Mini Project Report on

Heart Disease Predictor System Using Deep Learning

Submitted in partial fulfilment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by

Akanksha Dubey **University Roll No. 2016600**

Under the Mentorship of

Dr. Vikash Tripathi
Professor
Department of Computer Science and Engineering

Department of Computer Science and Engineering
Graphic Era (Deemed to be University)
Dehradun, Uttarakhand
January 2023

CANDIDATE'S DECLARATION

I here by certify that the work which is being presented in the project report entitled **"Heart Disease Predictor Using Deep Learning"** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Vikash Tripathi, Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Akanksha Dubey **University Roll no**.:2016600

signature

Table of Contents

Abstract

This report represents the mini project assigned to fifth-semester students for the partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun. Heart disease or Cardiovascular disease is an atypical function of the heart or blood vessels. It can lead to, an increase in risk for cardiac arrest, and sudden death. Heart Diseases are the most common cause of death worldwide over the last few decades in different parts of the world. Early detection of heart diseases and continuous supervision of doctors can reduce the risk of death or life loss that can happen because of late detection of heart diseases. However, it is not practical to monitor patients 24x7hrs in all scenarios accurately, and sometimes consultation with a patient for 24 hours by doctors is not available which may lead to a delay in the diagnosis of a disease which can delay treatment and hence increase the risk for the patient's health. In this project, I have researched and developed a model for heart disease prediction with an accuracy of 86% through various parameters such as age, sex, blood pressure rate, etc., and detected upcoming risk of heart disease using Multilayer Perceptron(MLP) neural network(Deep Learning Network) which have more than one layer of neurons. Learning of models in this type of network is usually done through the backpropagation error algorithm and Adam algorithm for optimization to handle noise problems on the dataset available publicly on the Kaggle website, further evaluating the results using a confusion matrix and cross-validation. This type of model for early diagnosis of heart-related problems can reduce the complications and can lead to an important contribution to the field of healthcare.

Keywords: *Deep Learning, Adam , Cross-Validation, Backward Elimination, Confusion Matrix, Cardiovascular Diseases.*

Chapter 1

Introduction

According to the World Health Organization, Cardiovascular Disease (CVDs) or heart disease is the leading cause of death globally. An estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths. Out of these deaths, 85% were due to heart attack and stroke. Heart disease is even considered a silent killer which leads to the death of a person without any serious symptoms. Most cardiovascular diseases can be prevented by addressing behavioral risk factors such as physical inactivity, sugar level, type of chest pains, and other common lifestyle parameters. It is important to detect the cardiovascular disease as early as possible to make decisions on lifestyle changes in high-risk patients and in turn, can reduce complications. This project aims to predict future Herat Disease by analyzing data of patients which classifies whether they have heart disease or not using deep learning algorithms.

1.1 Problem Definition

The most crucial challenge in heart disease is its early detection. For the early stage of heart disease detection, there are instruments available but either they are expensive or not efficient to calculate the chances of heart disease in humans. Early detection of heart disease can decrease the mortality rate and other complications in humans. However, it is not practical to monitor patients 24x7hrs in all scenarios accurately and sometimes consultation of a patient for 24 hours by doctors is not available which may leads in delay in the diagnosis of disease which can delay in treatment and hence increase the risk for the patients' health. Since we have a good amount of dataset in today's tech era, we can use various machine learning and deep learning algorithms to analyse the dataset for hidden patterns. These hidden patterns in dataset can be used for health diagnosis in healthcare dataset.

1.2 Motivation

In recent years, Deep learning is a new area of Machine Learning research regarded as a sophisticated and mathematically complex evolution of machine learning algorithm. The major motivation behind this research-based project was to explore the data preparation, processing, and features selections behind training models in the deep learning. Also, the healthcare sector has long been one of the prominent adopters of modern technologies to service itself. Hence this project is carried out with the motivation to explore the deep learning methods and further implement model to train the obtained data. Moreover, as the whole deep learning is motivated to develop an appropriate computer-based system and decision support that can help in early detection of heart disease. In this project I have developed a model which classifies if patient have a heart-related problem or not based on various features using multilayer perceptron network. This type of models for early diagnosis of heart related problems can reduce the complications and can leads in an important contribution in the field of healthcare.

1.3 Objectives

The main objective of developing this project are:

1.To develop deep learning model to predict possibility of heart disease.

2.To determine significant risk factors based on medical dataset which may lead to heart disease.

3.To analyse feature selection methods and understand their working principle.

Chapter 2

Literature Survey

With growing development in the medical science and health sector by the side of machine learning and deep learning various research and experiments have been carried out in recent years publishing relevant papers.

