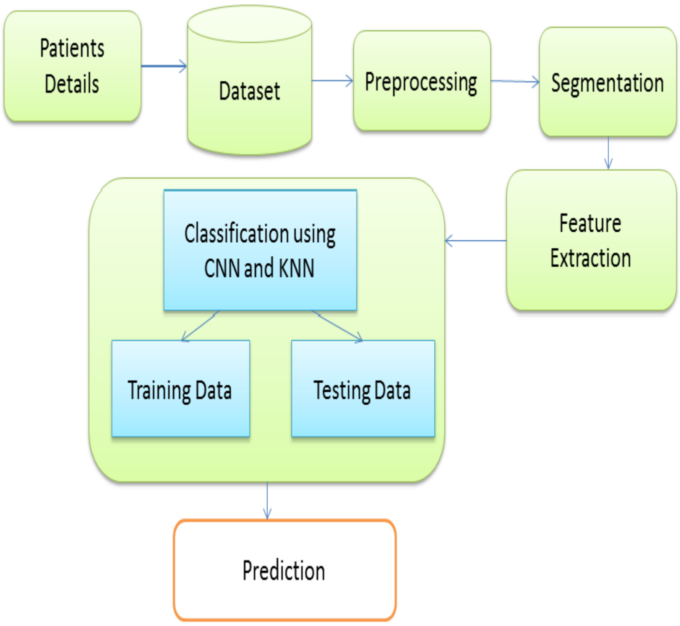


Figure 2: Architecture of the proposed system

Then the dataset's attributes are given to the convolutional neural network as shown in Fig 2. The CNN takes these raw data as inputs and applies a series of convolutional filters and pooling layers to learn relevant features. The filters learn to detect important patterns in the data, such as specific combinations of medical attributes that are indicative of heart disease. The pooling layers then reduce the dimensionality of the learned features by summarizing them in a more compact form.

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

Once the CNN has learned relevant features from the raw data, these features can be used as inputs to the classifier to predict the presence or absence of heart disease.

C. Convolutional Neural Network

Convolutional Neural Network is generally used for image processing but then it could also be used for prediction by using a modified pattern recognition using categorical values. In our system, we initialize the process by converting the categorical values of the dataset into a vector form, which bypasses the pixel conversion of the image process. Then, the contents of the input layer is given to the convolution layers, where the process is followed by the application of kernel, which is a set of filters used for feature extraction. Then, it processes the values accordingly and extracts the feature map. Stride could also be used for better efficiency. Also, border problem solving is done using padding with 0s. The pooling layer is used to reduce dimensions provided to reserve all the important features of the given data. Max pooling is used here with 2*2 dimensions, but average pooling could also be used. The feature map is flattened and made into a 1-dimensional array, which is then input into the fully connected layer, where classification is carried out using the pre-trained patterns in the datasets. If the value of an attribute ranges within a specified limit it recognizes the value pattern and classifies the patient as susceptible or not and the output is displayed accordingly. Filter size 64 would be optimal for the system to achieve maximal performance.

D. K-Nearest Neighbour

The KNN algorithm works by finding the K nearest neighbors to a new instance (i.e., a patient with medical attribute values) in the training set, and then using the class labels of those neighbors to predict the class label of the new instance. The value of K is a hyperparameter that is set prior to training the model.

In the context of heart disease dataset, the KNN algorithm can be applied by using the medical attribute values of each patient as the feature vector, and the class label (i.e., presence or absence of heart disease) as the target variable. The dataset is split into training and testing sets, with the training set used to fit the KNN model and the testing set used to evaluate its performance.

To make a prediction for a new instance in the testing set, the KNN algorithm finds the K nearest neighbors in the training set based on the Euclidean distance between the feature vectors. The class label of the new instance is then predicted as the majority class label of the K nearest neighbors.

