Heart Disease Prediction using Convolutional Neural Network

Hifju Huma khanam Redhi (C191247)
Shanjida Rashid Priya (C191250)
Rufaida Hossain (C191279)

Computer Science and Engineering

Abstract : The popular big cause of death worldwide is heart disease. We can see that in study, we present a method called Cardio help used a algorithm of deep-learning which is called convolutional-neural-network to determine a patient's risk for developing cardiovascular disease. The CNN model created in this work performs better accuracy tests while using substantially less processing power than previous pre- trained models With CNN, we were able to attain a training accuracy of 94%, which is an excellent result for the dataset being used. Here is a big plus because it gives medical experts who might not have access to high- performance computing resources better access to CNN. Overall, this study's findings demonstrate how CNNs may be used to precisely diagnose hearing loss and enhance patient outcomes. CNN technology may become a more useful tool in the medical industry as it develops and becomes more widely available. The suggested approach focuses on modeling temporal data and employs CNN for early- stage HF prediction We constructed the heart disease dataset, compared the outcomes using cutting-edge techniques, and got good findings.

Keywords -Heart disease, machine learning, convolution neural network (CNN) ,Prediction

I.INTRODUCTION

Heart-disease is most of the ordinary disease. The body depends on the heart capacity to beat efficiently.Heart conditions come in a variety of forms including Chestpain, RestingBP, Cholestor, FastingBS, RestingEC etc. There are several different methods for predicting heart disease. Computers can learn from practice data using machine learning, Cholestor, FastingBS, RestingEC etc. There are several different methods for predicting heart disease.Computer can learn from practice data using machine learning, a type of Al artificial intelligence and make predictions about test based on the data learnt information. The underlying concept behind the machine learning is finding patterns in data and making prediction. Machine learning has several uses, including in bioinformaticsmedical diagnostics recommender systems, and other areas. In machine learning there are essentially three different types of learning: supervised learning, unsupervised learning, and inforcement learning.To make predictions based on recent or historical data,predictive analytics employs a number of statistical techniques from data mining,machine learning(ML) and predictive modeling Predictive analytics used in are used in many different industries, including the healthcare sector, customer relationship management, and many more. The predictive analytics are greatly influenced by deep learning [2]. There are numerous models used in predictive modeling including Navie Bayes,Logistic Regression ,Neural Network,Support vector Machine classfication and Regression Trees etc.Artificial Neural Network is one of the mathematical or computational approaches. It resembles the neurons in the human brain The artificial neural network has separate layers, attachment and propagation direction.. Every layer consists of nodes with arrows indicating the relationships between them. We used a combination of supervised and unsupervised learning strategies to train our CNN after dividing the dataset into training and testing subsets. Additionally, we evaluated how well our CNN- based method performed in comparison to other cutting-edge machine learning techniques for heart disease prediction[3]. The input layer of a neural network contains a large number of nodes.With these input layer nodes are the hidden layer nodes connected.Each input receives a weighting.The network's input nodes provide information to the nodes in the hidden layer, that process it by carrying out various operations or calculations &

before sending the results to the output node. [4] The output layer contains the node that generates the final product.This is a description of how a neural network works.

II .PREVIOUS WORK

The studies numerous have been done that which focus on the illness of heart-disease. They applied various techniques of data-mining for the diagnosis & earned different probabilities for different methods. (Raju,Durga prasad & Scholar of Tech,2017) proposed the prediction of heart-disease using the MRM which is Multiple Regression Model and it proved that MLR which is Multiple Linear Regression is appropriate for the predicting of heart-disease chances and the work is performed using the data set which us training and it consists of thress thousand (3000) instances with 13 various contents which has mentioned previous. And the data set is divided into two main parts. One is 70% of the data are uses for only training and 30% used for only testing (Dipika & Seema,2017) focuses on on techniques that can envision the disease of chronic by the data mining which the data containing in records of historical-health using Naive-Bayes, Dicision-Tree, SVM(support vector machine), and Artificial Neural Network (ANN) [5]. A comparative study which is performed on classifiers for measure the good performance on an ac(Beyenee & kamat 2018) recommended various algorithms like NB(Naive Bayes,Classification Tree;CT, KNN,LR(Logistic Regression) ,ANN & SVM [12]. The LR (Logistic Regression) gives better accuracy compared to other all algorithms. (Beyene & Kamat, 2018) suggested the Heart-disease-Prediction system using data mining techniques for the accurate rate and also suggested multi disease prediction using the data mining techniques. [6]]

