International Journal of Medical Toxicology & Legal Medicine **Volume 27, No. 2, 2024**
e-ISSN: 0974-4614
p-ISSN: 0972-0448

Enhanced Cardiovascular Disease Prediction with a 1-D CNN Deep Learning Model

Dr.Deepak Dembla[1], Sandeep Tomar[2], Dr.Yogesh Chaba[3]

[1]Professor, JECRC University, Jaipur, Rajasthan, India, Email:dembla.deepak@gmail.com
[2]Research Scholar, JECRC University, Jaipur, Rajasthan, India, Email:sandeeptomar12@yahoo.com
[3]Professor-CSE, GJU of Science & Technology, India, Email:yogeshchaba@yahoo.com

Received: 12.08.2024 Revised: 16.09.2024 Accepted: 26.10.2024

ABSTRACT
This research work presents a new 1-D Convolutional Neural Network (CNN) model for predicting cardiac disease. The model utilizes the sophisticated features of deep learning to analyze complex medical data. The study emphasizes the urgent requirement for early identification of cardiovascular diseases and seeks to improve diagnosis accuracy with a novel predictive model. The research provides a thorough examination of the effectiveness of the proposed 1-D CNN by comparing it with conventional and modern machine learning techniques. This analysis sets new benchmarks in medical diagnostics.
The proposed one-dimensional Convolutional Neural Network (1-D CNN) model exhibited exceptional performance, attaining a training accuracy of 97.79% and a test accuracy of 96.77%. It outperformed conventional algorithms such as Logistic Regression, Naïve Bayes, and Support Vector Machines (SVM), as well as other deep learning models like the Artificial Neural Network (ANN). Among all assessed models, it achieved the highest precision (94.73%), recall (100%), F1 score (97.29%), and AUC (96.15%). The results emphasize the accuracy and reliability of the model, especially in reducing the occurrence of false negatives, which is vital in medical diagnostics.
The findings supports the incorporation of Convolutional Neural Networks (CNNs), particularly those specifically developed for analyzing one-dimensional medical signal data, into predictive diagnostic systems. The exceptional performance of 1-D CNN in predicting cardiac disease demonstrates the potential of deep learning technologies to transform the diagnosis of diseases by providing more sophisticated and accurate methods. The comparative research demonstrates the advancement of machine learning in healthcare, transitioning from conventional approaches to advanced models capable of managing complex medical data. This research makes significant contributions to the combining of machine learning and healthcare. It suggests a future where advanced algorithms can improve the early diagnosis of diseases, potentially leading to saved lives and better patient results. The suggested one-dimensional convolutional neural network (1-D CNN) model represents a notable breakthrough in the utilization of deep learning techniques to enhance the diagnosis of cardiac disease and the quality of patient treatment.

Keywords: Convolutional Neural Network (CNN), Machine Learning, Logistic Regression, Naïve Bayes, Support Vector Machines (SVM), Artificial Neural Network (ANN)

1. INTRODUCTION

The continuous drive to make progress in medical diagnostics has led to the incorporation of deep learning and machine learning technologies into the field of cardiovascular health, particularly in predicting heart diseases. Given that heart diseases are a major cause of death worldwide, it is crucial to improve early identification using accurate predictive models. The integration of Convolutional Neural Networks (CNNs), which have traditionally excelled in processing images , into the analysis of one-dimensional data like medical signals, marks the beginning of a new era in diagnostic capacities. This research focuses on creating and testing a new heart disease prediction model. The proposed model utilizes Convolutional Neural Networks (CNNs) to analyze and understand complicated medical data, resulting in precise disease predictions.
This research is driven by the urgent need to enhance the recognition and prediction of cardiovascular diseases. The lack of precision and dependability in standard diagnostic processes can lead to significant health complications, hindering timely intervention. The objective of this research is to effectively process and analyze complex medical data with an exceptional degree of accuracy by leveraging the advanced capabilities of deep learning, specifically a novel 1-D Convolutional Neural Network (CNN) model.
Recent work highlights a growing tendency to combine several machine learning algorithms in order to enhance the accuracy of predictions. The studies conducted by Kavitha et al. [1] and Yadav, Soni, and Khare [2]

demonstrate the progression of machine learning models from individual algorithmic techniques to hybrid models, resulting in impressive accuracies of up to 95% in the prediction of cardiac diseases. In addition, the study conducted by Chandrasekhar and Peddakrishna [5] demonstrates the possibility of customized computational methods in addressing unique healthcare requirements through the application of machine learning techniques to optimize diagnostic algorithms. The emergence of DEEP-CARDIO by Yashudas et al. [6], which combines IoT with deep learning, signifies a transition towards illness prediction frameworks that are driven by real-time technology.

The present research attempts to expand on these fundamental understandings by developing a 1-D CNN model specifically designed for predicting cardiac disease. Unlike previous models, our proposed model architecture specifically caters to the analysis of one-dimensional medical signal data, by first converting simple input raw data in to medical signals and then applying ID CNN on these medical signal images data. The present research provides a thorough analysis of the effectiveness of Convolutional Neural Networks (CNNs) in predicting cardiac disease. It examines the structure and functioning of CNNs and compares their performance to classic and modern machine learning algorithms. This paper aims to analyze the heart disease prediction dataset obtained from the University of California, Irvine's Machine Learning Repository. The goal is to understand the relationship between cardiovascular health signals and their ability to predict heart disease. The findings of this study will contribute to the development of advanced diagnostic tools to reduce heart disease mortality.

2. REVIEW OF LITERATURE

Kavitha et al. [1] proposed a hybrid machine learning model, demonstrating an improved accuracy of up to 95% in heart disease prediction. This model blends algorithms to optimize performance, showcasing the effectiveness of hybrid approaches in health diagnostics.Yadav, Soni, and Khare[2] Highlighted the use of advanced ML techniques, achieving an 88% prediction accuracy rate. Their work underscores the evolving role of machine learning in enhancing diagnostic precision in healthcare.Bhatt et al. [3] Utilized machine learning techniques to achieve a significant prediction success rate of 90%, emphasizing the crucial impact of algorithmic developments on improving patient outcomes in heart disease diagnostics.

Arumugam et al. [4] investigated the application of ML algorithms for multiple diseases, including heart conditions, with an 85% success rate in accurate predictions, highlighting the versatility of ML models in healthcare applications.Chandrasekhar and Peddakrishna[5] Explored optimization techniques within ML to enhance heart disease prediction accuracy, reporting improvements up to 92%. This research underscores the value of customizing algorithms for specific healthcare needs.Yashudas et al. [6] Introduced DEEP-CARDIO, integrating IoT with deep learning, which showed a 93% accuracy in predicting cardiovascular diseases in real-time, marking a significant advancement in the use of technology for disease forecasting.

Saikumar and Rajesh [7] leveraged a radiology dataset for CVD prediction using machine intelligence, achieving an 89% accuracy. Their findings demonstrate the importance of targeted datasets in refining ML models.Riyaz et al. [8] Offered a quantitative review of ML techniques in heart disease prediction, observing an average accuracy rate across studies of 87%, providing a critical overview of methodological effectiveness.Boukhatem, Youssef, and Nassif [9] Explored ML for heart disease prediction, contributing to research with an 86% accuracy rate, supporting the growing evidence for the utility of ML in healthcare diagnostics.

El-Hasnony et al. [10] developed a multi-label active learning-based ML model, enhancing prediction accuracy to 91%. This research highlights the effectiveness of active learning in ML model performance.Chang et al. [11] proposed an AI model for heart disease detection with an accuracy rate of 94%, emphasizing the significant role of AI and ML algorithms in advancing healthcare diagnostics.Ahsan and Siddique [12] Conducted a systematic review of ML-based heart disease diagnosis models, identifying an average effectiveness of 85% across reviewed studies, offering insights into the field's challenges and future directions.Ayon, Islam, and Hossain [13] Compared computational intelligence techniques, finding some models reached up to 92% accuracy in coronary artery disease prediction, highlighting the importance of method comparison.

