iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** XII **Month of publication:** December 2020

**DOI:** https://doi.org/10.22214/ijraset.2020.32671

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue XII Dec 2020- Available at www.ijraset.com*

# Heart Disease Prediction using CNN, Deep Learning Model

Nilam Harkulkar[1], Swati Nadkarni[2], Dr. Bhavesh Patel[3]
[1]*M.E. Student, Information Technology Dept., Shah and Anchor Kutchhi Engineering College, Mumbai.*
[2]*Head of Department, Information Technology, Shah and Anchor Kutchhi Engineering College,Mumbai.*
[3]*Principal, Shah and Anchor Kutchhi Engineering College [SAKEC], Mumbai.*

***Abstract:*** ***Heart disease is one of the most serious health threat growing among worldwide, for which mortality rate around the world is very high. Early detection of heart disease could save many lives, accurate detection of heart disease is crucial among the health care persons through regular clinical data and its analysis. Artificial intelligence is the effective solution for decision making and accurate heart disease predictions. Medical industry showing enormous development in using information technology, in which artificial intelligence play major role. In the proposed work, deep learning based approach on heart disease is done on Cleveland dataset. However existing studies are handled in Machine learning technique. The proposed work detects heart disease based in Convolutional Neural Networks. Experimental results shows our proposed work achieves high level of accuracy in prediction of heart disease.***

***Keywords:*** ***Cleveland Heart Disease Database, Decision Trees, Random forest, Hybrid algorithm, Machine learning***

## I. INTRODUCTION

Heart disease is one the most life threatening disease all over the world, due to changes in life style and lack of physical exercise obesity is more common among every lives. As per World Health Organisation (WHO) data, the deaths are nearly 31% of overall and out of which 85% is due to heart failure and stroke. The obesity, smoking, junk food, lack of exercise, living habits among people, high blood cholesterol and high blood pressure are the most common causes for heart disease. There about 17 million people losing their life every year due to heart disease. Early detection of this disease can save many lives and proven clinical analysis and accurate detection of disease is needed. Machine learning and deep learning approaches does detections accurately and reduces the diagnosis costs.

Various data mining concepts have been explored for understanding and analysing heart disease data, Similarly Neural network approaches also used identifying severity of disease among people. Premature death is the main cause of this disease if unattended on time. Data mining classification are most significant concept for heart disease and many syndrome detections in clinical laboratories. As the high number of attributes for learning may create very high number of combinations to learn and process, thus over the time, along with machine learning, feature selection techniques such as Genetic Algorithm (GA), Particle Swarm Optimisation (PSO) and Recursive Feature Elimination (RFE) are most commonly used, which selects the best attribute/features for machine learning, thus some of the work with improved accuracy is seen overall. The impact of heart disease is very high on our population, there is a high demand for more powerful tool for prediction, as many powerful treatments like Carotid Artery Stenting (CAS) and open heart surgeries are available. The most common techniques covered under data mining are clustering, association rule mining and classifications. There are plenty of algorithms available for implementing these data mining techniques. Though there are tool like weka are available for simulations, Python programming are emerging with these algorithm inbuilt with scikit learn packages. Thus the real time implementation of data mining concepts are more reliable than ever. Neural Network are the deep learning models are more powerful techniques for heart disease and tumour predictions. Posterior probabilities are considered in the hidden layers for the effective predictions. Thus in this proposed work, convolutional neural network (CNN) is used for heart disease prediction. These advancements in detection may highly reduces clinical tests costs. The following figure shows the overview of heart disease analysis through CNN model.

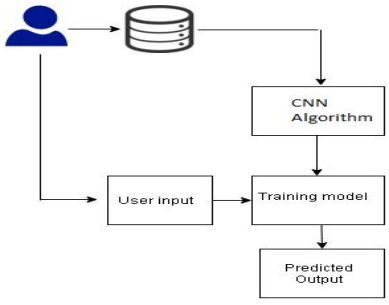



Figure: 1 Overview of Heart Disease prediction

©IJRASET: All Rights are Reserved

The objective of this study is to accurately predict heart disease based on deep learning model. Convolutional Neural Network, CNN algorithm is used for disease classification algorithm in Cleveland dataset. As like all neural networks, CNN also have three layers namely, input layer, hidden layer and output layers.

