iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.65851

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: 7.538
Volume 12 Issue XII Dec 2024- Available at www.ijraset.com

Analysis of Heart Disease Prediction Algorithms

Aayush[1], Dr. Yogesh Mohan[2]
Department of Computer Science, Himachal Pradesh University, India

Abstract: ***This paper investigates the accuracy of three machine learning algorithms for heart disease prediction: convolutional neural networks (CNNs), XGBoost, and logistic regression. The experiment was conducted on a dataset from Kaggle using Google Colab. All three algorithms achieved good results, but the CNN outperformed the others with an accuracy of 93.7%. Logistic regression followed with 83.46% accuracy, and XGBoost achieved 80.31% accuracy. The findings suggest that CNNs are a powerful tool for heart disease prediction, even when compared to established algorithms like logistic regression and XGBoost.***

Keywords: Artificial Intelligence, Heart Disease Prediction, Convolution Neural Network, XGBoost, Logistic Regression.

I. INTRODUCTION

A heart disease is an illness that impacts the heart or blood arteries. Smoking, high blood pressure, high cholesterol, poor lifestyle and being obese are major causes of heart diseases. Heart attacks, strokes, and chest pain are all symptoms of coronary artery disease, the most prevalent type of heart disease. [1] Globally, heart disease is the leading cause of death. Many cutting-edge technologies are employed to treat cardiac diseases. Hospitals are now able to do automatic diagnosis by analysing a patient's many health metrics and anticipate presence of heart disease. [2] The use of AI and machine learning in heart disease prediction has a rich history, dating back to the early 1990s. Early models focused on simple statistical techniques and decision tree algorithms to identify patterns and predict the risk of heart disease. As computing power and data availability increased, more sophisticated machine learning algorithms, were developed. These models were able to handle complex relationships between variables and achieve higher levels of accuracy in predicting heart disease risk. In recent years, the integration of deep learning techniques has further revolutionized the field. Deep learning models, such as convolutional neural networks can extract intricate features from complex data sources, such as electrocardiograms and medical images. This has enabled more accurate and precise prediction of heart disease, even in the early stages of the disease. This ongoing journey continues to shape the future of heart health.[3] Early Detection and Prevention, Improved Diagnosis and Treatment, Research and Innovation, etc are some applications of heart disease prediction.[4] There are many prediction algorithms. The more common ones are machine learning and deep learning. Different models have different effects. In this paper three prediction algorithms are compared to check the accuracy. Three algorithms used are convolution neural network, XGBoost and logistic regression.

II. TAXONOMY

Let’s discuss the three heart disease prediction algorithms in brief:

A. Logistic Regression (LR)

Logistic regression was developed by a French mathematician Pierre François Verhulst in the 19th century to describe population expansion and the path of many chemical reactions that are autocatalytic.[5] Logistic regression, sometimes referred to as a logit model is a statistical analysis technique that uses past observations of a dataset to forecast a binary result, such as yes / no, 0 /1, or true/false. A logistic regression model examines the relationship between one or more independent factors to predict a dependent data variable. For instance, a political candidate's chances of winning or losing an election could be predicted using logistic regression.[6] Logistic regression can be of different types, such as binomial (binary), multinomial, or ordinal depending on the nature of outcome variable. When the outcome variable can have only two categories, (e.g., disease present vs. disease absent, dead vs. alive) then binomial or binary logistic regression is used, if the outcome variable have more than two categories (e.g., drug A, drug B, and drug C) which are not ordered then multinomial logistic regression is used and if the outcome variable is ordered (e.g., poor, fair, good, very good, excellent) then ordinal logistic regression is used.[5] Logistic regression is simple and easy to implement, good for binary classification, robust to outliers, works well with small datasets and evaluate the significance of each feature in predicting the target by using the statistical significance measurements. [7]Logistic regression is widely used in many different fields, including marketing, banking, and medicine, because to its ease of use, interpretability, and efficiency in solving binary classification problems.[6]

