Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

An Efficient Deep Learning Technique for Heart Disease Prediction

Oviya K[1], Deepika G[2], Hemavathi V[3], Prema P M. M.E[4].,

[1,2,3] Students,and [4] Faculty
Dept. of Information Technology
Engineering, Panimalar College of
Engineering, Chennai, India.
oviyakalidass93@gmail.com

Abstract-*Extensive research has been conducted on predictive algorithms that support advance detection and treatment of cardiovascular disease, which remains a major global health problem. In this research machine learning methods such as Support Vector Devices(SVD), Decision Tree, Random Forest and Logistic Regression are thoroughly analysed, and it is compared with the deep learning algorithm of Artificial Neural Networks (ANN) in the context of heart disease prediction. The accuracy and efficiency of these algorithms were thoroughly evaluated in this work using carefully selected data sets. The research identified an algorithm that displayed the highest level of accuracy in diagnosing heart disease through extensive testing and cross validation methods. As they elucidate the most effective computational techniques for anticipating cardiovascular illness, the outcomes of the research have important significance for medical practitioners.*

Keywords- ANN(Artificial Neural Networks) , Logistic Regression, DecisionTree ,SupportVector Devices(SVD) Random Forest.

1.INTRODUCTION

Accurate prediction and early heart disease detection is extremely important because cardiovascular disease has long been globally, it is one of the leading causes of death. Machine learning and Deeplearning made great strides in recent years in creating predictive models that support detection of cardiovascular disease . To determine which algorithm is most accurate at predicting heart disease, this study compares and contrasts four well- known Machine Learning algorithms which includes Random forests, Decision trees, logistic regression, support vector Devices (SVDs), with the Deep learning algorithm like ANN. The choice of algorithm is very important when building predictive models to identify heart disease, as it has a direct influence on the accuracy and reliability of the diagnostic results. It is important to test the algorithm's performance on a consistent dataset, as each algorithm has different strengths and weaknesses. To do this, the research uses a carefully selected dataset containing a variety of health- related parameters, aiming to accurately represent a wide range of patient health records. The research aimed to evaluate and compare the accuracy of various algorithms to determine the best predictive ability and algorithms were able to diagnose heart disease. In doing so, this research aims to make relevant contributions to the decision-making process of choosing the best heart disease prediction algorithm, helping physicians improve patient

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

outcomes and make informed decisions. This paper will review the approaches used for data preprocessing, algorithm selection, and performance evaluation in the subsequent sections of this research. The implementation of each method will be discussed in detail after results and comparisons are available.

2.METHOHOLOGY

a)DATA COLLECTION:

For the early cardiovascular disease prediction in patients, the Cleveland Heart Disease dataset, sometimes referred to as the"Cleveland dataset," is a popular dataset in the fields of machine learning and healthcare. The clinical and demographic information in this dataset can be used to develop algorithms that predicts the existence of cardiovascular disease. Here are some details regarding the dataset: Dataset Origin: The UCI Machine Learning Repository initially gathered and made available the Cleveland Heart Disease dataset. It may be accessed by anybody for study and instruction. Description of the data the dataset includes 14 properties (features), which include category and numerical variables, and a total of 303 occurrences (patient records).

The dataset's most popular version contains the following characteristics:

1. Age: Age of patient.

2. Cp(Type of chest pain):The particular type of chest pain of patient (one of the following categories: 1, 2,3,4).

3. Fasting blood sugar (Fbs) :(0 = false,1 = true).

4. TRESTB(Resting blood pressure):values in mm/hg

5. Chol (Serum Cholesterol):values in mg/d).

6.Thalach (Heart Rate Achieved maximum):Values from 1 to 3

7.Oldpeak (Activity Relative to Rest is induced by ST Depression): values

8.Slope: Slope of the ST section of the peak workout (2-flat,1-up,3-down)

9. CA (Number of vessels): values from 0to 3

10. Thal (Thalassemia): type 7-Reversable3-normal, 6- fixed defect.

11. Sex: Patient’s gender (0 – females and 1Male)

12. Exan (Exercise angina): 0 - yes, 1-no

13.Target(fu=inal result): 1-presence of Disease,0- normal.