In the following chapters literature survey and related work is studied. In chapter 3, the proposed system precisely explained. In chapter 4, the demonstration results are discussed with results statistics. In chapter 5, this work is concluded and enhancements are discussed.

## II. RELATED WORK

There are many existing works have been studied by the researchers on cardio vascular disease predictions and analysis. Few works have provided good novelty on the topic, such works are discussed below.

The author in [1] analyzed heart disease prediction by novel technique called Hybrid algorithm. They analyzed all classification / machine learning algorithm and arrived accuracy for Cleveland dataset. Then proposed Hybrid model of two machine learning algorithm, to optimize the results. Hybrid models work in such a way that, the detection probabilities of one machine learning algorithm is given as input for the other machine learning algorithm. The experimental results proved that hybrid model is optimized and brings highest accuracy of heart disease detection.

In [2], Mohammed et al. discussed the implementation of data science concepts for heart disease detection using hybrid techniques. The author proposed hybrid model by combining three algorithms. Firstly, they used ANN (Artificial Neural Network) classifier, then the results are given to Support Vector machine (SVM) classifier, and finally the results are given to Naive Bayes classifier to get the prediction results. The proposed model proved high accuracy on hybrid algorithm around 88% has been achieved.

Data science framework was proposed by authors in [3] has discussed heart disease prediction. The author used only few attribute among the 13 available features, they have considered only age, sex, fasting blood sugar level and chest pain. They used a classification and a regression algorithm. For classification SVM was used and for regression, logistic regression was used. For this study, Cleveland dataset was used and achieved good precision values in logistic regression method.

Cardiovascular disease detection is done through Deep Neural Networks algorithm was proposed in [4]. Deep neural network with various types of optimization is used. Adam, Adagrad, adadelta and RMSpro are used and analyzed. The high accuracy was achieved in Adagrad optimization techniques.

Heart disease detection using deep learning models was propsoed in [5], the author used Cleveland dataset. Deep learning algorithm with different number of hidden layers was experimented in this study. Also the model evaluated with categorical and without categorical model. The accuracy achieved is high experimented in the results.

Heart disease forecast was performed in [6], the author implemented Decision tree algorithm and SVM and Artificial Neural Network algorithm on UCI dataset.

The prediction model has a superior execution for the test set and can be used for functional clinical use. The author evaluated ensemble model for prediction. Experimental results shown excellent results in ensemble model.

The author in [7] proposed heart disease prediction through machine learning processes. The author made an aim of increasing the performance measure of existing algorithms. Cleveland UCI data is exploited for the study. They evaluated exponential results in KNN classifier and other machine learning algorithm.

The author in [8] proposed HDPM through SMOTE and ENN algorithm. The evaluated two different dataset namely Cleveland and Statlog for the study. The author proposed this as a Decision making tool for clinical support. They used strong pre-processing model including feature selection, clustering through DBSCAN. On the pre processing end, outlier detection and removal also done. XGBoost algorithm is applied for the prediction of heart disease. The deployed model highest accuracy as high as 98.4 percentage.

It is inferred from the existing works, that there is a need for more efficient model, which could learn and acts as human. The existing works discussed here were used machine learning models and hybrid algorithms. There were very few studies covered under deep learning algorithm. Thus this study motivated to propose an effective model than the exiting one. Hence the Convolutional Neural Network (CNN) was proposed.

## III. PROPOSED WORK

The proposed work is implemented in anaconda 3 with Tensorflow environment created and the other libraries used were Keras, pandas, matplotlib. Dataset used on this study is from uci.edu, named Cleveland dataset. The work proceeded as binary classes classification model of heart disease prediction. CNN algorithm is applied for prediction.