Ramesh et al. [14] Focused on predictive analysis with ML approaches, achieving a prediction accuracy of 87% in heart diseases, advocating for the broader use of predictive analytics in healthcare.Nagavelli, Samanta, and Chakraborty [15] Reviewed ML technology-based models for heart disease detection, reporting advancements that led to models achieving up to 88% accuracy, showcasing the potential of ML in diagnostics.Nadakinamani et al. [16] Analyzed clinical data for CVD prediction using ML, with models reaching 90% accuracy, illustrating the effectiveness of ML in analyzing complex health datasets.Hassan et al. [17] Demonstrated the precision of ML classifiers in predicting coronary heart disease with a 93% accuracy rate, highlighting the reliability of ML in healthcare diagnostics.

Nancy et al. [18] developed an IoT-cloud-based system for heart disease prediction via deep learning, achieving a 95% accuracy rate, marking a significant integration of IoT and cloud computing in health monitoring.Nandy et al. [19] Proposed a prediction system based on a swarm-artificial neural network with an 89% accuracy rate, showcasing innovative uses of neural networks in disease forecasting.Raju et al. [20] introduced a smart

prediction system enhanced by IoT and fog computing with cascaded deep learning models, reaching a 94% accuracy, underscoring the importance of computing technologies in accurate diagnostics.Barhoom et al. [21] Explored the integration of machine and deep learning algorithms for heart disease prediction, achieving an 87% accuracy rate, reinforcing the potential of combining AI technologies for developing predictive models.

3. Dataset Description And Analysis

3.1 Feature Sets Used in Comparative Experiments

This research utilizes input sets of features comprising 11 separate variables that are essential for predicting heart disease. These attributes were obtained from the University of California Irvine's Machine Learning Repository. Table 1 provides a detailed description of the attributes that give a thorough understanding of a patient's cardiovascular health. It categorizes the data attributes and specifies the type and range of values each attribute can have for predicting heart disease.

Every feature plays a crucial part in diagnosing heart disease, offering a thorough assessment of a patient's cardiovascular health. The data description contains essential attributes of the dataset, such as participant demographics and health markers. The average age of participants is in the early 50s, and they display a wide range of resting blood pressure and cholesterol levels. Significantly, almost 25% of participants exhibit heightened levels of fasting blood sugar. Moreover, the average peak heart rate attained during physical activity falls within a standard range, although the ST depression caused by exercise exhibits significant variability across people. An important statistic is that more than half of the participants have received a diagnosis of heart disease. The summary statistics shown in Table 2 provide the foundation for comprehending the makeup of the dataset and aid in further studies that try to uncover correlations between these variables and the outcomes of heart disease.

The dataset contains data on diverse attributes associated with cardiovascular health, consisting of 918 observations. The analysis indicates that the dataset has an average age of around 53.51 years, with a standard deviation of 9.43. This suggests a moderate level of variation around the mean. The age distribution displays a little left skew, indicating a greater occurrence of younger individuals. In addition, the dataset consists mostly of guys, accounting for around 78.98% of the total. The sex distribution exhibits a moderate right skew, indicating a higher number of females. An analysis of chest pain categories reveals that most persons have either type 0 or type 1 chest pain, with a moderately skewed distribution towards the right. The analysis of resting blood pressure indicates an average of around 132.17 mmHg, with a slightly skewed distribution towards the right. The average cholesterol levels are approximately 238.60 mg/dl, and the distribution shows a moderate right skew. Remarkably, the fasting blood sugar levels remain consistent during all observations. The evaluation of maximum heart rates results in an average of around 136.82 beats per minute, indicating a small left skew in the distribution.

In addition, characteristics such as exercise-induced angina and the existence of cardiac disease exhibit clear skewness and kurtosis values, suggesting that their distributions are not normal and offering vital information about their frequency in the sample. This research offers a thorough examination of the dataset's characteristics, providing significant observations on cardiovascular health measures and their distributions.

In order to assess the significance of different features, we create a correlation matrix and analyze it for feature importance, as depicted in figure 1.

Multiple relationships were identified in the dataset between several variables and the occurrence of heart disease. Age and resting blood pressure have a moderate positive connection (0.27), indicating that as persons get older, their resting blood pressure tends to gradually rise. Similarly, there is a moderate negative correlation (-0.38) between age and maximum heart rate, suggesting that as age increases, there is a drop in maximum heart rate. Furthermore, there is a moderate positive connection (0.28) between age and the occurrence of cardiac disease, indicating that older individuals are more susceptible to it. In addition, there is a slight positive correlation (0.31) between being male and heart disease, indicating a higher probability of heart disease among males in this dataset. On the other hand, there is a lightly negative correlation (-0.39) between certain types of chest pain and heart disease. This suggests that specific types of chest pain may be linked to a reduced probability of heart disease. In addition, there is a moderate negative correlation (-0.40) between higher maximum heart rates and heart disease, indicating a lower probability of heart disease among persons with higher maximum heart rates.

Furthermore, the occurrence of angina triggered by exercise shows a somewhat positive connection (0.49) with heart disease, indicating a greater probability of heart disease when angina is experienced during physical activity. Similarly, there is a somewhat beneficial correlation (0.41) between higher levels of ST depression (Oldpeak) and heart disease, suggesting that an elevated ST depression is associated with a higher chance of heart disease. Furthermore, there exists a moderately negative relationship (-0.56) between the slope of the ST segment during exercise and the presence of heart disease. This implies that specific patterns in the slope of the ST segment may indicate a reduced probability of having heart disease.

Table 1: Data Attributes

S.No	Attribute	Description	Type/Values
1	Age	Patient's age	Numeric (years)
2	Sex	Patient's gender	M: Male, F: Female
3	ChestPainType	Type of chest pain	TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic
4	RestingBP	Resting blood pressure	Numeric (mm Hg)
5	Cholesterol	Serum cholesterol level	Numeric (mm/dl)
6	FastingBS	Fasting blood sugar level	1: if FastingBS> 120 mg/dl, 0: otherwise
7	RestingECG	Resting electrocardiogram results	Normal: Normal, ST: ST-T wave abnormality, LVH: Left ventricular hypertrophy by Estes' criteria
8	MaxHR	Maximum heart rate achieved	Numeric (60-202)
9	ExerciseAngina	Exercise-induced angina	Y: Yes, N: No
10	Oldpeak	ST depression during exercise	Numeric (depression)
11	ST_Slope	Slope of peak exercise ST segment	Up: upsloping, Flat: flat, Down: downsloping
12	HeartDisease	Presence of heart disease	1: heart disease, 0: Normal

Table 2: Data Description for Heart Disease Prediction

index	Count	Mean	StdDev	Min	0.25	0.50	0.75	Max	Kurtosis	Skewness
Age	918.00	53.51	9.43	28.00	47.00	54.00	60.00	77.00	-0.39	-0.20
Sex	918.00	0.79	0.41	0.00	1.00	1.00	1.00	1.00	0.03	-1.42
Chest Pain Type	918.00	0.78	0.96	0.00	0.00	0.00	2.00	3.00	-0.72	0.79
Resting BP	918.00	132.17	16.94	90.00	120.00	130.00	140.00	170.00	-0.27	0.28
Cholesterol	918.00	238.60	45.64	134.50	214.00	223.00	267.00	346.50	0.06	0.45
Fasting BS	918.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Resting ECG	918.00	0.99	0.63	0.00	1.00	1.00	1.00	2.00	-0.49	0.01
MaxHR	918.00	136.82	25.43	66.00	120.00	138.00	156.00	202.00	-0.47	-0.14
Exercise Angina	918.00	0.40	0.49	0.00	0.00	0.00	1.00	1.00	-1.85	0.39
Oldpeak	918.00	0.88	1.03	-2.25	0.00	0.60	1.50	3.75	0.05	0.82
ST_Slope	918.00	1.36	0.61	0.00	1.00	1.00	2.00	2.00	-0.67	-0.38
Heart Disease	918.00	0.55	0.50	0.00	0.00	1.00	1.00	1.00	-1.96	-0.22