Mentioned that the data-mining plays great role in predicting the multiple diseases and by using the data mining techniques,. the total number of tests can be reduced also. And this paper focuses on predicting the heart-disease, breast-cancer,diabetes7etc.Reddy&Sai,2017) proposed the heart-disease prediction using ANN(Artificial Neural Network) algorithm in the mining of data.For the increasing expenses of heart disease diagnosis diseases,there were must need to develop unique system which able to predict heart-disease. The model of prediction is used to predict the condition of the patient after that the evaluation on the different parameters like the rate of heart beat,Blood pressure and cholesterol etc. And the main thing is for the system the accuracy proved in java which proposed an analysis of CD means Cardiovascular-Disease[8]. This presenting paper proposed the data-mining techniques to predict the disease and it provide the survey of present techniques to extract the informations from dataset and this will useful and bring benefit for the practitioners of Health-Care and the performance can be based of obtained on the time taken to build the tree of decision for overall the system and the objective primary is to the predict disease[7]

Overall, although previous research have demonstrated positive outcomes in the prognosis of heart disease, their performance lags below cutting-edge techniques. Larger datasets and enhanced machine learning algorithms may help to get over these restrictions and boost the precision and effectiveness of heart disease prediction.

III. PROPOSED METHODOLOGY

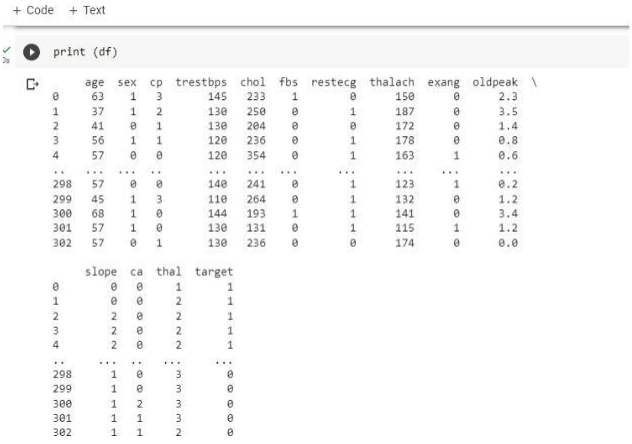
It is appropriate to refer to the challenge a CHD in a patient as a binary classification task.The neural network has been shown to be a successful classifier in specific situations when discussing supervised learning.Neural networs with application specific settings,such as several hidden layer,have significantly improved in a number of areas,according to current research.Image processing,audio processing and time series prediction are three appliucations of neural network that have shown considerable success[10]. A variety of deep learning architectures have been rigorously trained and fine-tuned using relatively larger datasets. The transfer of input data via hidden layers and assessment of error at the output layer make up how an artificial neural network The back propagated error by the output layer is then used by a gradient descent method for an iterative

update of the layer weights.Data augmentation is one of those frequently employed methods that artificially populates fresh,tiny datasets based on the instances that already exist[3]. However,when we are explicity discussing a biological application like clinical datasets, such techniques are not believable even though this

technique produces substantially superior instances. .The increased measurements of a CHD phenotype ,such as platelet count,might not match the patients probable range of readings ,to give an example. This situation arises fundamental distinction between the concepts underlying platelet count measurements and those governing statistical generation.