©IJRASET: All Rights are Reserved

## IV. DATASET DETAILS

The dataset collected with total 13 features/attributes are considered for the study. Screen shot of the dataset is shown as follows.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slop | ca | thal | pred_attribute | |
| 63 | 1 | 1 | 145 | 233 | 1 | 2 | 150 | 0 | 2.3 | 3 | 0 | 6 | 0 | |
| 67 | 1 | 4 | 160 | 286 | 0 | 2 | 108 | 1 | 1.5 | 2 | 3 | 3 | 2 | |
| 67 | 1 | 4 | 120 | 229 | 0 | 2 | 129 | 1 | 2.6 | 2 | 2 | 7 | 1 | |
| 37 | 1 | 3 | 130 | 250 | 0 | 0 | 187 | 0 | 3.5 | 3 | 0 | 3 | 0 | |
| 41 | 0 | 2 | 130 | 204 | 0 | 2 | 172 | 0 | 1.4 | 1 | 0 | 3 | 0 | |
| 56 | 1 | 2 | 120 | 236 | 0 | 0 | 178 | 0 | 0.8 | 1 | 0 | 3 | 0 | |
| 62 | 0 | 4 | 140 | 268 | 0 | 2 | 160 | 0 | 3.6 | 3 | 2 | 3 | 3 | |
| 57 | 0 | 4 | 120 | 354 | 0 | 0 | 163 | 1 | 0.6 | 1 | 0 | 3 | 0 | |
| 63 | 1 | 4 | 130 | 254 | 0 | 2 | 147 | 0 | 1.4 | 2 | 1 | 7 | 2 | |

Figure2: Data set Sample View

The dataset variable names are described below

Table 1: Data set variable description

| Variable name | Attribute Description |
|---|---|
| Age | Age of patient |
| Sex | Sex, 1 for male |
| CP | chest pain |
| Trestbps | resting blood pressure |
| Chol | serum cholesterol |
| Fbs | fasting blood sugar larger 120mg/dl (1 true) |
| Restecg | resting electroc. result (1 anomality) |
| Thalach | maximum heart rate achieved |
| Exang | exercise induced angina (1 yes) |
| Oldpeak | ST depression induc. ex. |
| Slope | slope of peak exercise ST |
| Ca | number of major vessel |
| Thal | no explanation provided, but probably thalassemia (3 normal; 6 fixed defect; 7 reversable defect) |
| Num | diagnosis of heart disease (angiographic disease status) |

Data science study give one important feature is data visualization, which give more knowledge to the data scientist to understand and gain knowledge on data. Cleveland dataset is visualized as follows to get a histogram of number of normal and affected instances, whereas '0' represents normal and '1' represents abnormal.

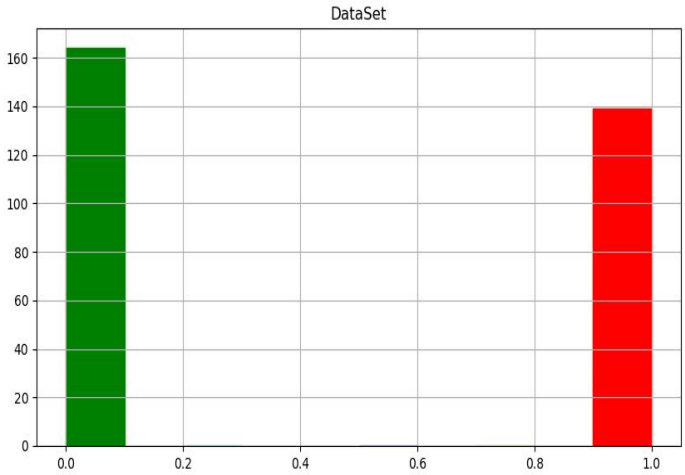



Figure 3: Data Visualization of heart Disease showing normal and abnormal

©IJRASET: All Rights are Reserved

The below historam represents nature of data values in each attributes of Cleveland data.