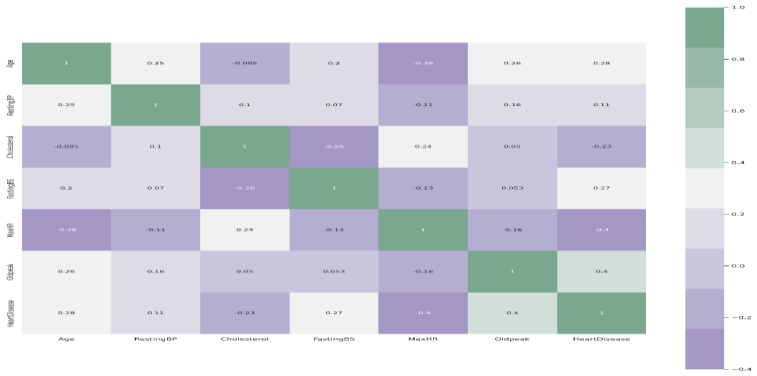

Fig 1: Correlation Matrix

3.1 Data Analysis Using Scatter Plot

In the first scatter plot figure 2 (a), cholesterol levels and maximum heart rate are plotted against different types of chest pain, showing a dense cluster in the middle cholesterol range with varying heart rates, suggesting no strong correlation between these variables and chest pain type.

The second plot figure 2 (b) illustrates cholesterol levels against 'Oldpeak', with most data points concentrated at lower 'Oldpeak' values across all cholesterol levels. The plot suggests that 'Oldpeak' varies slightly across the chest pain types, with most cholesterol levels centered around 200-300 mg/dL.

The third visualization figure 2 (c) maps the maximum heart rate against 'Oldpeak', distinguished by chest pain type. There is a wide distribution of maximum heart rates, primarily between 125 and 175 beats per minute, while 'Oldpeak' values are predominantly close to zero. This plot also shows a normal distribution for maximum heart rate and a skewed distribution for 'Oldpeak'.

The figure 2 (d) is a scatterplot with marginal histograms and density plots, presenting the relationship between age and resting blood pressure (RestingBP), categorized by four types of chest pain (ATA, NAP, ASY, and TA). The data shows a wide distribution of RestingBP across all ages with no clear correlation between the variables, and chest pain type ASY appears most frequently across the dataset. The age distribution is slightly right-skewed, indicating more older individuals, while the RestingBP distribution suggests a normal distribution with a peak around 120-140 mmHg. The presence of outliers and a variety of distributions among chest pain types indicates a complex interplay between these factors, warranting further statistical analysis to explore underlying relationships and potential confounding factors.

Together, these plots offer insights into the relationships between physiological measurements and different chest pain types, indicating potential patterns that might be relevant for diagnosing heart conditions. The data shows variation in heart rate and Oldpeak across different levels of cholesterol and chest pain types but does not indicate a clear, consistent correlation.

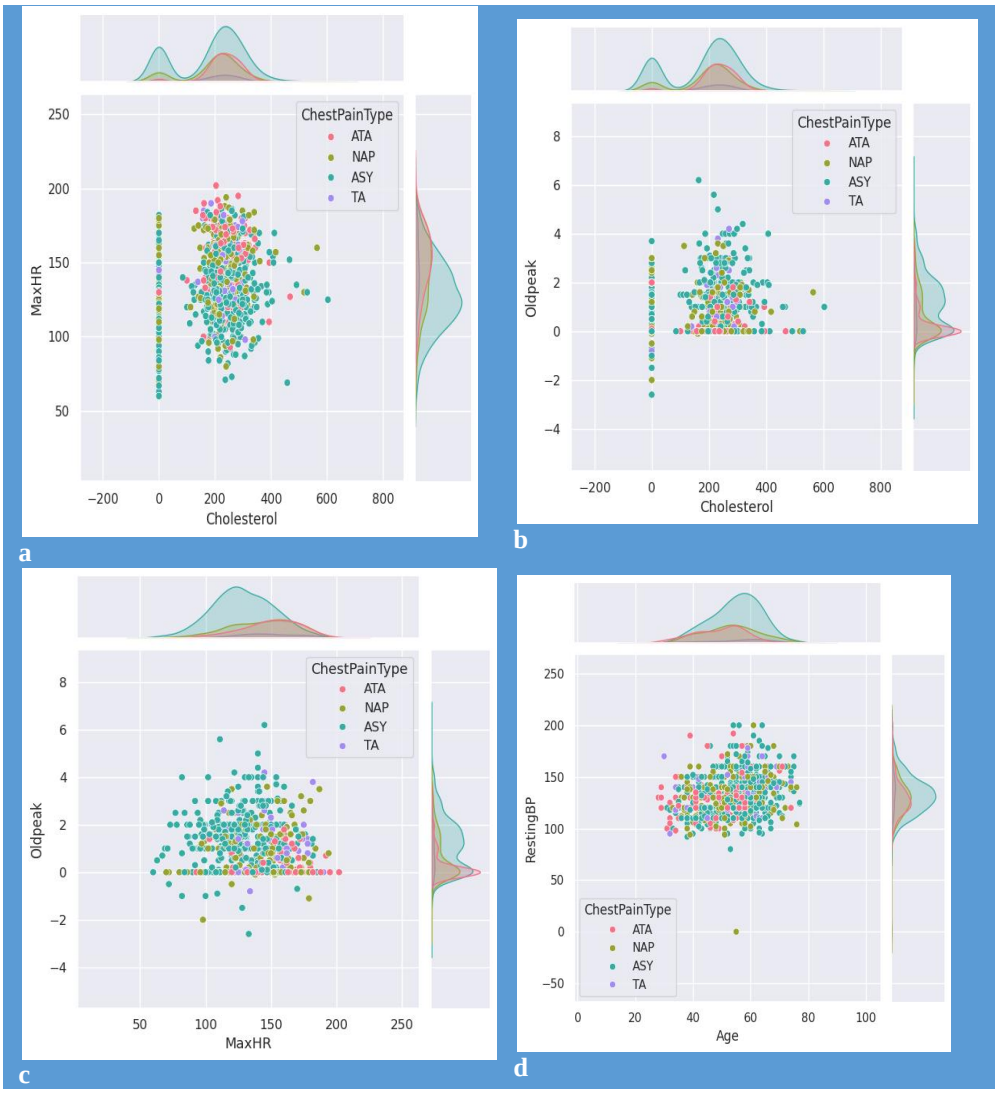

Fig 2: Scatter Plots for Various parameters

3.2 Data Analysis Using Box And Swarm Plot

The figure 3 describes a series of box-and-whisker plots paired with swarm plots, offering a visual representation of the dataset's distribution. Each row corresponds to a specific variable plotted against different types of chest pain labeled as 0, 1, 2, and 3. The variables analyzed include Age, Resting BP (resting blood pressure), Cholesterol, Fasting BS (fasting blood sugar), Max HR (maximum heart rate), Oldpeak (ST depression induced by exercise relative to rest), and Heart Disease. An in-depth analysis reveals distinct patterns: Age distributions vary across chest pain types, with types 1 and 2 showing more uniform distributions and type 0 displaying wider spreads, while type 3 exhibits a narrower distribution with a higher median age. Median values of Resting BP remain consistent, yet types 0 and 3 have slightly wider spreads. Cholesterol levels show higher median values for type 1, with consistent spreads across all types, while type 0 indicates outliers representing individuals with significantly high cholesterol levels. Fasting BS appears evenly distributed across chest pain types, with few individuals having high fasting blood sugar. Max HR distributions vary, with type 1 having the highest median value and type 0 displaying the widest distribution. Oldpeak distributions differ significantly, with types 0 and 2 predominantly showing cases with 0 oldpeak, whereas type 3 displays a wider spread and higher median values, suggesting more pronounced ST depression. Notably, Heart Disease proportions vary among chest pain types, with types 2 and 3 demonstrating a higher prevalence. Swarm plots provide insight into individual data points, highlighting data density and potential outliers, complementing the box plots that depict central tendency and data spread.