IV Dataset

In this part, we provide an explanation of the dataset used in this study's experimental work. As previously noted, we employ a cutting-edge dataset made accessible at specifically for the purposes and this set of attributes, which is a subset of a dataset prepared by medical professionals, of medical and shows the list of the attributes, which is subset of the attributes & big description of each one and, where relevant, a range of possible values.



	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
..	...	...	..	...	...	...	...	...	...	...	...	..	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

Fig 1

V Proposed Methodology

A model of new hybrid for prediction- based on convolutional neural network and Bi-LSTM approaches which developed to address the aforementioned issues and outperform existing work in the field of heart disease. The initial phase of the suggested strategy is to gather and load the Kaggle data repository with the Cleveland Heart disease data. Data on Cleveland Heart disease is taken into account when conducting tests. The processing of data before further processing is the next stage. In this stage, missing values have been verified and updated. High values have also been validated and normalized within the specified range using a min-max scaler. A later supplementary tree classifier that determines the relevance of each feature was used for feature selection. Once the features have been chosen, divide the dataset into the train and test sets. Following that, CNN and the Bi-LSTM approach were used to classify data in order to get the best outcomes. The outcome was then projected after calculating the assessment matrix. In the suggested method, we use a deep learning strategy based on a convolutional neural network (CNN) to increase the accuracy of heart disease prediction. In order to expand our dataset and strengthen the generalizability of our model, we will also use data augmentation approaches.

The following steps make up the suggested method :

Formatting: The UCI repository provided the data set utilized for implementation; might have some attributes with uncertain name also include a few unrelated attributes that are not helpful for the planned work's increased performance. The following R command, Dataset$Thal<-Null, was used to remove the attribute "Thal" from the dataset.

Cleaning: This phase of pre-processing involves deleting or correcting any entries that are missing from the data frame.This step is advised to eliminate the row comprising these incomplete columns as well as certain superfluous elements in the data frame.

Sampling: sampling is always worked on the dataset to update, the performance of the code however doing so may cause the algorithm to run more slowly.

D. Architecture: The six main sections that make up the building blocks of the proposed work each have their own functions.

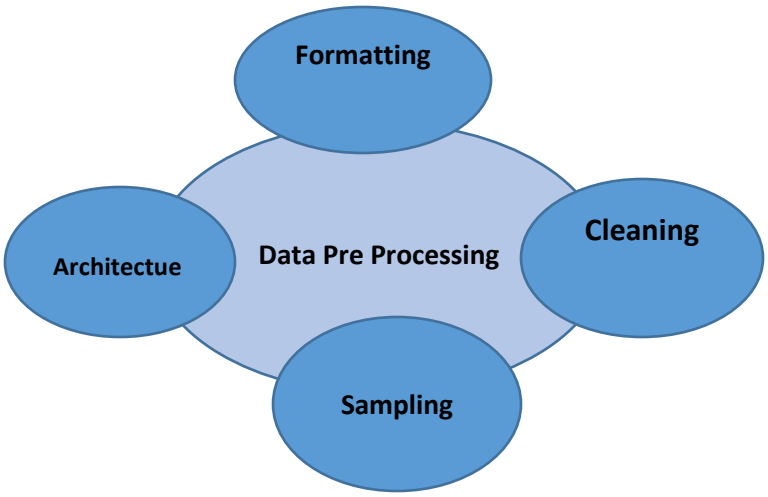


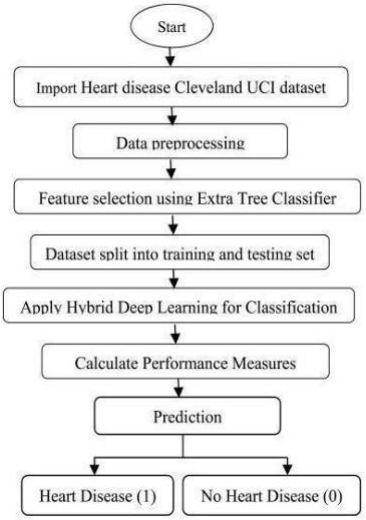
Fig 2 –Data Pre processing

Data selection : Cleveland is the name of the dataset that was chosen from the UCI repository. Discussion of the implementation dataset type has already been covered before in this part.