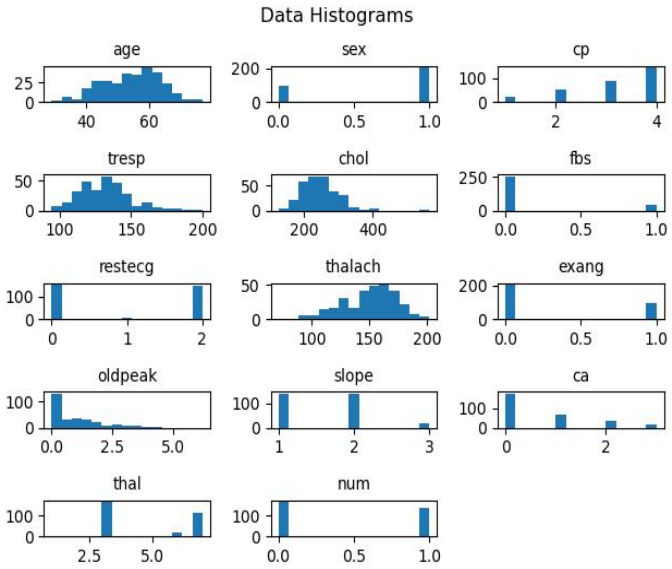



Figure 4: Data Visualization of Histogram plot for heart Disease dataset

Implementation of heart disease prediction is done through methodologies, including data collection, data visualization, date splitting, training the model and prediction. The results on data visualization is highly useful for better understanding the data nature and distributions. Many of data visualization graphs are added to the proposed system and some of the are shown in this section.

The following figure shows the data visualization of age wise histogram plot with normal '0' and heart disease '1' patients.

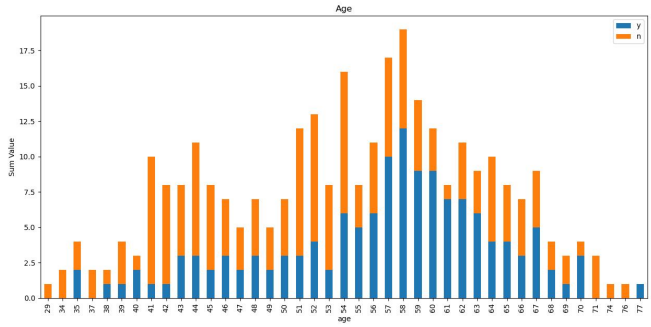



Figure 5: Data Visualization Age-wise normal and abnormal

The following figure shows the data visualization of Sex-wise histogram plot with normal '0' and heart disease '1' patients plotted in different colours.

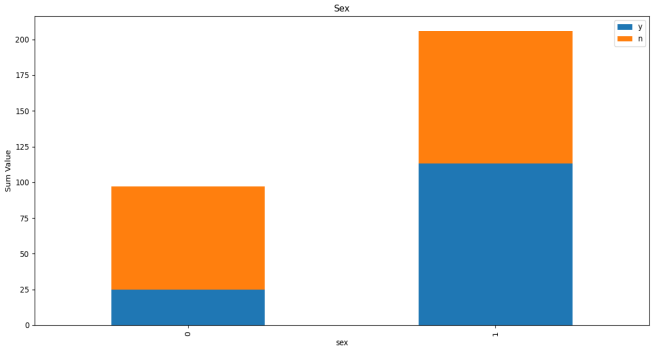



Figure 6: Data Visualization Sex-wise normal and abnormal

©IJRASET: All Rights are Reserved

The proposed architecture is given as below, user can give dataset as input the machine learns through CNN algorithm.

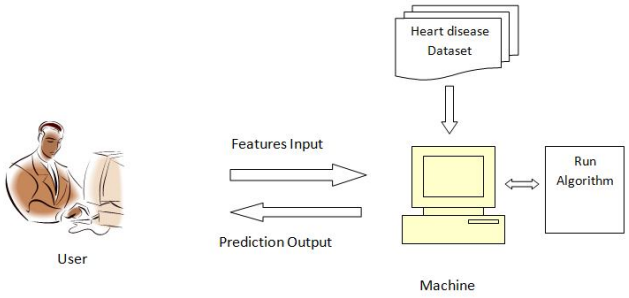



Figure 7: Architecture of Heart Disease prediction

The above figure shows the architecture of proposed system for heart disease prediction through machine learning algorithm models, which is briefly explained below. Deep learning is a part of machine learning based on multiple levels of learning and error reduction in middle layers, it more or less behaves as human brain. These techniques have inbuilt computation models, with several hidden layers to learn and predict the data. These models can itself learn the important features by feature learning process in the series of stages in middle layer.