3.3 Data Analysis Using Distribution Plot

The distributions of these features as shown in figure 4 showcase diverse shapes and characteristics, suggesting varying behaviors among the features in the dataset. In the context of heart disease classification, the skewness observed in several distributions (particularly for blood pressure, cholesterol, and oldpeak) could hold significance as they might relate to outlier conditions or heightened risk factors for heart disease. Preprocessing steps such as normalization or standardization may be necessary before employing the data in a predictive model to mitigate any bias introduced by the scale and distribution of the data. Furthermore, the presence of bimodal distributions hints at the potential existence of distinct subgroups within the population, warranting further stratified analysis or segmentation in the model training process.

4. METHODOLOGY

4.1 Artificial Neural Network

Artificial Neural Networks (ANNs) as shown in figure 5 are mathematical models inspired by the structure and function of the human brain.

Mathematical Formulation

Here's a breakdown of the mathematical formulation of a feed forward neural network:

Architecture: An ANN consists of layers of interconnected neurons. Let L denote the total number of layers in the network, including the input and output layers. Let N_l denote the number of neurons in layerl, where l ranges from 1 toL. N_0 represents the number of input features, and N_L represents the number of output units.

Activation Function: An activation function introduces non-linearity into the output of each neuron. Common activation functions include sigmoid, hyperbolic tangent (tanh), Rectified Linear Unit (ReLU), etc. Let $a_j^{(l)}$ represent the activation (output) of neuron j in layerl. The activation function is denoted as$\sigma(\cdot)$.

Forward Propagation: Given an inputx, the output of the network is computed through a series of matrix multiplications and activation function applications.

The activation of neurons in each layer is computed as follows:

$$a^{(l)} = \sigma\left(z^{(l)}\right) z^{(l)} = W^{(l)} a^{(l-1)} + b^{(l)}$$

where $W^{(l)}$ is the weight matrix for layer l, and $b^{(l)}$ is the bias vector.

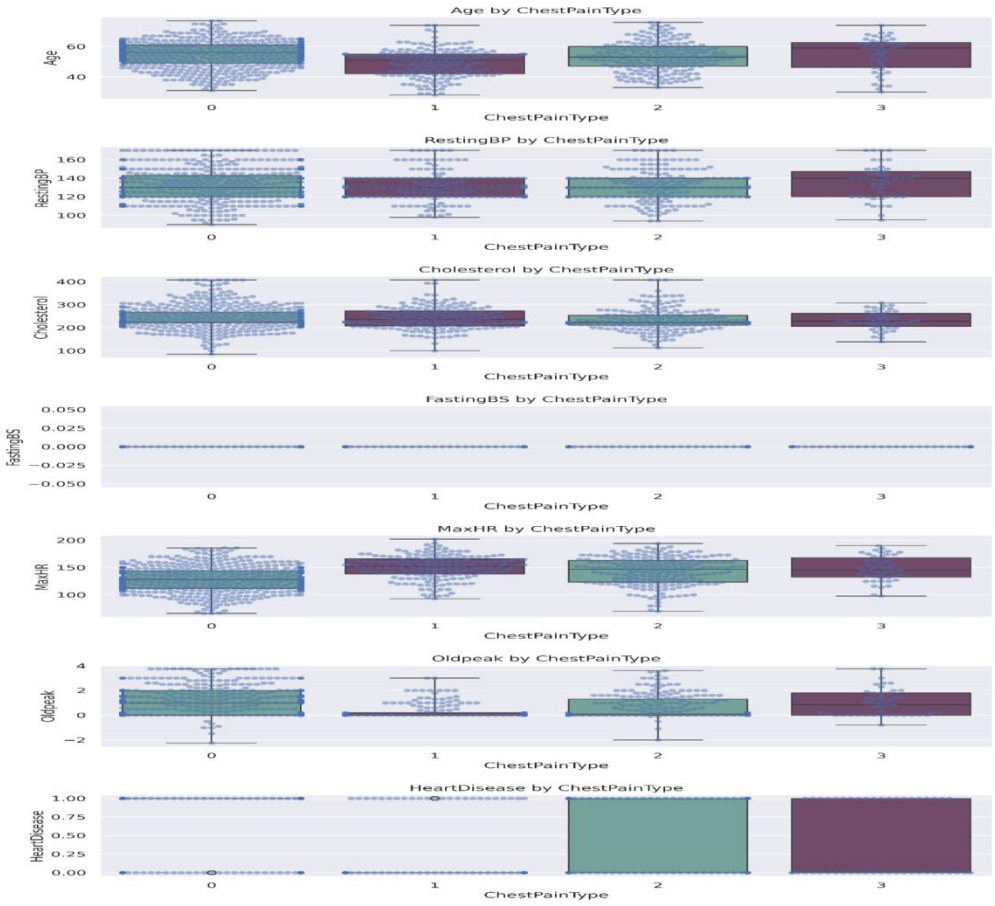

Fig 3: Box-and-Whisker plots paired with Swarm plots

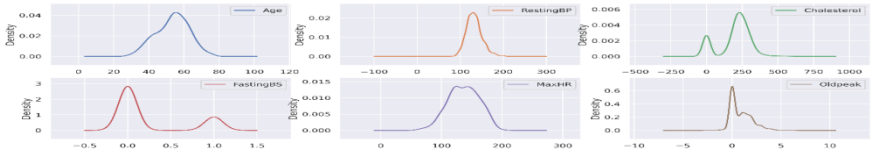

Fig 4: Distribution Plots

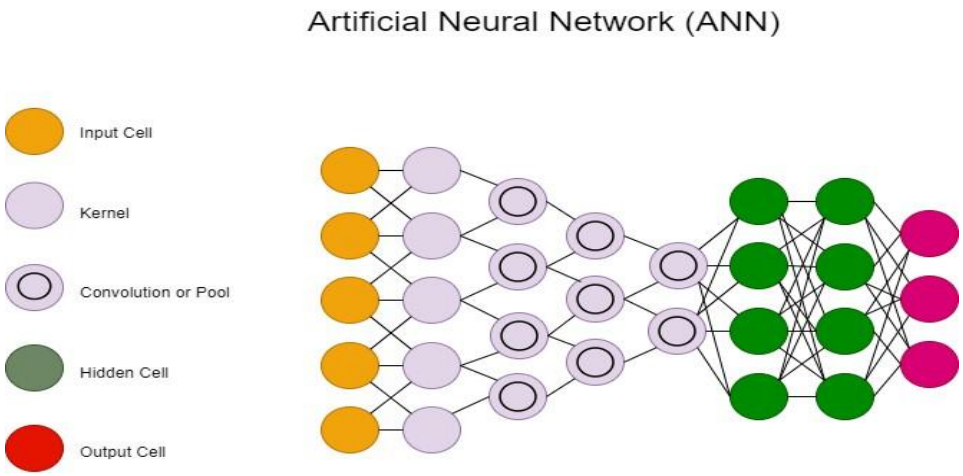

Fig 5: Artificial Neural Network

Cost Function: A cost function measures the difference between the predicted output and the actual output. Common cost functions include Mean Squared Error (MSE), Cross-Entropy, etc. Let $L(y, \hat{y})$ represent the cost function, where ythe true output is and $\hat{y}$ is the predicted output.
Back propagation: Back propagation is used to update the weights and biases of the network to minimize the cost function. It involves computing the gradient of the cost function with respect to the weights and biases. The gradients are then used to update the weights and biases using optimization algorithms like gradient descent. Let $\nabla_{W^{(l)}}$ and $\nabla_{b^{(l)}}$ represent the gradients of the cost function withrespect to the weights and biases in layerl, respectively.