©IJRASET: All Rights are Reserved

B. XGBoost

The XGBoost algorithm is a well-known machine learning algorithm because of its remarkable accuracy and performance in numerous machine learning tasks. XGBoost was created by Tianqi Chen and the DMLC(Distributed Machine Learning Community) team and it expands on gradient boosting. [8] XGBoost, which stands for "Extreme Gradient Boosting" combines the predictions of several weak models. [9] XGBoost uses regularization approaches and decision trees as base learners to improve model generalization. The computational efficiency of XGBoost is well known; it provides smooth handling of missing values, smart feature importance analysis, and efficient processing. It is the preferred approach for many different tasks, including as ranking, classification, and regression.[10] XGBoost has shown itself to be an effective instrument in a variety of domains like healthcare, financial sector, retail industry, detecting DDoS attacks, predicting stock market, etc. XGBoost is also used to forecast cardiac diseases. Early detection and prevention are made possible by the ability to recognize patterns and risk factors in patient data.[11] XGBoost has many advantages like it delivers excellent outcomes in a variety of machine learning tasks, it is appropriate for big datasets, also XGBoost is quite flexible as it is loaded with a large number of hyperparameters that may be changed to maximize performance. Real-world data frequently contains missing values. But, XGBoost's integrated support for handling missing values makes working with real world data simple. Despite these advantages, there are few drawbacks of XGBoost like it is less appropriate for systems with limited resources because it can be computationally demanding. Also, XGBoost is less appropriate for systems with constrained memory resources because it can be memory-intensive, particularly when working with huge datasets.[9]

C. Convolution Neural Network (CNN)

A deep learning network architecture that learns directly from data is called a Convolution Neural Network (CNN or CovNet). CNNs are especially helpful for identifying patterns in pictures so that items, classifications, and categories may be identified. [12]By finding patterns in medical data, Convolutional Neural Networks (CNNs), can be used to predict cardiac diseases. CNNs are an effective way to diagnose cardiac disease since they can learn both high-level and low-level properties.[13]CNN is to blame for the current popularity of deep learning. The primary benefit of CNN over its forerunners is that it does everything automatically and without human supervision, making it the most popular. [14]Key features of convolution neural network are that it is easy to learn and learning hierarchical features from unprocessed input is a strength of convolutional neural networks.[15] Another important feature of CNN is that it identifies hidden patterns. Healthcare data can contain hidden patterns and relationships that can be found using Convolutional Neural Networks (CNNs). [13]Convolutional Neural Networks (CNNs) are used to automatically identify early indicators of cardiac disease by examining medical pictures such as X-rays and ECGs. [15]A CNN have dozens, hundreds, or even thousands of layers, depending on how complicated its intended use is. Each layer builds on the outputs of the layers before it to identify intricate patterns. [16] The first two layers-convolution and pooling layers-perform feature extraction, while a fully connected layer and final layer translate the acquired features into the final output.[17]

Following are layers of convolution neural network:

1) Input Layer: The input layer is the first layer of a CNN. It takes in the raw data and passes it on to the next layer for further processing. [18]
2) Convolutional Layer: This layer extract features like edges, textures, and patterns by applying filters to the incoming data. Similar to specialty lenses, these filters are made to capture particular facets of the data. These filters are continuously improved during the training process.[19]
3) Pooling Layer: Pooling layers are essential for minimizing the spatial dimensions of feature maps. There are various benefits to this dimensionality reduction. First of all, it dramatically lowers the network's computational cost, making it possible to train CNNs that are deeper and more intricate. Second, pooling layers increase the network's resilience to minute changes in the input data, including rotations or tiny shifts. Lastly, pooling layers assist avoid overfitting. CNNs' overall performance and efficiency are enhanced by pooling layers. [16]
4) Fully Connected Layers: The last levels of a CNN are fully connected layers. They blend the local features—the output of the convolutional and pooling layers—to discover global patterns in the data.[20]
5) Final Layer: Making the final prediction is the responsibility of a CNN's last layer, also referred to as the output layer.

III. ENVIRONMENT

This experiment is written on google colab using python language. Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. Colab is especially well suited to machine learning, data science, and education.[21]

©IJRASET: All Rights are Reserved

Experiment is conducted using a 64 bit windows 11 system, the processor is Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz and the memory is 8.00 GB RAM. The dataset used for the purpose is taken from the following website: https://www.kaggle.com/ on 3rd December 2024 at 08:00 PM. The dataset used contains numerical values which shows different values related to patients which are required for heart disease prediction. Dataset contains record of 1026 patients. There are 14 parameters in dataset.

IV. RESULTS AND ANALYSIS

The experiment is carried out in order to compare the accuracy of three heart disease prediction algorithms, mainly convolution neural network, XGBoost and logistic regression. All the three algorithms are tested on same dataset and in same environment using same system.

A. Confusion Matrix

A confusion matrix is a table that is used to define the performance of prediction algorithms. A confusion matrix visualizes and summarizes the performance of a prediction algorithm.