### *A. Convolutional Neural Network Algorithm*

Designed CNN algorithm architecture is show below. One Dimensional Convolutional Neural Network (Conv1D) is used for the proposed work. In the convolutional layer, each points on input data is calculated with kernel computation.

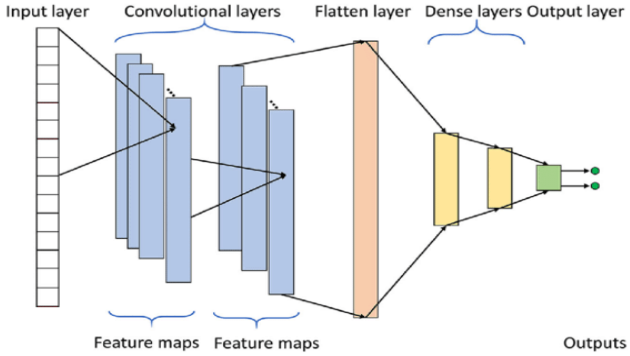



Figure 8: CNN Architecture

As given in the architecture, the input layer to the algorithm is given with dense input of 128 neurons are used. Rectified linear Unit (Relu) is the activation layer used for the work. Maxpooling with pool size 2 is added in hidden layer one and Pool size one is added to the second hidden layer. Adam optimizer is used for optimisation to handle the noise and sparse gradient problems. Mean squared error is used for error calculations and accuracy metric is evaluated.

## V. RESULTS AND DISCUSSIONS

The proposed work is implemented in Python 3.6 enabled Keras and Tesorflow packages and other mandatory libraries such as matplotlib, pandas. Dataset used for experimental analysis is downloaded from uci.edu, Cleveland dataset with 303instances. Deep learning algorithm, CNN is applied for experiment. Experimental result shown Heart disease prediction using CNN algorithm given a high accuracy of 75.2% and less error values are listed below.

| Algorithm | Accuracy (%) |
|---|---|
| MAE | 0.38 |
| MSE | 0.38 |
| RMSE | 0.61 |
| R-Squared | 0.53 |
| Accuracy | 75.2 |

Table 2 : Experimental Results

©IJRASET: All Rights are Reserved

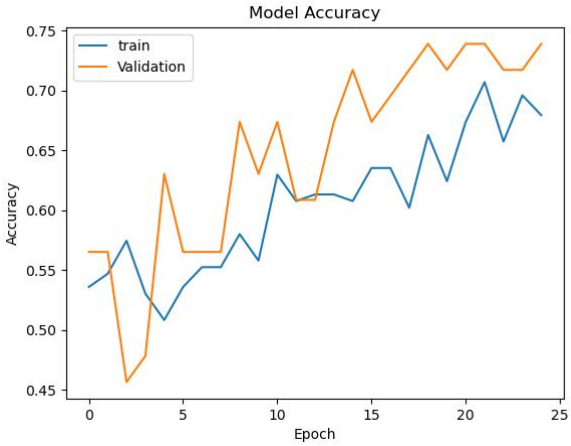



Figure 9: Heart Disease prediction Accuracy for CNN

The below figure, Figure 9, shows the MAE (mean absolute error) error values for training and validation sets for CNN algorithm. Similarly, figure 10, represents the loss for training and validation sets for CNN algorithm.