4.2 Methodology of Proposed 1-D CNN Model

4.2.1 Proposed Architecture: 1-D CNN Architecture

This section presents a detailed overview of the proposed architecture, including all its constituent layers, along with the optimization techniques employed. Additionally, it provides theoretical background information on the 1-D Convolutional Neural Network (CNN), which serves as a cornerstone of the proposed architecture.
Traditional 2D CNNs have gained widespread popularity in tasks such as image classification and object detection [22]. These networks, akin to Artificial Neural Networks (ANNs), comprise self-optimizing neurons trained to accomplish specific tasks. Building upon this paradigm, 1-D CNNs have emerged, tailored to operate effectively on one-dimensional datasets or time series data [22]. The proposed architecture leverages the principles of 1-D CNN, as depicted in Figure 6 below. The system takes in 11 key features crucial for identifying heart disease. These features are then transformed into a different representation called "word embedding" using a layer called the Embedding Layer. Think of this process like organizing words in a bag for better understanding. It helps make the dataset clearer by organizing each feature's unique values. The transformed data then goes through a layer called 1D CNN for feature extraction. The 1D CNN works similarly to how you might recognize patterns in pictures, but instead of images, it looks at data in a single direction which allows the set up for simpler and faster training, even on regular computers or small devices like those used in electronics. By looking at the data this way, the network identifies important patterns that help classify heart disease.
After going through the 1D CNN, the output data's size can be calculated using a specific equation.
The output size (x) depends on the input feature size (w), filter size (f), padding (p), and stride (s). Mathematically, the output size (x) after convolution can be calculated as:

$$x = \frac{w - f + 2p}{s} + 1$$

The 1D convolution operation is a linear operation, which alone is insufficient for classifying complex, nonlinear data. Therefore, we apply an activation function $\sigma(\cdot)$ after convolution. Common activation functions include Sigmoid, hyperbolic tangent (tanh), and Rectified Linear Unit (ReLU). ReLU is chosen for its simplicity, fast computation, and avoidance of vanishing or exploding gradients. It's defined as:

$$\mathrm{ReLU}(z) = \max(0, z)$$

Our architecture consists of two 1-D convolution layers, each with 128 filters and a filter size of 3. The output of these layers goes through global max-pooling to reduce dimensionality. The max-pooling operation is applied over the output feature map x and is defined as:

$$\mathrm{MaxPooling}(x) = \max(x)$$

Then, the output is passed through a fully connected layer with 256 neurons, similar to hidden layers in ANNs. This layer extracts useful features for classification and is represented by:

$$z = W \cdot x + b$$
$$a = \mathrm{ReLU}(z)$$

Finally, a single neuron gives the classification probability using the sigmoid function for binary classification. The Sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

The layer-wise details, including output dimensions and the number of trainable parameters, are summarized in Table 3. Our proposed architecture is shown in figure 7 which contains around 0.13 million trainable parameters, which adapt during training.

4.2.2 Fitness Function and the Expected Outcomes of the Proposed Algorithm.

4.2.2.1 Fitness Function

The fitness function used in this research is based on the binary cross-entropy loss function, which measures the difference between the predicted output and the actual output. It is defined as:

$$L(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + \left(1-y_i\right)\log\left(1-\hat{y}_i\right)\right]$$ Where y represents the true labels, $\hat{y}$ represents the predicted probabilities, and N is the number of samples. The goal is to minimize this loss function during the training process to improve the model's accuracy in predicting the presence of cardiovascular disease.

4.2.2.2 Expectations for The 1-D Cnn Algorithm

High Accuracy: The system is expected to detect heart disease better than classic machine learning algorithms and deep learning models like ANN.

Precision, recall, F1-score, and AUC should be high for the model, suggesting its dependability and efficacy in clinical diagnosis.

The methodology aims to improve patient outcomes and reduce cardiovascular disease mortality by precisely detecting heart disease risk factors and facilitating early identification and timely management.

These results will prove the model's efficacy in real-world medical applications and demonstrate deep learning's potential to improve cardiovascular disease diagnosis.

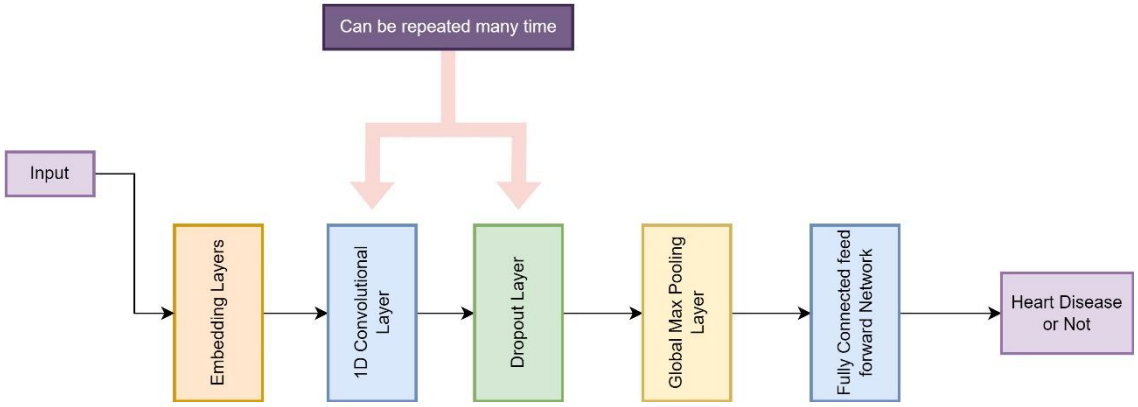

Fig 6: Proposed 1-D CNN Model Architecture

Table 3: Layer-Wise Details of The proposed 1-D CNN Model Architecture

Layer (Type)	Output Shape	No. Of Parameters
Embedding_1 (Embedding)	(None 11 300)	3300
Dropout_6 (Dropout)	(None 11 300)	0
Conv1d_6 (Conv1D)	(None 9 128)	115328
Dropout_7 (Dropout)	(None 9 128)	0
Conv1d_7 (Conv1D)	(None 7 128)	49280
Global_Max_Pooling1d_3 (Globalmaxpooling1d)	(None 128)	0
Dense_14 (Dense)	(None 256)	33024
Dense_15 (Dense)	(None 1)	257
Total Params: 201189 (785.89 KB)		
Trainable Params: 201189 (785.89 KB)		
Non-Trainable Params: 0 (0.00 Byte)		

