Journal of Science and Technology
ISSN: 2456-5660 Volume 9, Issue 4 (April -2024)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2024.v9.i4.pp1-9*

HARNESSING DEEP NEURAL NETWORKS FOR HEART DISEASE PREDICTION

Bobburi Naga Vardhana[1] | Basati Rohitha[2] | Gunti Anusha[3] | Bulla Sneha Latha[4]| Chilaka Aiswarya[5]| Mr. R. Sudha Kishore[6]

[1,2,3,4,5] B. Tech , Students, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur

[6]Associate Professor, Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur

A. *To Cite this Article*

Bobburi Naga Vardhana. Basati Rohitha, Gunti Anusha, Bulla Sneha Latha, Chilaka Aiswarya, Mr. R. Sudha Kishore "Real-Time Text Detection and Recognition Based on Optical Character Recognition(OCR)"
Journal of Science and Technology, Vol. 09, Issue 04 - April 2024, pp1-9

B. *Article Info*

Received: 28-02-2023 ***Revised: 05 -03-2024*** ***Accepted: 20-03-2024*** ***Published: 5-04-2024***

ABSTRACT

Making forecasts and diagnosing ailments has never been simple for medical professionals when it comes to heart conditions. Cardiovascular disease medical professionals have always found it difficult to predict and diagnose. As a result, being able to people all around the world can take the necessary actions to treat cardiac disease before it becomes severe if it is discovered in its early stages. The main causes of heart disease, a severe problem in recent years, are drinking alcohol, smoking cigarettes, and not exercising. A significant amount of data generated over time by the health care sector has allowed machine learning to offer efficient results in decision-making and prediction. Healthcare is basic to human well-being, and the industry collects an expansive sum of psychiatric information. Machine learning models are being utilized to move forward the precision of heart illness forecast. These models permit people to be dependably classified as sound or unfortunate. Our think about presented a comprehensive system that gets it the standards included in anticipating patients' chance profiles utilizing clinical information parameters. The proposed appear utilizes a Significant Neural Orchestrate to effectively address issues of underfitting and overfitting. This illustrate outflanks on both test and planning data. The model's effectiveness was encouraging inspected utilizing both Profound Neural Arrange (DNN) and Manufactured Neural Arrange (ANN) approaches, coming about in exact expectations of the nearness or nonappearance of heart illness.

KEYWORDS: Deep Neural Network (DNN), Heart Disease Prediction, Artificial Neural Network (ANN), Machine Learning, Psychiatric data.

1. INTRODUCTION

As per the latest survey by the WHO, 17.9 M people pass away every year. There is no surprise that by the year 2030 it will increase to 75 million. American Heart Association considers approximately every 40 seconds, and an American will have a heart attack. Many cardiovascular diseases (CVD) exist to kill humans by using detectable hazards such as tobacco usages, unwanted eating habits, physical dormancy, and deadly usage of liquor in a range of contexts. Human beings who are having CVD are at high risk of cardiac. The disease requires early detection and guidance on the use of short medications, as set out below. Overall, CVD comes to an end with fatty stores' production within the ducts and blood groups' output. It can also be linked to an injury to tissues, such as the head, eyes, heart, and kidneys. CVD is a major leading cause of death and injury in the United Kingdom [4] but can be stopped daily to a wide degree by maintaining a good lifestyle. Cardiac cases and strokes are usually caused by powerful events and are mostly caused by a clot that prevents blood flow to the mind or heart. The most commonly known aim behind this is the creation of the most inward-looking greasy shops. This issue had created a lot of seriousness between researchers; one of the critical tasks in this is to predict the disease present in the human body. Even doctors are also not efficient in predicting the disease [5]. However, they need a support system to predict the disease. Some of the algorithms are supported but need to improvise the system's performance beyond the existing system. Therefore, to help medicos, there is enormous research scope in predicting CVD disease in humans as support of medicos. However, they need a support system to predict the disease. Some of the algorithms are supported but need to improvise the system's performance beyond the existing system. Therefore, to help medicos, there is enormous research scope in predicting CVD disease in humans as support of medicos.