Data Pre-processing : Regards to the suggest work, data pre-processing is broken down into subpart. scalling and formatting It was also covered in the paragraph before this one.

Validation: Validation is necessary for any neuron-based technique since, in early phases of epochs, there was a significant discrepancy between the actual and anticipated results. To remedy this inaccuracy, ongoing work is required, and validation sets can help.

Overall , The proposed model explained in this section.As illustrated in Fig. 1 the main procedures include data preparation, CNN classification,Bi-LSTM feature selection and extra tree feature selection .In the sections that follow we will go through each of these. procedures in further detail details in the sections that follow.



VI. Result and Analysis

The hidden layer and the weight assigned to them are very important in deep neural networks. According to figure 3, , it can be seen that each neuron is given different activation function weight parameter with a different activation function value at each level.In order to achieve the best outcome, activation functions and weight values are both changed throughout model training. For the purpose of assessing the suggestions for each parameter work, particularly in relationthe mean square error and precission and a carefulstudy of the output Before evaluating the results, the dataset is carefully examined so that its skewness may be asseses [3].To determine the dataset's bias , evaluation is conducted on it. Fig-04 displays the epochs activation function value,and biasness of the data under consideration.

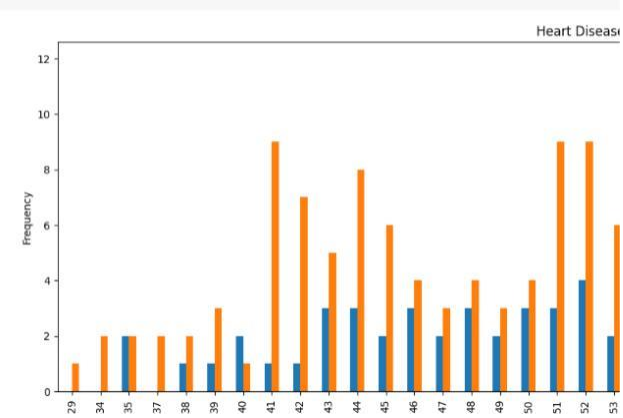


Fig 3 : Skewness in dataset

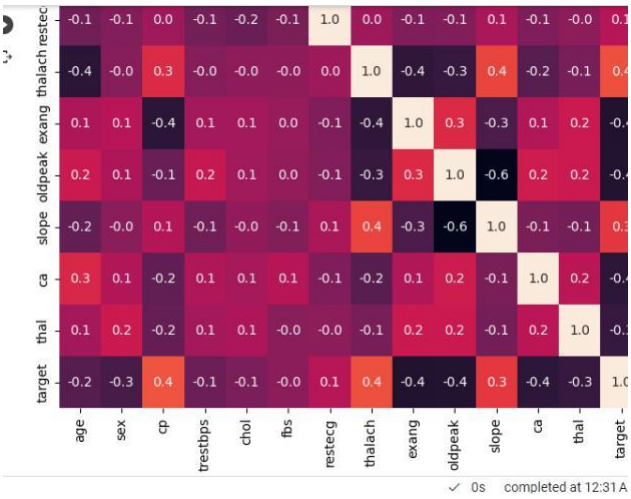
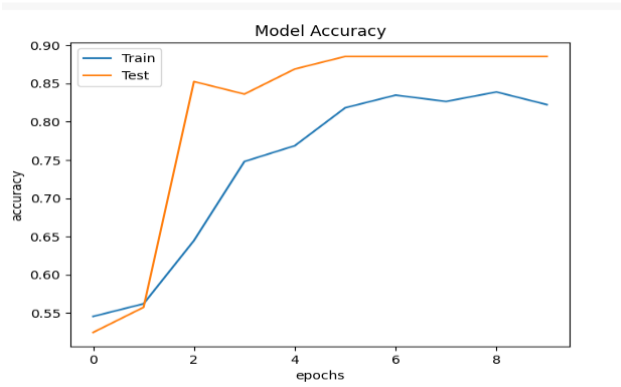