14.Restec(Electrocardiographic resting Results): 2-HT ,0- Normal,1-ST-T

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

Purpose: Building prediction algorithms that can divide patients into two groups— those with and without cardiac disease—is the main goal of this dataset. This dataset is used by academics and data scientists to create and assess machine learning algorithms for prediction of heart disease

Availability: The Cleveland cardio Disease dataset is freely accessible for research and instruction since it is available in many machine learning libraries and repositories. It is frequently employed in research projects, academic courses, and tutorials on healthcare analytics and predictive modelling.

b) SPLIT DATA:

A crucial step in developing and accessing machine learning models is dividing the C level and Heart Disease at a set into training, validation, and test sets for heart disease prediction. The data splitting procedure is detailed below:

1.data Loading: Open your favourite data analysis programme(such as Python with p and as or R) and import the C level and Heart Disease dataset into it. Make sure the dataset is organized into the right columns and has an organized format.

2.Data pretreatment: Before dividing the data, perform the necessary data preparation tasks. A few instances of this include filling in blanks, encoding category variables (if any), and scaling or normalizing numerical traits. If the target variable is not already in binary values (0 for no heart disease and 1 for heart illness), you should convert it to that format.

1. Shuffling: Make sure the dataset is randomized and not arranged according to any particular standards. The data split is less prone to bias thanks to shuffling.

2. Dividing the Data
   a. Training Set (80%): Provide the majorityof the data to the training set . Using this set of data, your machine learning model is trained.

   b. Validation Set (10% to 15%): From there maintain data, construct a validation set. The Hyper parameters of the model are adjusted by the validation set and keep track of how well it performs throughout training.

   c. Test Set (20%): Set aside a specific chunk of datafor thetest set. After training and a hyper parameter adjustment, the test set is used to assess the final results of your model.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

b. PROPOSED WORK AND ALGORITHM

The collection of data and decision-making process for the most crucial attributes is the first stage of system operation. After pre-processing, the patient data is converted to the desired format. Afterward, training and test data are separated from the data. Models are trained using algorithms, and the system's accuracy is assessed using test data. For this system's implementation, the following modules are used Data gathering, Attribute selection, Pre-processing, Data balance, and Disease Prediction

MACHINE LEARNING MODEL:

This research has chosen machine learning and deep learning approaches to build a model for predicting heart disease. Details of these methods are as follows:

a) Logistic Regression:

Heart disease prediction relies heavily on a fundamental machine learning process known as logistic regression .As part of this research work, Logistic Regression is an important method for achieving an accurate diagnosis. Because it excels at binary classification problems, this method is an excellent choice for identifying people with and without heart disease. The relationship between the input data and the likelihood of a binary outcome, such as the presence or absence of heart disease, is predicted by logistic regression.

b)Support Vector Device:

A potent machine learning technology called support vector machines (SVMs) is used for predicting heart disease using imported datasets. SVMs excel in this task because they can divide data into different groups and handle feature spaces of large size. SVM is excellent at identifying complex patterns in medical data when used to predict heart disease, helping doctors more accurately identify patients at risk. In addition, their core technique allows SVM to detect small indicators of risk that other algorithms might miss by capturing complex non-linear correlations in the data

c)Decision Tree

A machine learning strategy that is applied to both classification and regression issues are called decision tree. It represents a flowchart-like structure, with each inner node representing a decision based on a particular characteristic, leading to child nodes representing possible outcomes and further decisions. Decision trees are popular for their simplicity the system is characterized by its ease of interpretation and ability to handle both categorical and numerical data by recursively partitioning it based on the most significant features with the goal of creating unique and homogeneous subsets that enable accurate predictions.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

d) Random Forest

Applications The machine learning method known as Random Forest is used in both classification and regression. During training, it is created by combining several decision trees in order to improve prediction accuracy and reduce overfitting. Aggregating the findings from each individual tree in prediction yields the output (classification or regression), which is frequently decided by major vote (for categorization)or mean (for regression). The robustness, adaptability, and capacity ofrandom forests to manage complicated datasets and high-dimensional feature spaces are well known.

Deep Learning Models

a) Artificial neural networks

The deep learning technique uses neural networks to evaluate complex medical data and forecast a person's chance of developing heart disease. This algorithm is particularly designed to manage the features of age, gender, blood pressure, cholesterol levels, and other input elements. Heart disease is predicted using a sophisticated computational framework called deep learning based on artificial neural networks (ANN).intricate patterns in medical data and make accurate predictions about whether a person will develop heart disease. This algorithm's multilayered neural network design, which was inspired by the interconnected neurons in the human brain making it suitable for processing a variety of input features, including patient medical history and lifestyle aspects.. As data moves through different stages, the system uncovers intricate relationships between input variables and the prevalence of cardiac disease. This learning is performed through iterative training on a dataset of labelled samples. Once trained, the ANN can quickly evaluate new patient data and predict the likelihood of heart illness, frequently as a probability score or binary classification.