Deep learning is a more popular machine learning method. It is not only when applying it in image classification tasks but also uses normal tabular data. In this model, we create a deep learning neural network model using Keras. Keras is a deep learning neural network library. Keras creates a high-level neural networks model. It developed for easy and fast experimentation. Keras supports both convolutional neural networks (CNN) and recurrent neural networks (RNN), as well as combinations of the two. It runs perfectly on CPU and GPU. This paper aims to achieve better accuracy and to make the system more efficient so that it can predict the chances of a heart attack.

2. RELATED WORK

Handling Significant Neural Frameworks (DNNs) for heart sickness desire talks to a basic movement inside the restorative field, indicating to overhaul early area and interventions methods for combating this unpreventable prosperity issue. Heart disease remains a driving cause of mortality around the world, underscoring the squeezing require for correct prescient models. Conventional hazard evaluation strategies frequently display impediments in taking care of the complexity of biomedical information and may not completely abuse the complex connections inside these datasets. Be that as it may, later advancements in profound learning, especially DNNs, offer a promising road for tending to these challenges.

In recent years, there has been significant progress in the application of machine learning and deep neural networks (DNNs) for heart disease prediction. A comparative analysis of machine learning models has been conducted by Miah et al. (2023) [1], focusing on improving cardiovascular disease prediction, specifically myocardial infarction. Their study contributes valuable insights into the effectiveness of different machine learning techniques in this domain. Similarly, Aburayya et al. (2023) [2] explored automated heart disease detection using a machine learning approach. Their work delves into the development of automated systems for efficient and accurate detection of heart diseases, showcasing the potential of machine learning in healthcare applications. Ramachandran et al. (2023) [3] introduced a novel approach utilizing 1D Convolutional Neural Networks (CNNs) for heart disorder prediction. This study highlights the efficacy of deep learning techniques, particularly CNNs, in handling complex medical data and predicting heart-related disorders.

In understanding the epidemiology of cardiovascular diseases, Bhatnagar et al. (2016) [4] provided valuable insights into the trends and patterns of these diseases, offering a broader context for the development of predictive models. Jabbar et al. (2012) [5] proposed a risk score prediction model for heart disease using associative classification and hybrid feature subset selection techniques. Their work demonstrates the integration of machine learning and statistical methods for accurate risk assessment. Nannapaneni et al. (2023) [6] developed a hybrid model integrating deep neural networks for heart disease prediction, showcasing the potential synergy between traditional machine learning and deep learning techniques. Jaiswal et al. (2023) [7] conducted an empirical analysis focusing on heart disease prediction using deep learning algorithms. Their study contributes to the growing body of research exploring the capabilities of deep learning in medical diagnosis and prognosis. Furthermore, Junaid and Kumar (2020) [8] discussed the application of data science in heart disease prediction, emphasizing the interdisciplinary nature of research in this field. Finally, Mohan et al. (2021) [9] explored supervised machine learning algorithms for heart disease prediction, highlighting the importance of algorithm selection and data preprocessing techniques in achieving accurate predictions.

These studies collectively provide a comprehensive overview of the advancements in heart disease prediction, emphasizing the role of machine learning and deep neural networks in improving diagnostic accuracy and patient outcomes.

3. METHODOLOGY

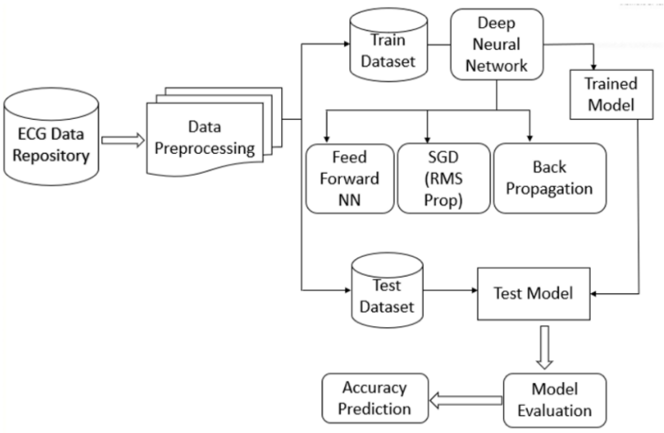

Figure 1. System Architecture

3.1 Data Source

The Electrocardiogram a deep learning repository, or ECG for short, is a great place to find free and open-source deep learning datasets. The ECG Profound Learning Store is the source of the dataset utilized in this examination to anticipate heart infection. ECG may be a bunch of datasets utilized to put Profound learning methods into hone. This dataset was gotten from a real dataset. 334 instances of data with the relevant 22 clinical parameters make up the dataset. The clinical parameter of the dataset pertains to tests that are performed in relation to heart illness, such as blood pressure measurement, type of chest discomfort, ECG result, and so forth.