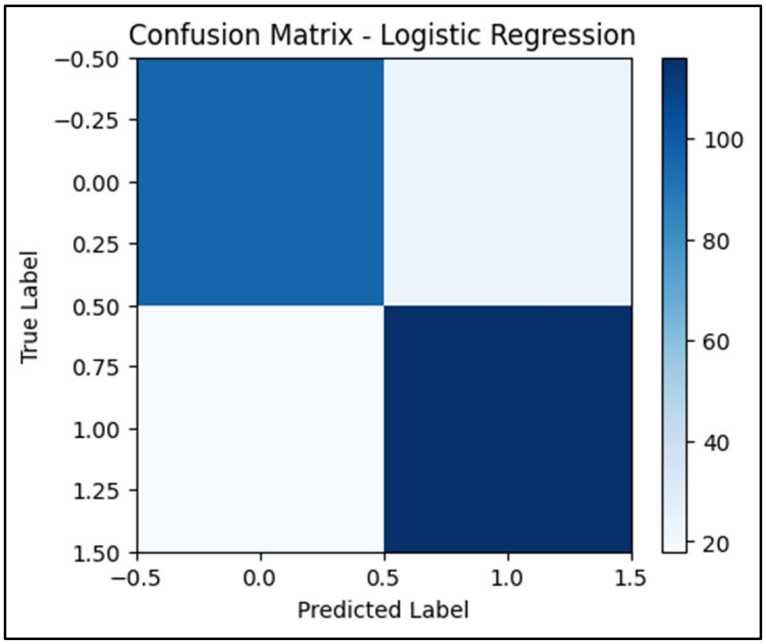


Fig.1 Logistic Regression Confusion Matrix

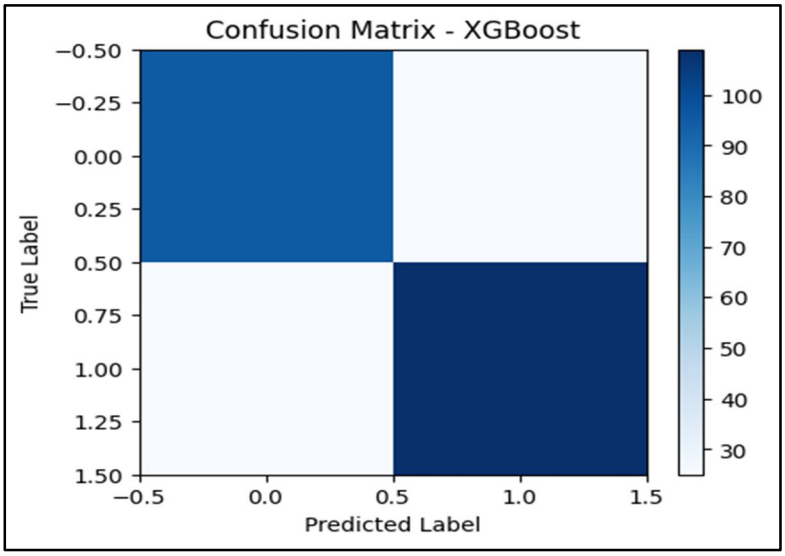


Fig.2 XGBoost Confusion Matrix

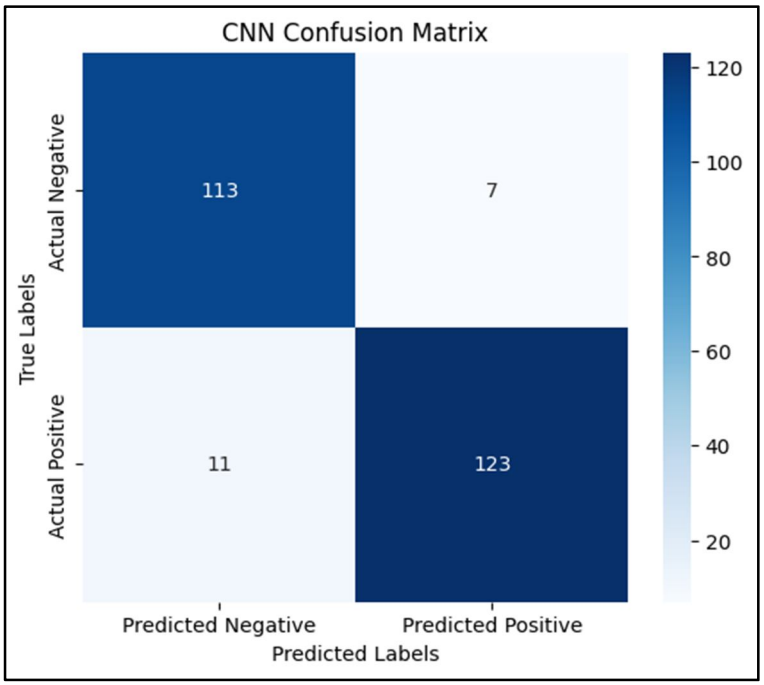


Fig.3 Convolution Neural Network Confusion Matrix

B. Classification Report:

The classification report visualizer displays the precision, recall, F1, and support scores for the model. In order to support easier interpretation and problem detection, the report integrates numerical scores with a color-coded heatmap. All heatmaps are in the range (0.0, 1.0) to facilitate easy comparison of classification models across different classification reports.

1) *Precision:* Precision can be seen as a measure of a classifier's exactness.
2) *Recall:* Recall is a measure of the classifier's completeness
3) *f1 score:* The F_1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0.0
4) *Support:* Support is the number of actual occurrences of the class in the specified dataset. Support doesn't change between models but instead diagnoses the evaluation process.[22]

TABLE 1
Classification Report for Logistic Regression

	Precision	Recall	F1-Score	Support
0	0.84	0.80	0.82	120
1	0.83	0.87	0.85	134
accuracy			0.83	254
macro avg	0.84	0.83	0.83	254
weighted avg	0.83	0.83	0.83	254

TABLE 2
Classification Report for XGBoost

	Precision	Recall	F1-Score	Support
0	0.79	0.79	0.79	120
1	0.81	0.81	0.81	134
accuracy			0.80	254
macro avg	0.80	0.80	0.80	254
weighted avg	0.80	0.80	0.80	254

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

TABLE 3
Classification Report for Convolution Neural Network

	Precision	Recall	F1-Score	Support
0	0.91	0.96	0.93	120
1	0.96	0.92	0.94	134
accuracy			0.94	254
macro avg	0.94	0.94	0.94	254
weighted avg	0.94	0.94	0.94	254