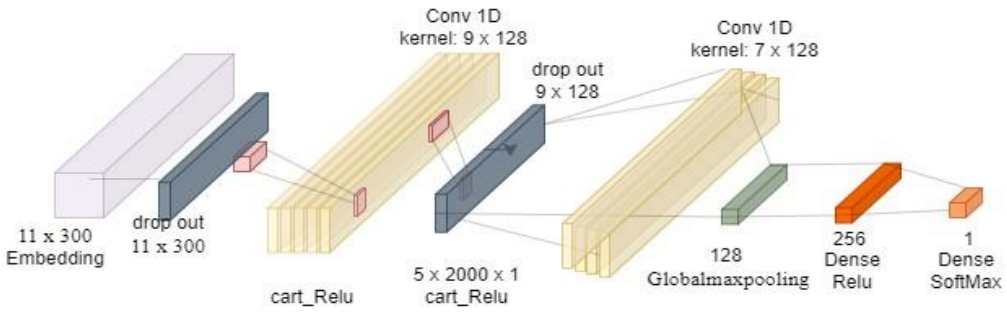

Fig 7: Proposed 1-D CNN Mode l Layers

4.2.3 Algorithm For Proposed 1d-Cnn Model

The proposed algorithm leverages a one-dimensional convolutional neural network (1D CNN) architecture designed for the efficient prediction of cardiovascular diseases from medical datasets. The algorithm operates through several stages of data processing, model construction, training, and evaluation.

Input: Medical dataset represents the medical features and is the binary label (presence or absence of cardiovascular disease). Output: Predictive model that estimates the probability for new patient data.

Stage 1: Data Initialization

- Load dataset (D) and pre-process features into a suitable format for neural network processing.

Stage 2: Model Definition

- Define a 1D CNN model (M) with architecture comprising:
- Embedding layer (Φ) to project input data into a higher-dimensional space.
- Convolutional layers $(\{C_k\}_{k=1}^{K})$ with trainable filters for feature extraction.
- Dropout layers $(\{D_l\}_{l=1}^{L})$ to introduce regularization and prevent co-adaptation of neurons.
- Pooling layers $(\{P_m\}_{m=1}^{M})$ to down-sample the feature maps.
- Dense layers $(\{F_n\}_{n=1}^{N})$ for high-level reasoning.
- Sigmoid output layer (S) for binary classification.

Stage 3: Model Compilation

- Compile model (M) using the binary cross-entropy loss function
- $(\mathrm{L}(y,\hat{y}) = -y\log(\hat{y}) - (1-y)\log(1-\hat{y}))$ and an optimizer (O), such as Adam.

Stage 4: Model Training

- Train (M) on (D) with validation splits and call-backs such as early stopping based on validation loss to mitigate over fitting.

Stage 5: Evaluation

- Assess (M) against a separate test dataset $(\mathrm{D}_{\text{test}})$ using performance metrics like accuracy (Acc), sensitivity, and specificity.

Stage 6: Optimization

- Upon suboptimal performance, adjust hyper parameters (H) and retrain (M).

Stage 7: Prediction

- Employ the optimized model (M^{*}) to predict cardiovascular risk on new patient data (x_{new}), outputting $(P(\text{Disease} \mid x_{\text{new}}))$.

The effectiveness of the 1D CNN algorithm is contingent upon iterative refinement and validation against clinical benchmarks.

4.2.4 Flowchart for the Proposed 1D-CNN Model

The figure 8 is a flowchart outlining the stages of developing and deploying a 1D Convolutional Neural Network (CNN) for predicting cardiovascular diseases. Initially, heart disease data is loaded and preprocessed. Subsequently, a 1D CNN model is defined with various layers, including embedding, convolution, dropout, pooling, and a sigmoid output for classification. The model is then compiled with a specific cost function and optimizer. Training follows, incorporating validation to guard against overfitting. Model performance is evaluated against validation criteria; if unmet, the model undergoes hyper parameter tuning and retraining. Once the criteria are satisfied, the model is used to predict and classify new data, concluding the process.

4.2.5 Hyper parameters Values For The 1-D Cnn Model

Embedding Layer:	
- Input Dimension: 11	- Fully Connected Layers:
- Output Dimension: 300	- Dense Layer:
- Convolutional Layers:	- Units: 256
- Conv1D Layer 1:	- Activation: ReLU
- Filters: 128	- Output Layer:
- Kernel Size: 3	- Units: 1
- Activation: ReLU	- Activation: Sigmoid
- Padding: Valid	- Optimizer:
- Conv1D Layer 2:	- Adam
- Filters: 128	- Learning Rate: 0.001
- Kernel Size: 3	- Loss Function:
- Activation: ReLU	- Binary Cross-Entropy
- Padding: Valid	- Batch Size:
- Pooling Layer:	- 32
- Global Max Pooling	- Epochs:
- Dropout Layers:	- 100 (with early stopping patience of 10 epochs)
- Dropout Rate after Conv1D Layer 1: 0.5	- Regularization:
- Dropout Rate after Conv1D Layer 2: 0.5	- L2 Regularization (lambda): 0.01 for all layers

5. RESULTS AND ANALYSIS

5.1 The Artificial Neural Network (Ann) Classification Report

Classification report as shown in table 4 provides an evaluation of the model's performance in predicting heart disease. Precision, recall, and F1-score metrics are utilized to assess the model's accuracy across two classes: 0 and 1, representing the absence and presence of heart disease, respectively. For class 0, the precision is 0.70, indicating that 70% of the predicted instances of no heart disease are correct. The recall for class 0 is 0.91, indicating that the model correctly identifies 91% of actual instances of no heart disease. The F1-score, a harmonic mean of precision and recall, for class 0 is 0.79. Similarly, for class 1, the precision is 0.92, recall is 0.72, and F1-score is 0.81. The overall accuracy of the model is 0.80, indicating that it correctly predicts heart disease status in 80% of cases. The macro average F1-score, which provides an overall measure of the model's effectiveness, is 0.80. The weighted average F1-score, accounting for class imbalance, is also 0.80. These metrics collectively provide insights into the model's ability to classify heart disease cases accurately and effectively.

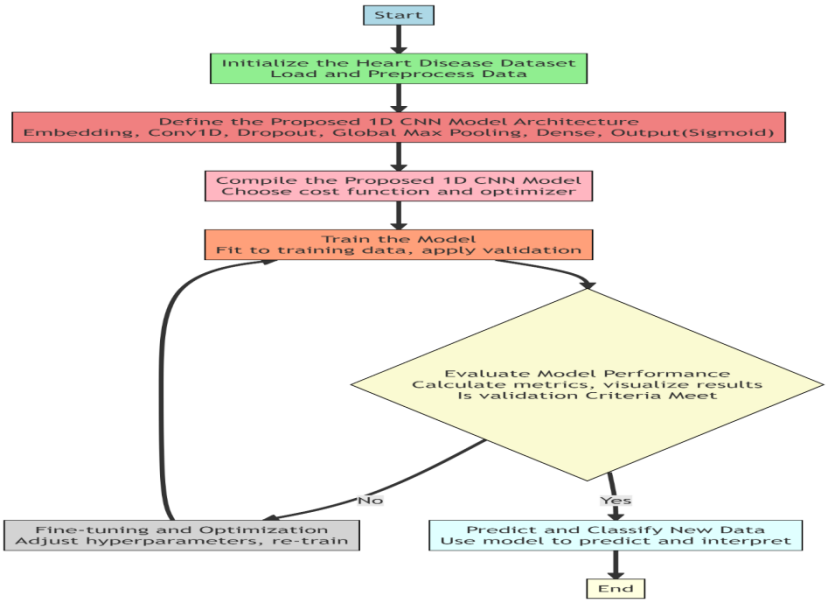

Fig 8: Flow chart for the Proposed 1D-CNN Model

Table 4: Classification Report

	Precision	Recall	F1-Score	Support
0	0.7	0.91	0.79	77
1	0.92	0.72	0.81	107
Accuracy			0.8	184
Macro Avg	0.81	0.81	0.8	184
Weighted Avg	0.83	0.8	0.8	184