4. ARCHITECTURE

The process begins with collecting patient details, including parameters like age, chest pain type, cholesterol levels, and maximum heart rate. These details form the dataset, which undergoes preprocessing to clean the data, handle missing values and outliers, normalize or standardize for consistency, and select the most relevant features for heart disease prediction. Once preprocessing is complete, the prepared dataset is used to train several machine learning and deep learning models: Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Artificial Neural Networks (ANN). Each algorithm learns patterns and relationships indicative of heart disease from the pre-processed data. After training, the models are used to make predictions about the likelihood of heart disease in patients. The prediction phase involves evaluating the trained models using a separate test set to ensure their performance is reliable and generalizes well to unseen data. Key performance metrics such as accuracy, precision, recall, and F1 score are used to measure the

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

accuracy of the predictions. These metrics help in determining how well each algorithm performs and in identifying the most effective model for heart disease prediction.

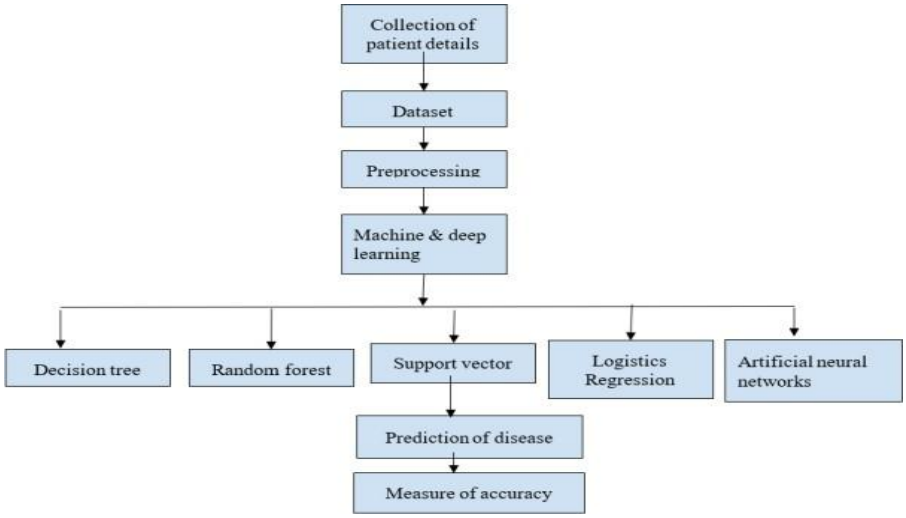


FIG.1.Architecture diagram.

5. FLOW CHART

1. Here We Collect raw data stored in a CSV file containing various features relevant to predicting heart illness (e.g., age, cp, chol, thalach).

2. Preprocess the raw data to prepare it for model training. This includes data cleaning, normalization/standardization, and feature selection.

3. Split the pre-processed data into two parts :Training set (70-80% of the data) for training the machine learning model and Test set (20-30% of the data) for evaluating the model's performance.

4. Train the machine learning model using the training set. This involves selecting an appropriate algorithm (e.g., logistic regression, decision tree, support vector device) and fitting it to the training data.

5. Evaluate the trained model using the test set. This includes making predictions on the test data and assessing performance using metrics such as accuracy, precision, recall, and F1 score to ensure the model generalizes well to unseen data and is not overfitting.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

6. Use the trained and evaluated model to make predictions on new data or the test set. The output provides the predicted likelihood or classification of heart illness, offering actionable insights based on the model's analysis.

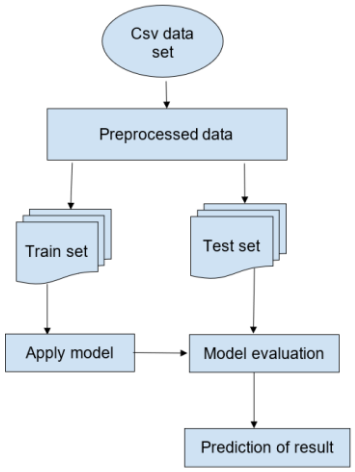


FIG.2.Flow Chart Diagram

6. PICTORIAL REPRESENTATION

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix, accuracy_score
from keras.models import Sequential
from keras.layers import Dense, Dropout
import matplotlib.pyplot as p