4. Experimental Results and Discussion

The Cleveland Heart Disease Dataset in the UCI repository provided the information for our investigation. There are 303 samples in this dataset, with 76 attributes. Among this most important 14 attributes are taken for this experiment. The data set does not contain much null values. So all the samples are taken into consideration. In this study, a convolutional neural network is utilised to predict cardiac disease. When we survey the research papers we have understood that when we use machine learning algorithms for classification, we need to apply feature selection algorithm separately inoder to derive optimal set of features. Without using feature selection algorithm, when we apply machine learning algorithm it is not possible to derived the optimal features. Selecting the optimal features leads to high compuational complexity and also it is time consuming. It generates more false predictions which may lead to misdiagnosis and inappropriate therapy thus such process is tedious and time consuming. In recent years, researchers have started working on evolutionary algorithms for generating optimal subset of features. However, the fundamental disadvantage of this strategy is that it takes a long time to train the model. And when comes to disease prediction optimal features play a vital role. So Neural Networks came into picture as it works like a human brain and can trainn the model effeciently in short span of time. Deep learning algorithm automatically extract features with out the human intervention and can also work on massive amount of complex data.

The Convolutional Neural Network (CNN) is a deep learning technology that helps in feature extraction and categorization of images. For heart disease prediction, we employed CNN to train and test categorical and numerical data. We used machine learning and deep learning algorithms in our research and found that deep learning outperforms machine learning even when dealing with numerical data.For the training and testing of the data, we used KNN and CNN

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

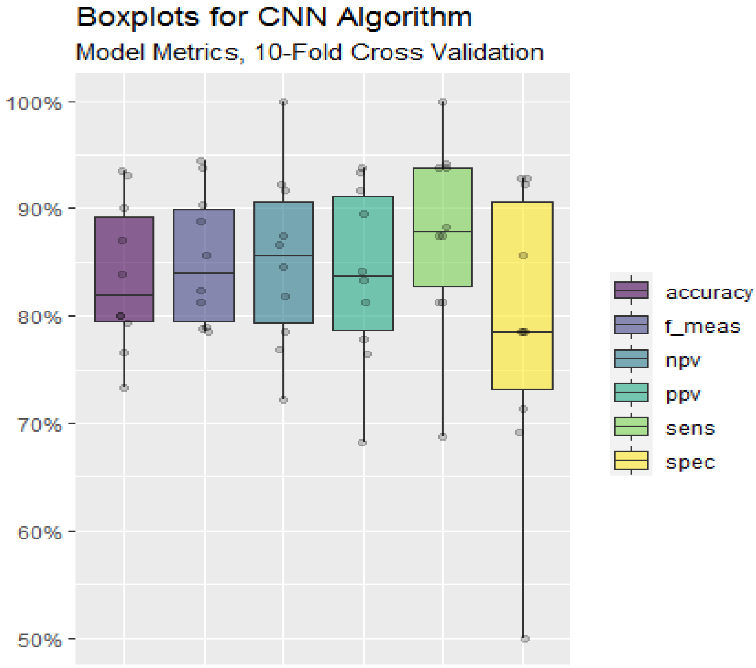


Figure 3: Performance measures for cleveland dataset using CNN

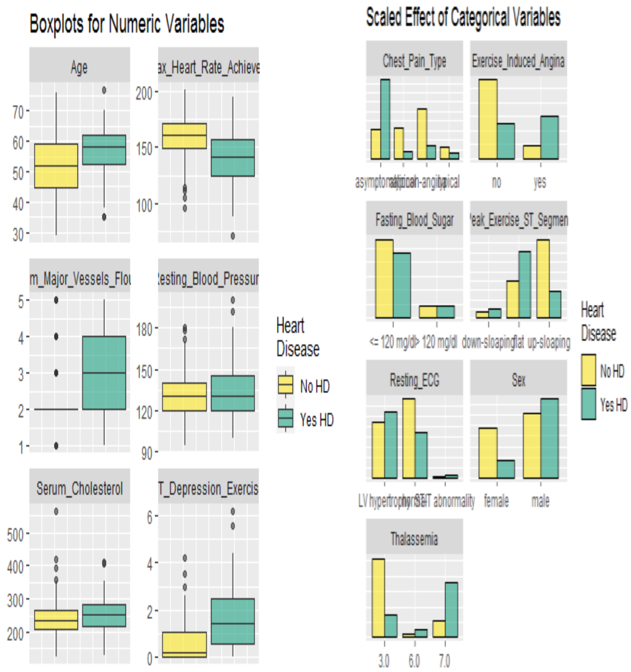


Figure 4: Characteristics that indicate whether or not a person has heart disease