TABLE 4
Final Accuracy of Algorithms

Algorithm	CNN	XGBoost	LR
Accuracy	93.7 %	80.31 %	83.46 %

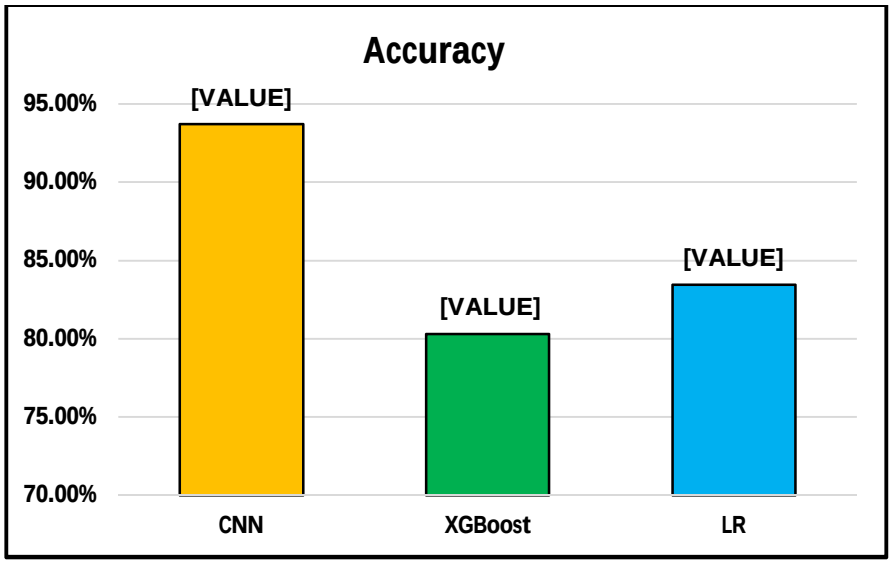


Fig.4 Algorithms Accuracy Comparison Bar Graph

V. CONCLUSION

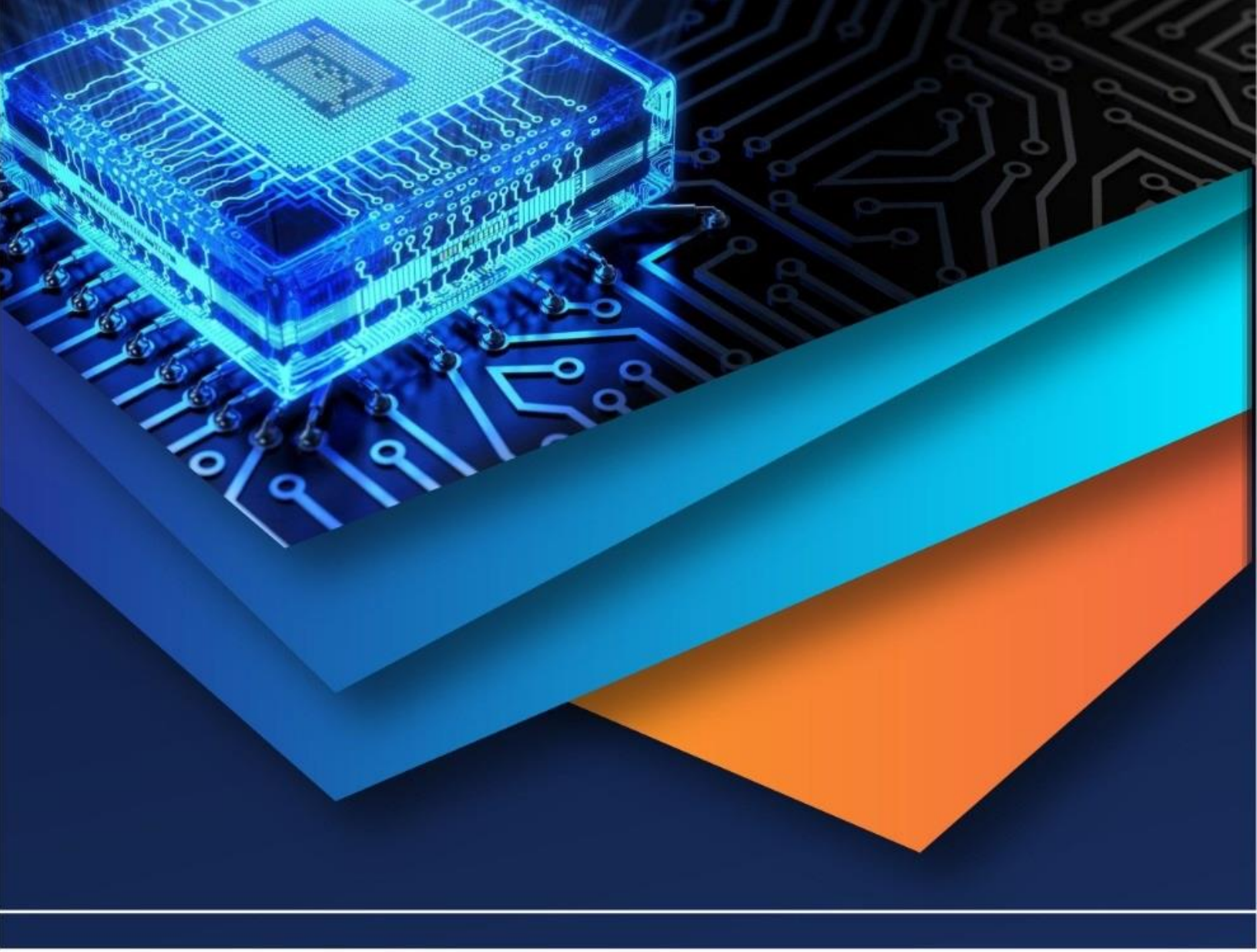
The research paper briefs about heart disease, heart disease prediction and benefits of heart disease prediction. Paper focuses on three heart disease prediction algorithms and gives a brief idea about them. The three heart disease prediction algorithms are convolution neural network, XGBoost and logistic regression. The results given by these algorithms when implemented in same environment and on same dataset are compared and analysed. The results of these prediction algorithms are expressed using confusion matrix, classification reports and graphs. On the basis of these results conclusion is drawn out that convolution neural network yields best results with accuracy of 93.7%. Second best results are given by logistic regression (83.46%) followed by XGBoost (80.31%).