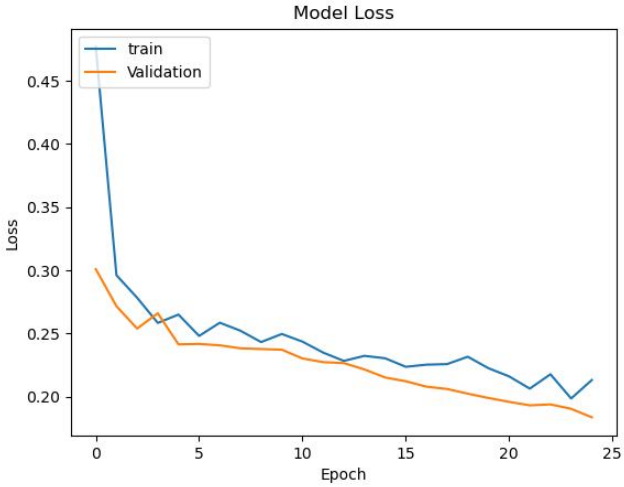



Figure 10: Heart Disease prediction Loss for CNN

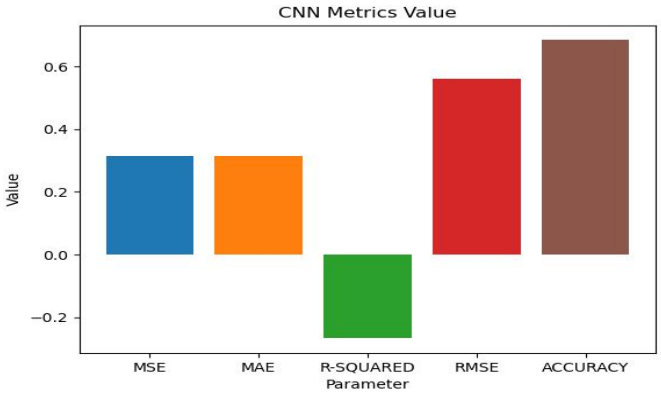



Figure 11: Heart Disease Evaluation metrics for CNN

The below figure, Figure 8, shows the MAE (mean absolute error) error values for training and validation sets for LSTM model. Similarly, figure 9, represents the loss for training and validation sets for LSTM model.