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

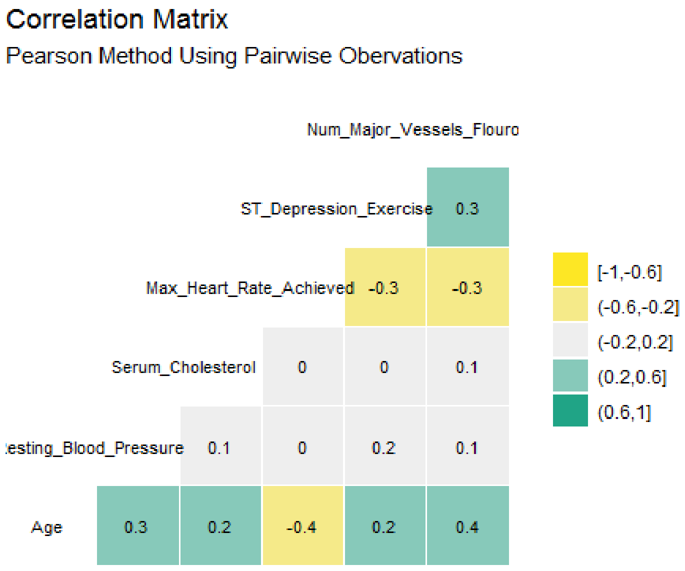


Figure 5: Correlation matrix describing features contribution using pairwise observations

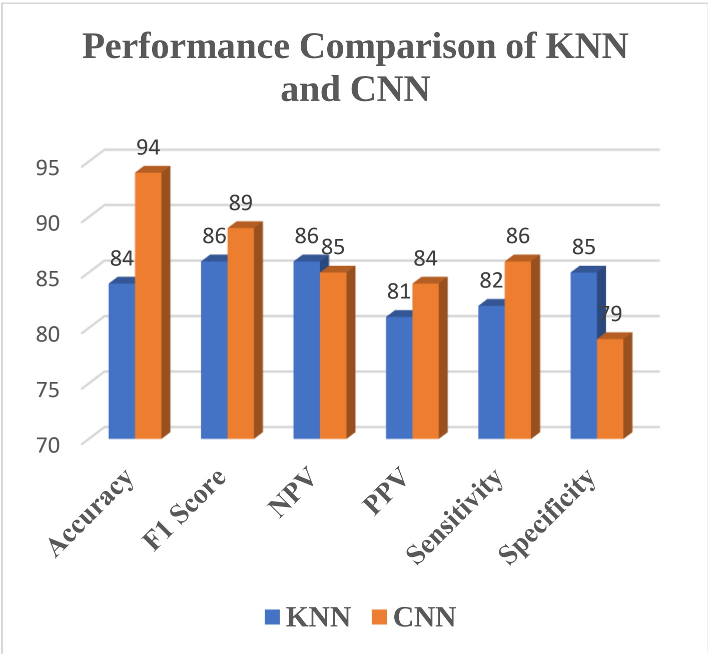


Figure 6: Comparison of Performance Metrics for CNN and KNN

In this system, we offer automated predictions about the patients heart condition so that further treatment can be made effective. The primary goal of our method is to enhance accuracy in predicting whether or not cardiovascular disease is present. We suggest using a shallow convolutional neural network (CNN), which is inspired by the success of deep networks and has the convolution layers placed between two fully connected layers, to get over the limitations of inaccurate prediction. The Convolution Neural Network (CNN) technique utilizes all features without any limitations of feature determination. By using CNN, important features can be automatically detected making it computationally efficient. Optimal accuracy is also achieved using CNN. 70 percent of the data is used for training, whereas 30 percent is used for testing. Figure 3 and 4 describes the box plot for various performance measures like accuracy, specificity, sensitivity etc.

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

Table 1: Confusion matrix for heart disease prediction

S.No	Prediction	Truth	n	Outcome
1	No Heart Disease	No Heart Disease	29	True Negative
2	Heart Disease	No Heart Disease	3	False Positive
3	No Heart Disease	Heart Disease	6	False Negative
4	Heart Disease	Heart Disease	21	True Positive

The variables for predicting heart disease are depicted in Figure 5. Age, gender, chest ache, ECG value, blood pressure, cholesterol, thalasemmia, and other factors are all utilised to forecast the risk of heart disease. The comparative analysis of KNN and CNN is shown in Figure 5. We can see from the graph that the accuracy provided by CNN is higher than the value of KNN. Figure 7 depicts the confusion matrix, which includes serum cholesterol, blood pressure, maximum heart rate reached, and the number of main arteries. Also, the other performance measures indicate the optimal accuracy of the system.