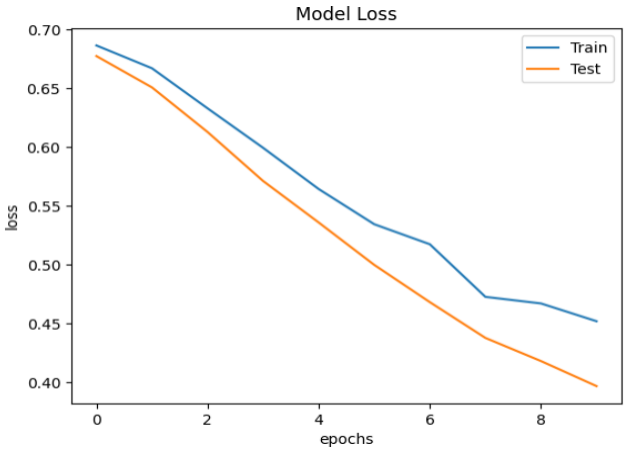
Fig. 4 correlation matrix among predictor variables

A neural network's first epoch involved using all of the training data to fine-tune the model's parameters. Epoch sizes can increase precision up to a point, after which the model starts to overfit the data.

	Precision	recall	F1 score	support
0	0.89	0.86	0.88	29
1	0.88	0.91	0.89	32
Accuracy			0.89	61
Macro Average	0.89	0.88	0.88	61
Weighted Average	0.89	0.89	0.89	61

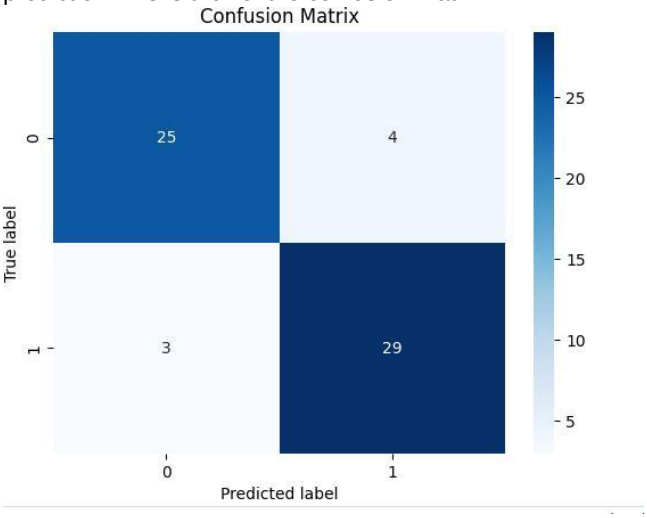
Fig 5 classification Report





Loss is the result of a bad prediction.Loss ,then is a measure of how poorly a single case was predicted by the model.If the models forecast right,there is no loss ,otherwise there is a greater loss.we can see that in this graph the blue mark shows that train accuracy and another one shows that test accuracy

A confusion matrix is a collection of the predicted results of a classification issue.For each class, count values are utilized to express the properties of accurate and incorrect prediction.This is trick of the confusion matrix.