# Load the dataset
df = pd.read_csv('heart.csv')

# Handle missing values
df = df.replace('?', np.nan)
df = df.dropna()

# Select features and target
X = df.iloc[:, :-1].values
y = df.iloc[:, -1].values

# Binarize the target variable
y = (y > 0).astype(int)

# Split the dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Feature scaling
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

FIG.3

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14

```
# Build the ANN model
model = Sequential()

# Input layer and first hidden layer with dropout
model.add(Dense(units=32, activation='relu', input_dim=X_train.shape[1]))
model.add(Dropout(0.2))

# Second hidden layer with dropout
model.add(Dense(units=16, activation='relu'))
model.add(Dropout(0.2))

# Third hidden layer with dropout
model.add(Dense(units=8, activation='relu'))
model.add(Dropout(0.2))

# Output layer
model.add(Dense(units=1, activation='sigmoid'))

# Compile the model
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Train the model
history = model.fit(X_train, y_train, epochs=150, batch_size=10, validation_split=0.1)

# Evaluate the model
y_pred = (model.predict(X_test) > 0.5).astype(int)

accuracy = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
```

FIG.4

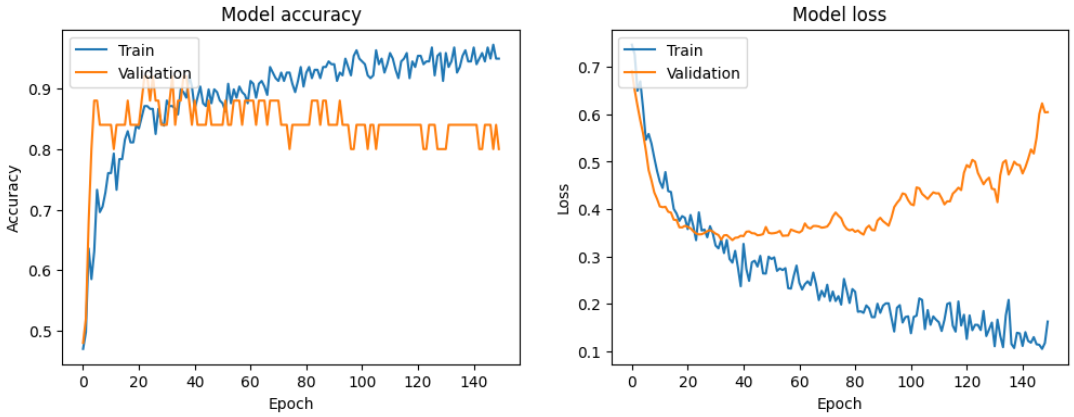
```
print("Accuracy:accuracy")
print("Confusion Matrix: \n", cm)

# Plot training & validation accuracy values
plt.figure(figsize=(12, 4))
plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'])
plt.plot(history.history['val_accuracy'])
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Train', 'Validation'], loc='upper left')

# Plot training & validation loss values
plt.subplot(1, 2, 2)
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Validation'], loc='upper left')

plt.show()
```

FIG.5

7. SAMPLE OUTPUT:

```
Accuracy:0.988888
Confusion Matrix:
 [25  4]
 [ 6 26]]
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
Volume7-Issue2, January 2024
Paper: 14



8 .CONCLUSION

The findings of this researched insight on the capabilities of different machine learning and deep learning algorithms, offering an important contribution to the field of heart disease prediction. The study additionally emphasizes the potential of predictive analytics in the medical industry, where rapid diagnosis and therapy may significantly enhance patient outcomes and the medical system. This paper emphasized the value of ongoing research toimproveandenhancethecapabilitiesofmachine learning algorithms for medical applications as we draw to a close. In order to further improve accuracy and interpretability, future work might investigate ensemble methods that incorporate the advantages of several algorithms as well as feature engineering and selection strategies.

9. **REFERENCES**

[1].Sumit Sharma, Mahesh Parmar,” Heart Diseases Prediction using Deep LearningNeural Network Model,”2020 International Journal of Innovative Technology and Exploring Engineering (IJITEE).

[2]Abhay Kishore, Ajay Kumar, Karan Singh, ManinderPunia, YogitaHambir”Heart Attack Prediction Using Deep Learning”, International ResearchJournal of Engineering and Technology (IRJET)

[3]Rohit Bharti ,Aditya Khamparia , Mohammad Shabaz “Prediction of Heart Disease UsingCombination of Machine Learning and Deep Learning”,2021 in Computational Intelligence and Neuroscience.