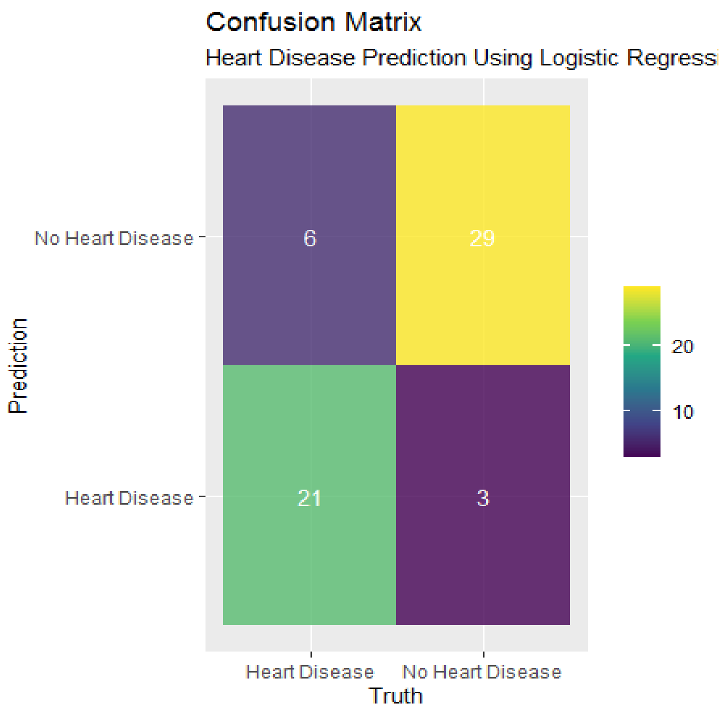


Figure 7: Confusion Matrix describing the end result of our system.

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

```
1    165
0    138
Name: target, dtype: int64
```

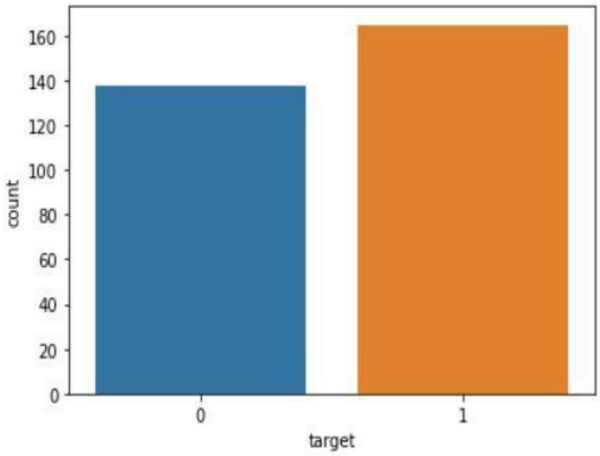


Figure 8: Count plot on target vs samples

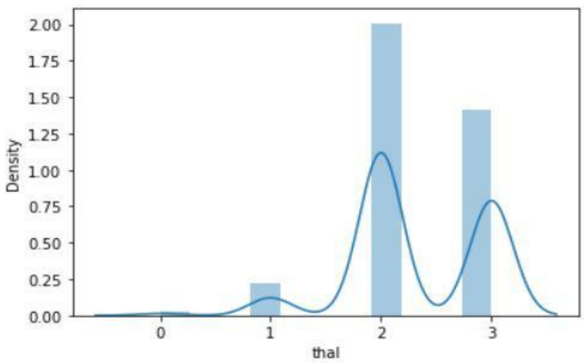
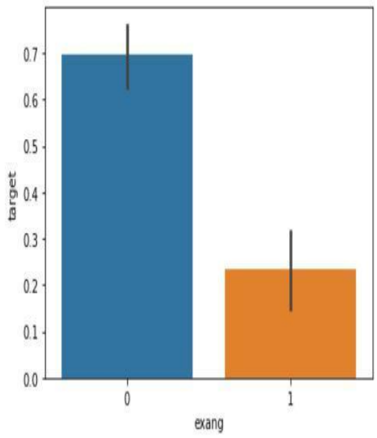


Figure 9: Plot on thalassemia vs density samples



People with exang=1 i.e. Exercise induced angina are much less likely to have heart problems