REFERENCES

[1] "Definition of heart disease - NCI Dictionary of Cancer Terms - NCI." Accessed: Nov. 09, 2024. [Online]. Available: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/heart-disease
[2] "(PDF) Heart Disease Prediction." Accessed: Nov. 09, 2024. [Online]. Available: https://www.researchgate.net/publication/349140147_Heart_Disease_Prediction
[3] S. Tian, W. Yang, J. M. Le Grange, P. Wang, W. Huang, and Z. Ye, "Smart healthcare: making medical care more intelligent," *Global Health Journal*, vol. 3, no. 3, pp. 62–65, Sep. 2019, doi: 10.1016/J.GLOHJ.2019.07.001.
[4] "How AI is improving diagnostics and health outcomes | World Economic Forum." Accessed: Nov. 10, 2024. [Online]. Available: https://www.weforum.org/stories/2024/09/ai-diagnostics-health-outcomes/

[5] "Guide to Confusion Matrices & Classification Performance Metrics | by Nima Beheshti | Towards Data Science." Accessed: Dec. 09, 2024. [Online]. Available: https://towardsdatascience.com/guide-to-confusion-matrices-classification-performance-metrics-a0ebfc08408e
[6] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An efficient convolutional neural network for coronary heart disease prediction," *Expert Syst Appl*, vol. 159, Nov. 2020, doi: 10.1016/j.eswa.2020.113408.
[7] "Everything You Need to Know About Logistic Regression - Spiceworks." Accessed: Dec. 09, 2024. [Online]. Available: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/
[8] "XGBoost: The King of Machine Learning Algorithms | by Luís Fernando Torres | LatinXinAI | Medium." Accessed: Dec. 05, 2024. [Online]. Available: https://medium.com/latinxinai/xgboost-the-king-of-machine-learning-algorithms-6b5c0d4acd87
[9] "XGBoost - GeeksforGeeks." Accessed: Dec. 05, 2024. [Online]. Available: https://www.geeksforgeeks.org/xgboost/
[10] "What is the XGBoost algorithm and how does it work?" Accessed: Dec. 05, 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/09/an-end-to-end-guide-to-understand-the-math-behind-xgboost/
[11] "What are some examples of real-world applications of XGBoost? - Data Science Engineering Analytics Visualization - Quora." Accessed: Dec. 05, 2024. [Online]. Available: https://bigdata.quora.com/https-www-quora-com-What-are-some-examples-of-real-world-applications-of-XGBoost-answer-Etienne-D-Noum%C3%A8n
[12] "What Is a Convolutional Neural Network? | 3 things you need to know - MATLAB & Simulink." Accessed: Dec. 05, 2024. [Online]. Available: https://in.mathworks.com/discovery/convolutional-neural-network.html
[13] A. A. Samir, A. R. Rashwan, K. M. Sallam, R. K. Chakrabortty, M. J. Ryan, and A. A. Abohany, "Evolutionary algorithm-based convolutional neural network for predicting heart diseases," *Comput Ind Eng*, vol. 161, Nov. 2021, doi: 10.1016/j.cie.2021.107651.
[14] "Convolutional Neural Networks (CNNs): A 2025 Deep Dive - viso.ai." Accessed: Dec. 05, 2024. [Online]. Available: https://viso.ai/deep-learning/convolutional-neural-networks/
[15] P. Pitchal, S. Ponnusamy, and V. Soundararajan, "Heart disease prediction: Improved quantum convolutional neural network and enhanced features," *Expert Syst Appl*, vol. 249, p. 123534, Sep. 2024, doi: 10.1016/J.ESWA.2024.123534.
[16] "What is a Convolutional Neural Network (CNN)? | Definition from TechTarget." Accessed: Dec. 05, 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network
[17] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Aug. 01, 2018, *Springer Verlag*. doi: 10.1007/s13244-018-0639-9.
[18] "Basic structure of CNN. 1. Input layer: The input layer is the input of... | Download Scientific Diagram." Accessed: Dec. 05, 2024. [Online]. Available: https://www.researchgate.net/figure/Basic-structure-of-CNN-1-Input-layer-The-input-layer-is-the-input-of-the-whole-CNN-In_fig1_333159107
[19] "Convolutional Neural Network (CNN) | NVIDIA Developer." Accessed: Dec. 05, 2024. [Online]. Available: https://developer.nvidia.com/discover/convolutional-neural-network
[20] "Convolutional Neural Networks (CNN) and Deep Learning." Accessed: Dec. 05, 2024. [Online]. Available: https://www.intel.com/content/www/us/en/internet-of-things/computer-vision/convolutional-neural-networks.html
[21] "colab.google." Accessed: Dec. 09, 2024. [Online]. Available: https://colab.google/
[22] "Classification Report — Yellowbrick v1.5 documentation." Accessed: Dec. 09, 2024. [Online]. Available: https://www.scikit-yb.org/en/latest/api/classifier/classification_report.html

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)