However, working with ECG data presents challenges due to noise, artifacts, and signal variations, often necessitating preprocessing techniques for improved accuracy. Despite these challenges, ECG datasets find wide applications in automated diagnosis, patient monitoring, and cardiac research. It's imperative to handle ECG data ethically, ensuring patient privacy and adhering to regulatory standards such as HIPAA and GDPR. Overall, ECG datasets play a pivotal role in advancing our understanding of cardiac health and disease, facilitating the development of innovative diagnostic tools and personalized treatment approaches.

The target variable in this dataset is the presence or absence of heart disease; it is commonly labeled as1 (heart disease present) or 0 (no heart disease).

3.2 Data Preprocessing

Data preprocessing is fundamentally the necessary crucial work that guarantees the data in the proper format and quality for analysis or deep learning. It requires a number of actions meant to improve the data quality and prepare it for additional analysis. By resolving data-related problems and improving the data usefulness for the considered use, it provides the improvement of the accuracy and effectiveness of data-driven tasks and models.

3.2.1 Data Cleaning

Data cleaning is the process of filling in the blanks in elements such as blood pressure, cholesterol or ECG readings. we have made the decision to either eliminate missing values or assign these missing values using statistical techniques such as mean, median, or predictive modelling. Extreme values have the potential to change the data, so handling outliers is very important.

3.2.2 Data Transformation

By placing variables like age and cholesterol on the same scale, organizing features makes the models results directly comparable. It is necessary to encode categorical variables. For example, to make the type of chest pain numerical and suitable for modelling, it might be encoded as 0, 1, 2, or 3. To capture more complex relationships in the data, feature transformation may involve the creation of new variables, such as risk score based on multiple attributes.

3.2.3 Data Normalization

Data normalization is performed using mean normalization, also known as standardization. This process involves centering the data by subtracting the mean and scaling it by dividing by the standard deviation.

The mathematical equation for mean normalization is

$X_{normalized} = X – mean(X) / std(X)$ where:

$X_{normalized}$ is the normalized (standardized) data matrix,

X is the original data matrix,

Mean(X) is the mean of each feature in X, Std(X) is the standard deviation of each feature in X.

3.2.4 Data Splitting

This involves creating training and testing sets out of the dataset. The testing set is used to assess the prediction model's performance after it has been trained using the training set.20% is usually set aside for testing and 80% for training. Enabling the model to learn from past data in order to make precise predictions on new, unseen data is the aim of using the training dataset.

3.3 Deep Neural Network Architecture

The DNN architecture is designed to learn complex patterns and relationships from the input features to predict the likelihood of heart disease in individuals, while employing various techniques like regularization and dropout to enhance the model's generalization and prevent overfitting. The architecture follows a feedforward neural network design with multiple hidden layers, each utilizing ReLU activation for introducing non-linearity. Dropout regularization is incorporated in each hidden layer and the output layer to prevent overfitting. L2 regularization is also applied to control the complexity of the model and improve generalization.

The deep neural network (DNN) architecture for heart disease prediction comprises an input layer with 20 neurons using ReLU activation to introduce non-linearity. This is followed by three hidden layers: The first has 64 neurons with ReLU activation and L2 regularization (0.001) to prevent overfitting, along with a 20% dropout rate. The second hidden layer has 32 neurons with similar activation and regularization settings, while the third has 16 neurons with ReLU activation and L2 regularization.