Figure 10: Countplot on exang vs target samples

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

The count plots for the total instances with the target outcomes are shown in Figure 8. In a total of 303 samples, 165 people are thought to have cardiac disease. Figures 9 and 10 show how the dataset's thalach and exang features correlate to the goal values. The graphical user interface for heart disease prediction is shown in Figure 11 a, b, and c. The user will be asked to enter their age, gender, blood pressure, blood sugar, forms of chest pain, exang, maximum heart rate reached, slope, CA, and ECG findings into this system. The system will predict and display whether the patient have heart related issues or not. The test findings indicate that our proposed strategy has a better ability to forecast heart disease than existing methodologies. This system could be used as a clinical decision support tool as it can help you to diagnose and to provide preventive care at the earliest and it can alert you to potential errors. Figure 12 a and b shows the performance metrics for the proposed system using KNN and CNN algorithm. The proposed system achieved accuracy 94%, F1 score 89%, sensitivity 86% and specificity 89%.

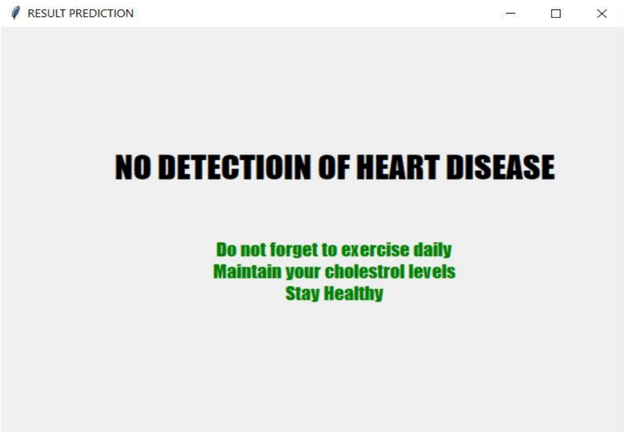


Figure 11 : User Interface featuring the 0 result prediction page.

Figure 12: User Interface featuring the prediction system

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

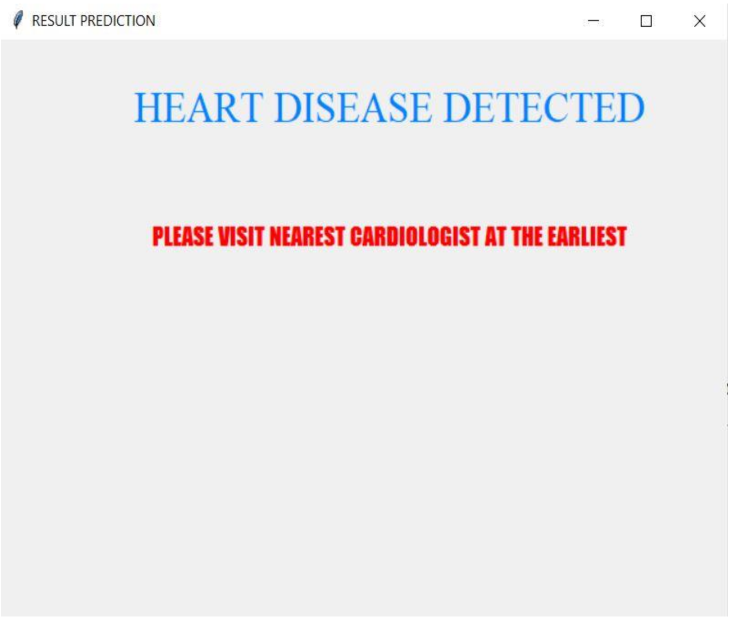


Figure 13: User Interface featuring the 1 result prediction page.