VII Conclusion & Future Work

The ability to detect heart disease applying machine learning and deep learning code has recently attracted a lot of attention from researchers. The indicated try aims to update accuracy using a deep learning method, despite the fact that much has already been agreed upon and much more has to be done in this field. Considering how little research a deep learning computation of heart attack prediction has been conducted, this project can provide a platform for additional accuracy improvements by utilizing different deep learning algorithms. This essay included all the major ideas in the work, as well as the fundamental terms and ideas that must be covered. The intricate details of each architecture component have also been explained, along with a complete architecture model. Due to this work's explanation of the interconnected workings of various architectures, it is imperative to understand how they function. This section has covered implementation specifics,the way the code is ordered and the package needed for execution, A comparison is needed to convey the accomplishment attained by the intended task since result analysis is one of the key areas that needs board attention. As a result, each outcome based on mean square error and accuracy is described in depth and graphically represented . When compared to other machine learninh algorithms that have been employed in the past to predict heart attacks, deep neural networks clearly outperform them all in terms of messiness, . As was previously discussed, there is a lot of room for improvement in utilizing the terms of the accuracy different deep learning algorithms the work of future on this topic also will use the additional deep-learning-algorithms in an effort to accuracy improve.

REFERENCES :

[2] A. M. I. M. I. N. N. Masud, "Prediction of Heart Disease Using CNN,"arabian journal for science and engineering, 2021.

[3] A. M. I. M. I. N. N. Masood, "Prediction of Heart Disease Using Deep ConvolutionalNeuralNetworks," Awaais Mahmod1·Monwar Iqbal1·Zahid Mahmod2·Aun Irtaza1·Marriam Naoaz1·Tahera Najir1·Mumina Masud pp. 1-4, 2021.

[4] M. A. A. R. H. Sharma3, "Improved Heart Disease Prediction Using Deep Neural Network," Assian jurnal *of Computer Science and Technology,* vol. 8, no. 2, pp. 49-54, 2019.

[5] M. A. R. a. H. S. Mohd Ashraf1, "Improved Heart Disease Prediction Using Deep Neural Network," 2019.

[6] M. K. M. J. Jasmine, "Neural Network based
Heart Disease Prediction," international jurnal of engineering research & technology (ijert), pp. 1-4, 2017

[7] A. M. I. M. I. N. N. Masood, "Prediction of Heart Disease Using Deep Convolutional Neural Networks," awaais Mahmud1·Monwar Iqbal1·Zahid Mahmud2·Aun irtaza1·Morium

Nawaz1·Tahira Nazir1·Momina Masood, pp. 1-4, 2021.

[8] M. A. A. R. H. Sharma3, "Improved Heart Disease Prediction Using Deep Neural Network," *Asian Journal of Computer Science and Technology,* vol. 8, no. 2, pp. 49-54, 2019.

[9] M. A. R. a. H. S. Mohd Ashraf1, "Improved Heart Disease Prediction Using Deep Neural Network," 2019.

[10] A. Rajkumar and G. S. Reena, "Diagonsis of heaer disease using data mining algorithm," Global Journal of Computer Science and Technology, vol. 10, pp. 38-43, 2010.

[11] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Graph based approach for heart disease prediction,” in Proc. Third International Conference on Trends in Information, Telecommunication and Computing, vol. 150, 2013, pp. 465-474.

[12] A. Damesh and A. S. Meena, "Diagonsis of heaer disease using data mining algorithm," Global Journal of Computer Science and Technology, vol. 11, pp. 40-70, 2011.

[13] L. A. Monish, B. L. Deekshatulu, and L. Chandra, “Network based approach for heart disease prediction,” in Proc. Third International Conference on Trends in Information, Telecommunication and Computing, vol. 150, 2013, pp. 465-474.

[14] L. Rajkumar and O. S. Urmi, "Enable approach of heart disease using data mining algorithm," Journal of Computer Science and Technology, vol. 12, pp. 36-40 2011.

[15] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Graph based approach for heart disease prediction,” in Proc. Third International Conference on Trends in Information, Telecommunication and Computing, vol. 152, 2015, pp. 464-476.

[16] P. Ramesh and S. Urmina, "Diagonsis of heart disease using Neural Network," Global Journal of Computer Science and Technology, vol. 11, pp. 46-75, 2013.

[17] L. A. Gonish, B. L. Deekshatulu, and L. Chandra, “Network based approach for heart disease prediction,” in Proc. Third International Conference on Trends in Information, Telecommunication and Computing, vol. 158, 2012, pp. 460-475