The output layer, with 2 neurons for binary classification, uses SoftMax activation to yield probabilities for heart disease or no heart disease, also with a 20% dropout rate. The model is compiled with categorical cross entropy loss for multiple classes, RMSprop optimizer (learning rate 0.001), and accuracy metric. Early Stopping with a monitoring metric of loss and patience of 15 epochs is employed to prevent overfitting and halt training if validation loss stagnates. This architecture aims to learn intricate patterns from input features while mitigating overfitting through regularization and dropout techniques, culminating in a robust model for heart disease prediction.

3.4 Algorithm used for Implementation

Step-1: Initialization

Initialize the weights and biases of the neural network randomly. Choose hyperparameters such as the learning rate (η), decay rate (β), and a small constant (ϵ) for RMSprop.

Step-2: Forward Pass

Input data is fed into the neural network. Neurons in each layer compute a weighted sum of inputs, apply an activation function, and pass the result to the next layer. This process repeats layer by layer until the output is produced.

Step-3: Loss Calculation

Compare the network's output to the ground truth labels. Calculate the loss using a suitable loss function.

Step-4: Backward Pass (Backpropagation)

Compute gradients of the loss function with respect to the network's parameters. Propagate gradients backward through the network using the chain rule of calculus. Update the weights and biases to minimize the loss using the computed gradients.

Step-5: RMSprop

Calculate the moving average of squared gradients for each parameter using exponential decay. Update parameters using a learning rate scaled by the root mean square (RMS) of recent gradients and the moving average of squared gradients.

Step-6: Repeat

Iterate over the training dataset multiple times (epochs), performing forward pass, loss calculation, backward pass, and RMSprop update at each iteration. Monitor the training loss to assess convergence.

Step-7: Prediction

Use the trained network to make predictions on new data. Input the new data, perform a forward pass, and interpret the output for the task at hand.

Step-8: Evaluation

Evaluate the trained network's performance using metrics such as accuracy, precision, recall, or AUC on a separate validation or test dataset. Compare the performance to assess the model's generalization ability and adjust hyperparameters if necessary.

4. RESULTS AND DISCUSSIONS

4.1 Bar graph

A bar chart for the variable "Target" shows how many observations there were for each value. The target variable most likely indicates the proportion of people with and without heart disease, depending on how it is coded in the dataset (0 for no heart illness, 1 for heart disease, etc.).

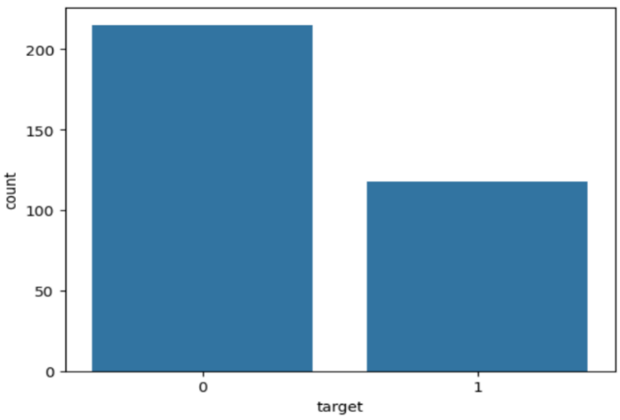

Figure 2. target vs count

A visual representation of the target variable's distribution within the dataset is offered by a count graph, which displays the relative frequencies of each category. This data can be shown visually using a bar graph, in which each bar reflects the total number of individuals with heart disease.

4.2 Histogram Grid

It shows a network of histograms, each of which delineates the esteem conveyance for a distinctive include within the dataset. The grid's columns each compare to unmistakable highlights, and each row's histograms appear the recurrence or number of perceptions for different include ranges.