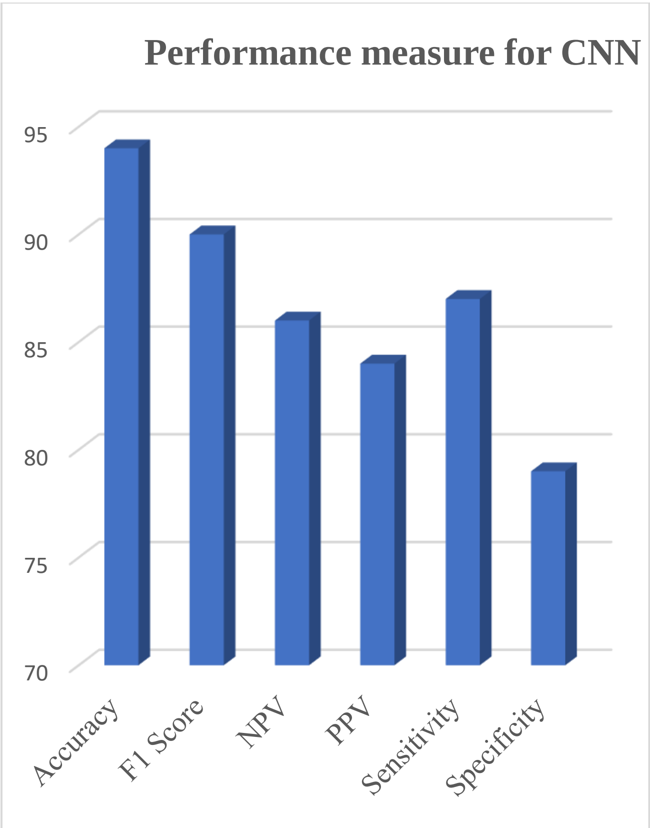


Figure 14: Performance measure for CNN using the proposed system

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

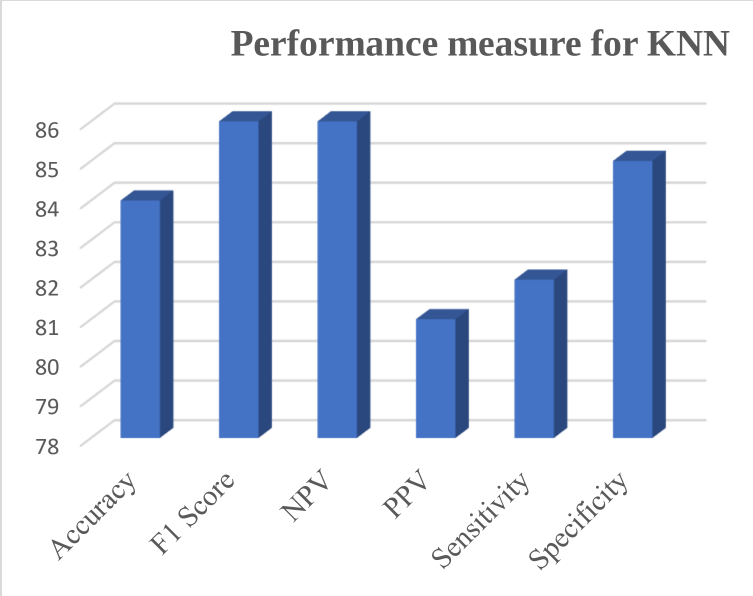


Figure 15: Performance measure for KNN using the proposed system

5. Conclusion

The systems primary goals are to increase the prediction accuracy in the presence or absence of Cardiovascular Disease, as well as to overcome incorrect prediction restrictions. This system could be used as a clinical decision support tool because it can assist you in diagnosing and providing preventive care as soon as possible, as well as alerting you to potential errors. Low false positives and false negatives are indicated by the F1 measure, indicating that real threats have been recognised. Other performance measures show that the system's accuracy is optimal, with a recall score of 87 percent and a specificity of about 79 percent. When compared to existing methodologies, the test results show that our proposed technique has a more grounded capability to predict heart disease. We discovered that increasing the convolutions to four layers allows us to achieve the best accuracy. The Convolutional Neural Network algorithm is a method for determining early heart disease risk, and the accuracy obtained using our model ranges up to 94 percent. More trials will be conducted in future to increase the attribute count and performance accuracy of these prediction systems for heart disease diagnosis. In the future, our system could be used as a clinical decision support tool to help you with treatment plans.