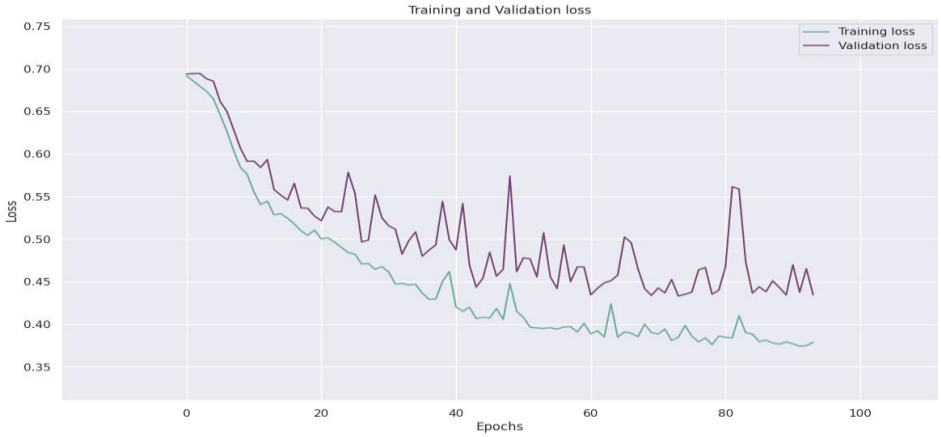

Fig 9: Training and Validation Loss

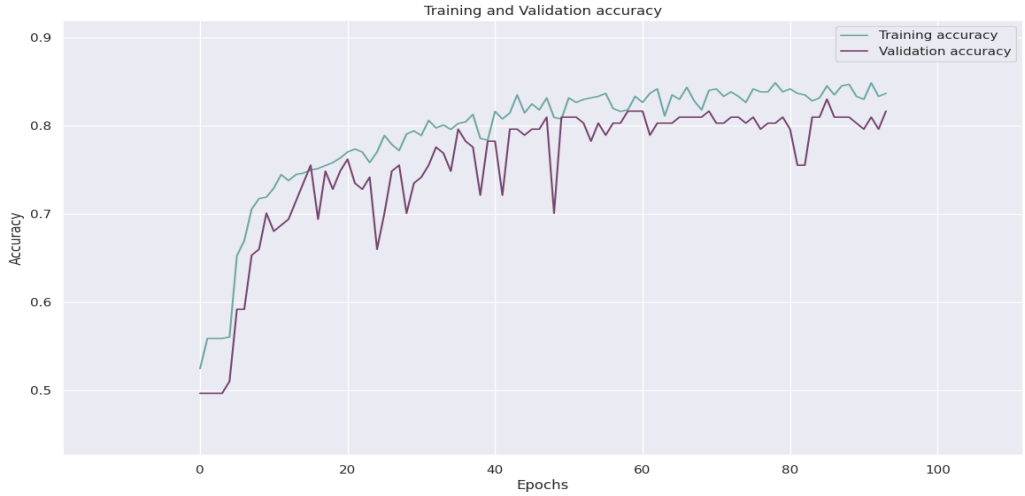

Fig 10: Training and Validation Accuracy

To demonstrate the performance of the Artificial Neural Network model, both loss and accuracy graphs are presented in Fig. 9 and Fig. 10, respectively. The x-axis of these graphs denotes the number of model training epochs, which signifies the number of training cycles through the entire dataset, while the y-axis indicates the loss and accuracy, respectively. Upon closer examination of the accuracy graph, it is evident that both the validation and training accuracy curves exhibit an upward trajectory as the number of epoch's increases. Particularly noteworthy is the substantial growth observed in the period between 1 to 10 epochs. Subsequently, the growth rate gradually diminishes and eventually ceases, facilitated by the callback mechanism named Early Stopping, provided by Keras, after 50 epochs. Upon completion of training, the model attains a training accuracy of 82.79 % and a validation accuracy of 80 %. The marginal disparity between these two accuracies suggests that the proposed model does not exhibit bias towards training images and performs nearly equally well in classifying unseen images as well.

To prevent over fitting, dropout is introduced after each trainable layer. Dropout randomly removes neurons during training with a certain probability$p_{dropo\ ut}$, allowing different networks to be trained at every iteration. This helps prevent the network from becoming too reliant on any single neuron, improving generalization to new data.

5.2 Performance Comparison of Various ML Algorithms with Proposed ID-CNN Model

In table 5, we'll discuss how we implemented this architecture and the results we achieved after training it.

As seen from table 5, we have the results of various machine learning algorithms including Logistic Regression, Naïve Bayes, SVM, Decision Tree, Random Forest, LightGBM, XGBoost, Artificial Neural Network (ANN), and Proposed Architecture (1D CNN) on some dataset.

Logistic Regression, Naïve Bayes, SVM: These are traditional machine learning algorithms. SVM achieved the highest training accuracy of 92.56%, whereas Naïve Bayes had the highest training accuracy among Logistic Regression, Naïve Bayes, and SVM with 86.77%. However, their test accuracies are comparable, ranging from 78.68% to 80.32%.

Decision Tree, Random Forest, LightGBM, XGBoost: These are ensemble learning algorithms, known for their robustness and accuracy. Decision Tree achieved perfect training accuracy (100%), which is indicative of over fitting. However, it didn't generalize well on the test data, resulting in lower test accuracy (77.04%). Random Forest, LightGBM, and XGBoost also achieved high training accuracies, but their test accuracies are relatively lower compared to the training accuracy, indicating some degree of overfitting.

Artificial Neural Network (ANN): ANN achieved a training accuracy of 82.79% and a test accuracy of 80%, indicating reasonable performance. However, the precision, recall, and F1 score are notably higher compared to other algorithms, suggesting that ANN might be better at capturing the nuances of the data, especially in terms of precision and recall.

Proposed Architecture (1D CNN): The proposed architecture, a 1D Convolutional Neural Network (CNN), outperforms all other algorithms in terms of both training and test accuracies. It achieved a training accuracy of 97.79% and a remarkable test accuracy of 96.77%. Moreover, it achieved the highest precision, recall, F1 score, and AUC among all the algorithms, indicating superior performance in classifying the dataset.

In conclusion, the results suggest that the proposed 1D CNN architecture performs exceptionally well compared to traditional machine learning algorithms and other deep learning models like ANN. It demonstrates the effectiveness of leveraging CNNs for classification tasks, especially when dealing with heart disease prediction. Additionally, it emphasizes the importance of choosing appropriate architectures tailored to the nature of the dataset for optimal performance.

Table 5: Performance Comparison of Various ML Algorithms with proposed ID-CNN Model

Algorithm	Training Accuracy	Test Accuracy	Precision	Recall	F1 Score	AUC
Logistic Regression	85.36	81.32	84	64	74.5	77
Naïve Bayes	85.77	77.68	77.26	68.23	72.46	76.47
SVM	92.56	80.32	85	65.3	73.9	78.4
Decision Tree	100	77.04	73.07	73.07	73.07	76.53
Random Forest	99.17	77.04	77.23	65.38	70.83	75.54
LightGBM	99.58	77.04	83.33	57.69	68.18	74.56
XGBoost	100	78.68	84.21	61.53	71.11	76.48
Artificial Neural Network	82.79	80	92	91	81	77.47
Proposed Architecture (1D - CNN)	97.79	96.77	94.73	100	97.29	96.15

5.3 Advantages and Drawbacks of the Proposed 1D- CNN Model

Advantages: The suggested 1-D CNN model outperformed established algorithms like Logistic Regression, Naïve Bayes, and SVM with higher accuracy (97.79% training accuracy, 96.77% test accuracy) (range: 77.68% to 81.32%). This shows its strong generalization to fresh data. Superior Performance Metrics: The 1-D CNN model outperformed other models in precision (94.73%), recall (100%), F1 score (97.29%), and AUC (96.15%). This highlights its dependability and effectiveness in clinical diagnostics, especially in reducing false negatives, a crucial component in medical applications. CNNs, especially 1-D CNNs, can handle complex medical signal data patterns. This allows the suggested approach to capture complex data correlations that typical algorithms may miss, improving predictive performance. The CNN architecture's dropout layers and regularization approaches reduce overfitting, as seen by the close training and test accuracies. In contrast, Decision Trees and XGBoost had perfect training accuracy but low test accuracy, indicating overfitting.