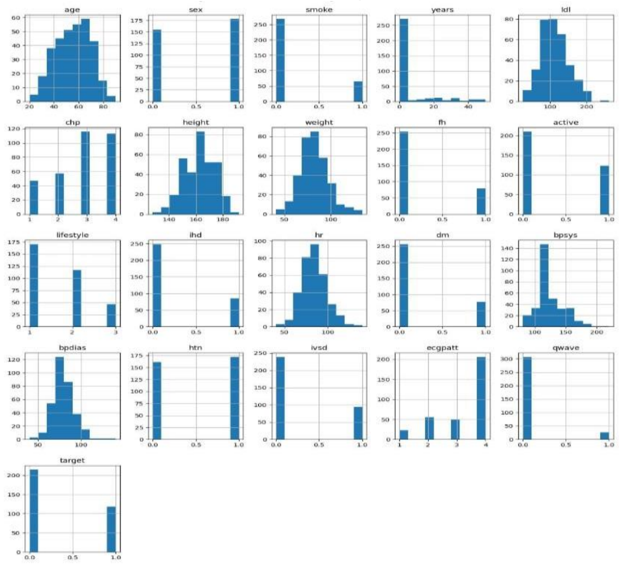

Figure 3. Data Distribution of All Participants

An outline of the distribution of each highlight within the dataset is given by this visualization, making it straightforward to spot patterns, inconsistencies, and conceivable issues like multimodality or skewness. It helps in comprehending the data's fundamental structure and can coordinate ensuing stages of information preprocessing and examination.

4.3 Crosstabulation-Based Bar Plot Display

The recurrence of heart infection cases over different age bunches is spoken to outwardly within the Crosstabulation Based Bar Plot Show for Heart Malady Recurrence for Ages. It computes a crosstabulation (cross-tab) of age and target factors (where 'target' signifies the nearness or nonattendance of heart illness) using the Pandas library, and after that plots the comes about as a bar plot.

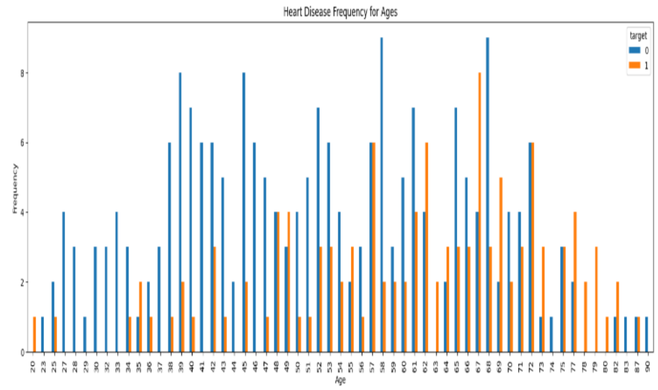

Figure 4. Heart Disease Frequency for Ages

Each age group will have a bar in the resulting plot, and the height of each bar will represent the frequency of heart disease cases within that age group. This visual aid makes it simple to compare the frequency of heart disease in various age groups and offers insights into the relationship between age and the dataset's prevalence of heart disease.

4.4 Correlation Heatmap

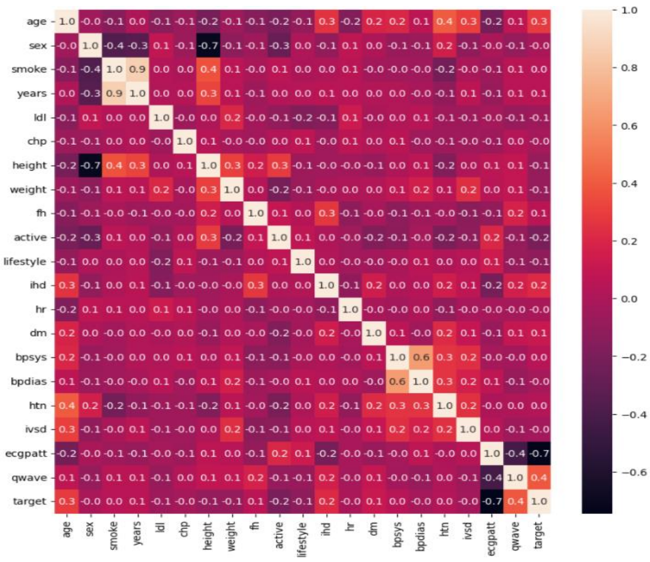

Figure 5. Histogram Equalization of the dataset

It's a heatmap with the relationship coefficient between two factors within the dataset spoken to by each cell. The exact relationship coefficient esteem is given by the comments interior each cell, making it basic to get it how the factors relate to one another. A positive straight relationship between factors is demonstrated by positive relationship coefficients (values closer to 1) and a negative straight relationship is shown by negative relationship coefficients (values closer to -1). There's no direct relationship between the factors when the relationship coefficient is 0. The heatmap encourages information investigation and comprehension by advertising a visual outline of the relationships inside the dataset. Using the Seaborn library, it shows the relationship network heatmap for the dataset. The heatmap employments a color scale to graphically portray the relationships between the factors; more grounded relationships (positive or negative) are shown by darker colors, and weaker relationships or no relationship are demonstrated by lighter colors.