References

[1] Shankar, V. V., Kumar, V., Devagade, U., Karanth, V., & Rohitaksha, K. :Heart Disease Prediction Using CNN Algorithm. SN Computer Science, 1(3), 821,2019

[2] Amin, M. S., Chiam, Y. K., & Varathan, K. D. :Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, 82–93,2019

[3] Samuel, O. W., Yang, B., Geng, Y., Asogbon, M. G., Pirbhulal, S., Mzuri Kwa, D., Idowu, O. P., Ogun dele, T. J., Li, X., Chen, S., Naik, G. R., Fang, P., Han, F., & Li, G. : A new technique for the prediction of heart failure risk driven by hierarchical neighborhood component-based learning and adaptive multi-layer networks. Future Generation Computer Systems, 110, 781–794, 2020

[4] Aziz, A., & Ur, A.: Detection of Cardiac Disease using Data Mining Classification Techniques. International Journal of Advanced Computer Science and Applications, 8(7), 221,2017

[5] Sharma, P., Choudhary, K., Gupta, K., Chawla, R., Gupta, D., & Sharma, A.: Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning. Artificial Intelligence in Medicine, 102, 101752

[6] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. :A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms. Mobile Information Systems, 1–21, (2018)

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023

[7] Ali, L., Rahman, A., Khan, A., Zhou, M., Javeed, A., & Khan, J. A. :An Automated Diagnostic System for Heart Disease Prediction Based on Statistical Model and Optimally Configured Deep Neural Network. IEEE Access, 7, 34938–34945

[8] M.Durairaj and.Revathi.: Prediction of heart disease using backpropagation MLP algorithm. Int.J.Sci.Technol.Res.,vol.4,no.8,pp.235–239,(2015)

[9] M.GandhiandS.N.Singh.: Predictions in heart disease using techniques of data mining. inProc.Int.Conf.Futuristic Trends Comput. Anal .Knowl Manage. (ABLAZE), Feb. 2015, pp. 520–525,(2015)

[10] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar.: Prediction of heart disease using machine learning. in Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1275–1278,(2018)

[11] B. S. S. Rathnayaka and G. U. Ganegoda.:Heart disease prediction with data mining and neural network techniques. in Proc. 3rd Int. Conf.Converg. Technol. (I2CT), Apr. 2018, pp. 1–6,(2018)

[12] Harkulkar, N. : Heart Disease Prediction using CNN, Deep Learning Model. International Journal for Research in Applied Science and Engineering Technology, 875–881. (2020)

[13] Chen, R., Lu, A., Wang, J., Ma, X., Zhao, L., Wu, W.,Du, Z., Fei, H., Lin, Q., Yu, Z., Liu, H. : Using machine learning to predict one-year cardiovascular events in patients with severe dilated cardioMyopathy. European Journal of Radiology, 178–183.Radiology, 117: 178-183, (2018)

[14] Beunza, J. J., Puertas, E., García-Ovejero, E., Villalba,G., Condes, E., Koleva, G., Hurtado, C; Landecho, M. F. : Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). Journal of Biomedical Informatics, 103257, (2019)

[15] Jabbar, M. A., Chandra, P., Deekshatulu, B.: Prediction of risk score for heart disease using associative classification and hybrid feature subset Selection. 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA).

[16] Rajesh, N., T, M., Hafeez, S., Krishna, H. : Prediction of Heart Disease Using Machine Learning Algorithms. International Journal of Engineering &Technology, 363,(2015)

[17] S.Irin Sherly, G.Mathivanan, "ECG Signal Noises versus Filters for Signal Quality improvement", 2021 International Conference in Advances in Electrical, Computing, Communication and sustainable Technologies, 10.1109/ICAECT49130.2021.9392621.

[18] S.Irin Sherly, G.Mathivanan, "An Ensemble based heart disease prediction using gradient boosting decision tree", Turkish Journal of Computer and Mathematics Education, Vol 12, No 10(2021), 3648-3660.

[19] Ashok Kumar M , Abirami A , Sindhu P , Ashok Kumar V D , Rani V, Modern Medical Innovation on the Preferred Information about the Medicine using AI Technique, Journal of Cognitive Human-Computer Interaction, Vol. 1 , No. 1 , (2021) : 8-17 (**Doi :** https://doi.org/10.54216/JCHCI.010102)

[20] K.Vengatesan , Raghvendra Vijay Naidu , Kunal Ganesh Joshi , ChaitanyaSantosh Tekane , Siddhant Ravindra Gore, Machine Learning Based Product Price Inference Using Price Elasticity of Demand Approach, Journal of Cognitive Human-Computer Interaction, Vol. 3 , No. 1 , (2022) : 08-16 (**Doi :** https://doi.org/10.54216/JCHCI.030101)

Doi : https://doi.org/10.54216/JCHCI.050102
Received: November 10, 2022 Revised: January 15, 2023 Accepted: March 04, 2023