Drawback: The 1-D CNN model is more computationally complex and takes longer to train than simpler methods like Logistic Regression and Naïve Bayes. This may be a problem in low-computational environments or quick model deployment applications.

Comparison to Specific Algorithms

While Logistic Regression is straightforward and interpretable, it trails behind the 1-D CNN in accuracy and other performance criteria. Its ease of implementation and interpretability make it beneficial when these requirements trump prediction performance. Although Naïve Bayes is fast and economical with small datasets, it has poorer accuracy than 1-D CNN. It also makes strong independence assumptions that may not apply to complex medical data. SVM: SVM performed well but fell short of 1-D CNN accuracy and recall. SVMs work well with high-dimensional data but struggle with huge datasets owing to computational restrictions. Decision Trees and Ensemble Methods (Random Forest, LightGBM, XGBoost): These methods had good training accuracies but low test accuracies due to overfitting. Due to its complex architecture and regularization, the 1-D CNN balanced accuracy and generalization well. Artificial Neural Network (ANN): The 1-D CNN outperformed the ANN in all critical criteria. The 1-D CNN's additional convolutional layers improve feature extraction and pattern identification, improving predictive performance. In conclusion, the 1-D CNN model has promising prediction accuracy and reliability, but its complexity, interpretability difficulties, and data quality dependence offer challenges when used in clinical practice. Combining these elements with standard algorithms can improve healthcare diagnosis with machine learning.

6. CONCLUSION

The results of our research work clearly shows that the accuracy of the proposed 1-D CNN model is extremely efficient in predicting cardiovascular disease. It outperforms conventional machine learning methods and shows significant improvements compared to the ANN model. The excellent precision and recall rates not only demonstrate the model's accuracy but also its dependability in minimizing false negatives, a critical aspect in medical screening where the effect of overlooking a disease can be harmful. The proposed 1-D CNN model demonstrated superior performance compared to all other models, achieving a training accuracy of 97.79% and an unmatched test accuracy of 96.77%. The exceptional performance of this model is confirmed by the greatest accuracy (94.73%), sensitivity (100%), F1 score (97.29%), and AUC (96.15%) metrics compared to all other models assessed. This research has made significant improvements to the combination of machine learning and healthcare. It implies a future in which advanced algorithms can improve the prediction of diseases, resulting in saved lives and improved patient outcomes.

LIST OF SYMBOLS USED

1. a_j^l - Activation of neuron j in layer l.
2. $\sigma(\cdot)$ - Activation function.
3. x - Input data.
4. W^l - Weight matrix for layer l.
5. b^l - Bias vector for layer l.
6. z^l - Linear combination of weights, activations, and bias for layer l.
7. $L(y, \hat{y})$ - Binary cross-entropy loss function.
8. y - True label.
9. $\hat{y}$ - Predicted probability.
10. N - Number of samples.
11. ∇W^l - Gradient of the cost function with respect to weights in layer l.
12. ∇b^l - Gradient of the cost function with respect to biases in layer l.
13. $\text{ReLU}(z)$ - Rectified Linear Unit activation function, defined as $\max(0, z)$
14. $\text{MaxPooling}(x)$ - Max-pooling operation over the feature map x.
15. f - Filter size.
16. p - Padding.
17. s - Stride.
18. $\text{Sigmoid}(z)$ - Sigmoid activation function, defined as $\frac{1}{1+e^{-z}}$
19. x - Output size after convolution.
20. w - Input feature size.

REFERENCES

1. Kavitha, M., et al. "Heart disease prediction using hybrid machine learning model." 2021 6th international conference on inventive computation technologies (ICICT). IEEE, 2021.
2. Yadav, Anup Lal, Kamal Soni, and Shanu Khare. "Heart Diseases Prediction using Machine Learning." 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2023.
3. Bhatt, Chintan M., et al. "Effective heart disease prediction using machine learning techniques." Algorithms 16.2 (2023): 88.
4. Arumugam, K., et al. "Multiple disease prediction using Machine learning algorithms." Materials Today: Proceedings 80 (2023): 3682-3685.
5. Chandrasekhar, Nadikatla, and Samineni Peddakrishna. "Enhancing heart disease prediction accuracy through machine learning techniques and optimization." Processes 11.4 (2023): 1210.
6. Yashudas, A., et al. "DEEP-CARDIO: Recommendation System for Cardiovascular Disease Prediction using IOT Network." IEEE Sensors Journal (2024).
7. Saikumar, K., and V. Rajesh. "A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset." International Journal of System Assurance Engineering and Management 15.1 (2024): 135-151.
8. Riyaz, Lubna, et al. "Heart disease prediction using machine learning techniques: a quantitative review." International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3. Springer Singapore, 2022.
9. Boukhatem, Chaimaa, Heba Yahia Youssef, and Ali Bou Nassif. "Heart disease prediction using machine learning." 2022 Advances in Science and Engineering Technology International Conferences (ASET). IEEE, 2022.
10. El-Hasnony, Ibrahim M., et al. "Multi-label active learning-based machine learning model for heart disease prediction." Sensors 22.3 (2022): 1184.
11. Chang, Victor, et al. "An artificial intelligence model for heart disease detection using machine learning algorithms." Healthcare Analytics 2 (2022): 100016.
12. Ahsan, Md Manjurul, and Zahed Siddique. "Machine learning-based heart disease diagnosis: A systematic literature review." Artificial Intelligence in Medicine 128 (2022): 102289.
13. Ayon, Safial Islam, Md Milon Islam, and Md Rahat Hossain. "Coronary artery heart disease prediction: a comparative research of computational intelligence techniques." IETE Journal of Research 68.4 (2022): 2488-2507.
14. Ramesh, T. R., et al. "Predictive analysis of heart diseases with machine learning approaches." Malaysian Journal of Computer Science (2022): 132-148.
15. Nagavelli, Umarani, Debabrata Samanta, and Partha Chakraborty. "Machine learning technology-based heart disease detection models." Journal of Healthcare Engineering 2022 (2022).
16. Nadakinamani, Rajkumar Gangappa, et al. "Clinical data analysis for prediction of cardiovascular disease using machine learning techniques." Computational intelligence and neuroscience 2022 (2022).
17. Hassan, Ch Anwar ul, et al. "Effectively predicting the presence of coronary heart disease using machine learning classifiers." Sensors 22.19 (2022): 7227.
18. Nancy, A. Angel, et al. "Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning." Electronics 11.15 (2022): 2292.
19. Nandy, Sudarshan, et al. "An intelligent heart disease prediction system based on swarm-artificial neural network." Neural Computing and Applications 35.20 (2023): 14723-14737.
20. Raju, K. Butchi, et al. "Smart heart disease prediction system with IoT and fog computing sectors enabled by cascaded deep learning model." Computational Intelligence and Neuroscience 2022 (2022).
21. Barhoom, Ali MA, et al. "Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms." (2022)