4.5 Accuracy Plot

The model accuracy trend over training epochs is displayed in a line plot as the output. While the accuracy values are displayed on the y-axis, the x-axis represents the epochs, or iterations over the training data. A training accuracy line and a validation accuracy line are usually displayed on the plot. In addition to showing whether overfitting or underfitting occurs, the plot helps to visualize how the model's accuracy changes or improves during training. An indication that the model is learning well is when the accuracies in both training and validation increase and converge.

.

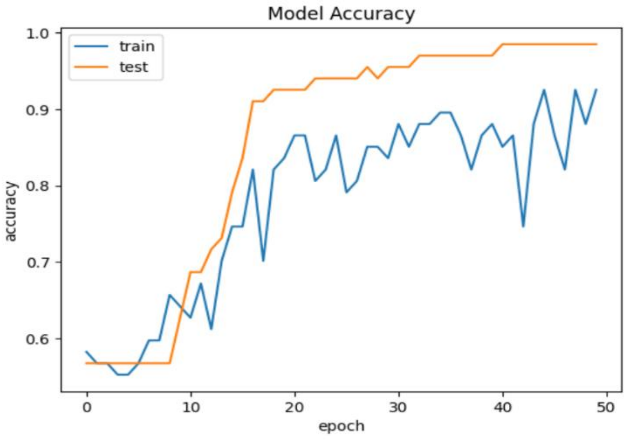

Figure 6. Model Accuracy

4.6 Training History Plot

The show loss over epochs for the preparing and validation datasets is appeared graphically as a line plot.
X-axis (epoch): Represents the number of preparing epochs, which are emphases over the complete dataset amid the preparing process.
Y-axis (loss): Represents the loss esteem, which may be a measure of how well the show is performing. The loss is regularly calculated employing a loss work, with lower values indicating way better performance.
Training loss (blue line): Shows the design of the model's loss for every epoch on the preparing dataset. It appears how well the preparing set of information is being fitted by the model.

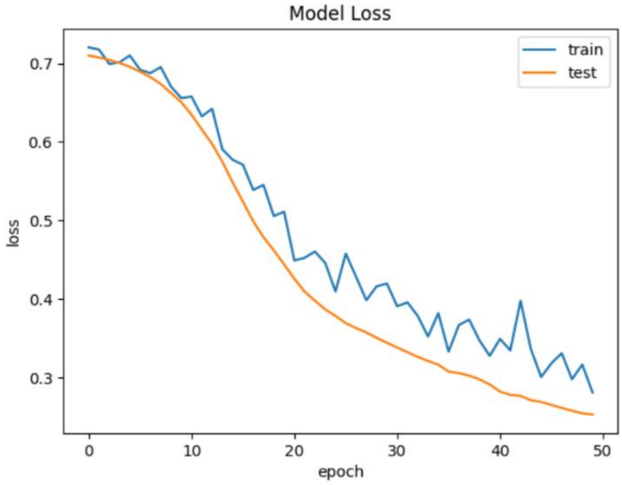

Figure 7. Model Loss

Validation loss (orange line): Shows the model's misfortune slant over each epoch on the approval dataset. It appears how well the show applies to fresh, untested information.