©IJRASET: All Rights are Reserved

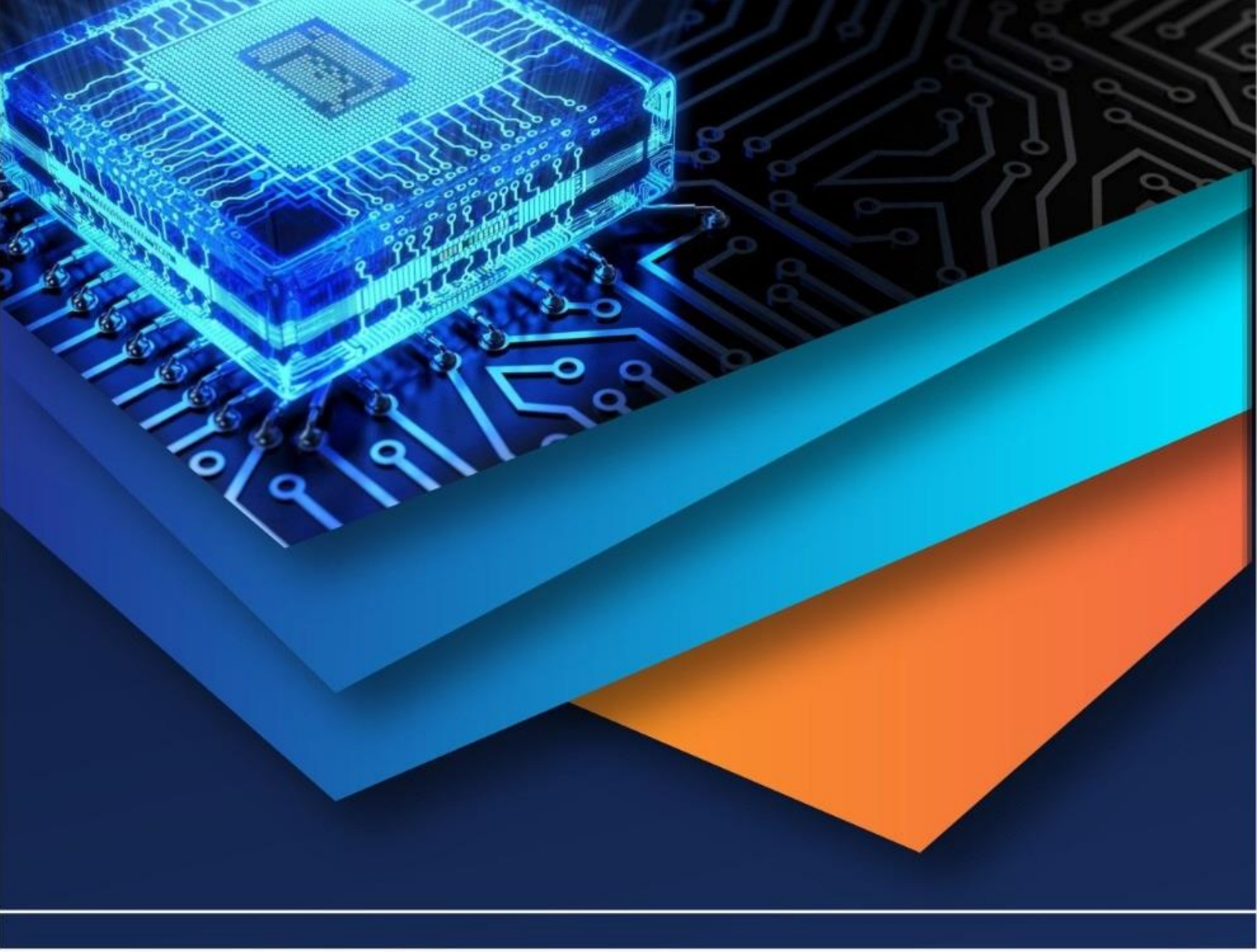
## VI. CONCLUSIONS

Coronary Disease is currently one of the life threatening disease, for which millions of lives losses occurs every year. The odds of anticipating coronary disease physically on danger factors are hard to evaluate. Many advanced diagnosis techniques are available in clinical industry, however, deep learning is considered to be the best of its choice in terms of accuracy. The experimental study shows that CNN algorithm has achieved highest accuracy. In future, the work can be extended to study ensemble models or combining different parameter for hidden layers. Along with deep learning, it can be extended with feature selection algorithms.

## REFERENCES

[1] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
[2] M. J. A. Junaid and R. Kumar, "Data Science And Its Application In Heart Disease Prediction," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.
[3] C. S. Prakash, M. Madhu Bala and A. Rudra, "Data Science Framework - Heart Disease Predictions, Variant Models and Visualizations," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020, pp. 1-4, doi: 10.1109/ICCSEA49143.2020.9132920.
[4] V. Sharma, A. Rasool and G. Hajela, "Prediction of Heart disease using DNN," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 554-562, doi: 10.1109/ICIRCA48905.2020.9182991.
[5] C. -H. Lin, P. -K. Yang, Y. -C. Lin and P. -K. Fu, "On Machine Learning Models for Heart Disease Diagnosis," 2020 IEEE 2nd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), Tainan, Taiwan, 2020, pp. 158-161, doi: 10.1109/ECBIOS50299.2020.9203614.
[6] X. Wenxin, "Heart Disease Prediction Model Based on Model Ensemble," 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2020, pp. 195-199, doi: 10.1109/ICAIBD49809.2020.9137483.
[7] B. Keerthi Samhitha, M. R. Sarika Priya., C. Sanjana., S. C. Mana and J. Jose, "Improving the Accuracy in Prediction of Heart Disease using Machine Learning Algorithms," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 1326-1330, doi: 10.1109/ICCSP48568.2020.9182303.
[8] N. L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," in IEEE Access, vol. 8, pp. 133034-133050, 2020, doi: 10.1109/ACCESS.2020.3010511.
[9] M. A. Alim, S. Habib, Y. Farooq and A. Rafay, "Robust Heart Disease Prediction: A Novel Approach based on Significant Feature and Ensemble learning Model," 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2020, pp. 1-5, doi: 10.1109/iCoMET48670.2020.9074135.
[10] P. S. Sangle, R. M. Goudar and A. N. Bhute, "Methodologies and Techniques for Heart Disease Classification and Prediction," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225673.
[11] P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithyavishnupriya, "Heart Disease Prediction Using Deep Neural Network," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 666-670, doi: 10.1109/ICICT48043.2020.9112443.
[12] M. Saw, T. Saxena, S. Kaithwas, R. Yadav and N. Lal, "Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-6, doi: 10.1109/ICCCI48352.2020.9104210.
[13] Y. Pan, M. Fu, B. Cheng, X. Tao and J. Guo, "Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 189503-189512, 2020, doi: 10.1109/ACCESS.2020.3026214.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)