Table 1. Classification Report

	Precision	Recall	F1-score	Support
0	1.00	0.97	0.99	38
1	0.97	1.00	0.98	29
Accuracy			0.99	67
Macro avg	0.98	0.99	0.98	67
Weighted avg	0.99	0.99	0.99	67

5. CONCLUSION

With an astonishing precision of 98.51%, we were able to foresee heart malady utilizing profound neural systems. Utilizing cutting-edge profound learning methods, we made a dependable model that can dependably decide whether a set of input highlights demonstrates the nearness or nonappearance of heart malady. Our profoundly neural organize engineering performed strikingly well in separating between individuals with and without heart illness. It was fastidiously planned and refined through broad testing. We were able to recognize complicated connections and designs within the information by utilizing profound learning, which permitted us to deliver amazingly exact forecasts. Our model's tall exactness offers a part of potential for viable healthcare applications. Our profound learning-based strategy has the potential to be a valuable instrument for the early conclusion and discovery of heart infection, helping restorative experts in making choices and improving persistent results with extra improvement and validation.

Further studies could concentrate on improving the model's interpretability, examining the significance of particular features, and considering possible extensions to manage a variety of datasets and clinical scenarios. The potential of deep neural networks to transform healthcare and advance the field of medical diagnostics is highlighted by our project, which opens the door to more precise, effective, and customized methods of illness prediction and management.

6.REFERENCES

[1] J. Miah, D. M. Ca, M. A. Sayed, E. R. Lipu, F. Mahmud and S. M. Y. Arafat, "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction," 2023 15th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 2023, pp. 49-54, doi: 10.1109/IIT59782.2023.10366476.

[2] R. A. Aburayya et al., "Automated Heart Diseases Detection Using Machine Learning Approach," 2023 6th International Conference on Engineering Technology and its Applications (IICETA), Al-Najaf, Iraq, 2023, pp. 108-114, doi: 10.1109/IICETA57613.2023.10351330.

[3] Ramachandran, Vedantham & Rekha, Sudha & Rajesh, Barige & Narayana, Y. & Sathish, Parasa. (2023). Heart Disorder Prediction Using 1D Convolutional Neural Network. 10.1007/978-981-99-3691-5_44.

[4] Bhatnagar, P., Wickramasinghe, K., Wilkins, E., Townsend, N. (2016). Trends in the epidemiology of cardiovascular disease in the UK. Heart, 102(24): 1945- 1952. https://doi.org/10.1136/heartjnl-2016-309573.

[5] Jabbar, M.A., Chandra, P., Deekshatulu, B.L. (2012). 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), Kochi, pp. 628-634. https://doi.org/10.1109/ISDA.2012.6416610.

[6] U. K. Nannapaneni, V. K. Koneni, Y. S. Reddy and J. Nalabothu, "A Hybrid Model for Heart Disease Prediction Using Deep Neural Network," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10306620.

[7] A. Jaiswal, M. Singh and N. Sachdeva, "Empirical Analysis of Heart Disease Prediction Using Deep Learning," *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2023, pp. 1-9, doi: 10.1109/ACCAI58221.2023.10201235.

[8] 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, UK, 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.

[9] N. Mohan, V. Jain and G. Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms," *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2021, pp. 1-3, doi: 10.1109/ISCON52037.2021.9702314.

[10] S. Chua, V. Sia and P. N. E. Nohuddin, "Comparing Machine Learning Models for Heart Disease Prediction," *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, 2022, pp. 1-5, doi: 10.1109/IICAIET55139.2022.9936861.

[11] A. Singh and A. Jain, "Prediction of Heart Disease using Dense Neural Network," *2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT)*, New Delhi, India, 2022, pp. 1-5, doi: 10.1109/GlobConPT57482.2022.9938354.

[12] I. Sutedja, "Descriptive and Predictive Analysis on Heart Disease with Machine Learning and Deep Learning," *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, Makasar, Indonesia, 2021, pp. 1-6, doi: 10.1109/ICORIS52787.2021.9649585.

[13] N. Sabri *et al*., "HeartInspect: Heart Disease Prediction of an Individual Using Naïve Bayes Algorithm," *2023 IEEE 11th Conference on Systems, Process & Control (ICSPC)*, Malacca, Malaysia, 2023, pp. 350-354, doi: 10.1109/ICSPC59664.2023.10420149.

[14] 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.

[15] N. Mohan, V. Jain and G. Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms," *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2021, pp. 1-3, doi: 10.1109/ISCON52037.2021.9702314.

[16] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in *IEEE Access*, vol. 8,pp.107562107582,2020,doi:10.1109/ACCESS.2020.3001149.