The paper [1] proposed by A H Chen, S Y Huang, P S Hong(Department of Medical Informatics, Tzu Chi University, Hualien, Taiwan) Heart disease prediction system In this paper, they develop a heart disease prediction system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. These approaches include three steps. Firstly, the selection of 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ECG, max heart rate, exercise-induced angina, old peak, slope, number of vessels colored, and thal. Secondly, developing an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction was 80%.

Hana H. Alalawi et. al. [2] used deep learning and machine algorithms to diagnose heart disease using a combination of two datasets which was collected from the Kaggle and Cleveland dataset for the heart. Using an Artificial Neural Network, he achieved an accuracy of 77%.

JayshrilS.Sonawane and D.R.Patil [3] Proposed another methodology for heart disease prediction. The network was trained using the technique called Vector Quantization Algorithm. They used three layers. The input layer has 13 neurons equal to the number of attributes in the dataset. The number of neurons in the hidden layer can be varied based on the error obtained. And finally, there is a single neuron in the output layer which indicates whether the heart disease is present or not [12]. The system's performance was improved by varying the epochs and the number of neurons. The result shows that the system can be obtained an accuracy of about 85.55%.

The main idea behind the proposed system after reviewing the above papers was to build an efficient heart disease prediction system grounded on the inputs as shown in Table 1. We analysed the classification algorithms namely Naive Bayes, Decision Tree, Random Forest, Logistic Regression, and ANN based on their Accuracy, Precision, Recall, and f-measure scores, and identified the best classification algorithm which can be used in heart disease prediction.

Chapter 3

Methodology

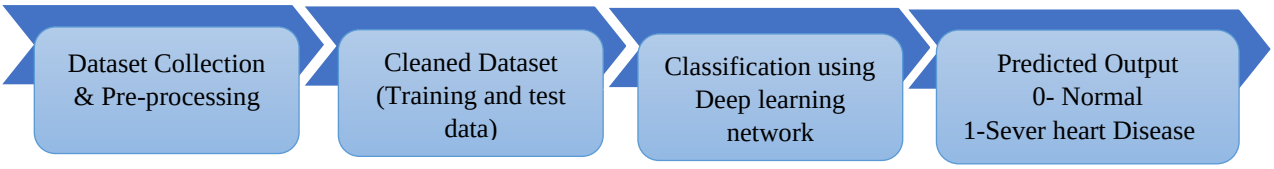

Figure 3.1*Methodology Followed for Heart Disease Prediction*

3.1 Dataset

It is the heart disease dataset from the UCI data repository. This is a multivariate dataset that provides or involves a statistical variable, multivariate numerical data analysis, and a variety of separate mathematical variables. It is composed of 14 attributes which are sex, age, resting blood pressure, chest pain type, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak — ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels and Thalassemia and target 0 or 1. The data set is in CSV (Comma Separated Value) format which is further prepared to data frame as supported by the pandas library in python.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
5	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
7	44	1	1	120	263	0	1	173	0	0.0	2	0	3	1
8	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
9	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1

Figure 3.2 *Original Dataset Snapshot*

3.2 Data Pre-processing

The Data pre-processing removes any ambiguous data from the collected dataset which can result in reduced accuracy and prediction rate. After the collection of the dataset, it is further pre-processed to remove missing or duplicate data from the dataset.

3.2 Methods and Functions Used

The main purpose of designing this prediction system is to predict the possibility of a person having some heart disease. We had used Multilayer Perceptron (MLPs) as a Deep Learning algorithm to train our model and various feature selection algorithms.

3.2.1 Deep Neural Network

The architecture of the model is a multilayer perceptron (feed-forward network) with a sequential model in which each layer is connected in a single input and single output manner. The heart-disease classification

attribute is a binary attribute that is classified as ‘1’ for patients having heart diseases, and ‘0’ for patients with the absence of heart disease. The model has 3 hidden dense layers. The 14 selected attributes were joined in the fully connected dense layers. The first dense layer contained 145 neurons, the next contained 120, and the last dense layer contained 70 neurons. The rectified linear unit(relu) was used as an activation function except for the last output layer. The sigmoid function was used as an activation function in the last output layer of our neural network model. The Adam optimizer was used with a learning rate of 0.001. Binary cross-entropy was set as a loss function. During the training phase, the number of epochs was set to 25.

3.2.2 Sigmoid Function

The sigmoid function is a mathematical function that has an S-shaped curve and can map any numerical value into a small range of numbers i.e., between 0 and 1. The sigmoid function is mostly used in neural networks, so it is also referred to as the logistic sigmoid function. It was used as the last output layer in our model to transform the outcome of our proposed model into a probability score. The formula for a sigmoid function and its graphical representation are described in Figure 3.2.

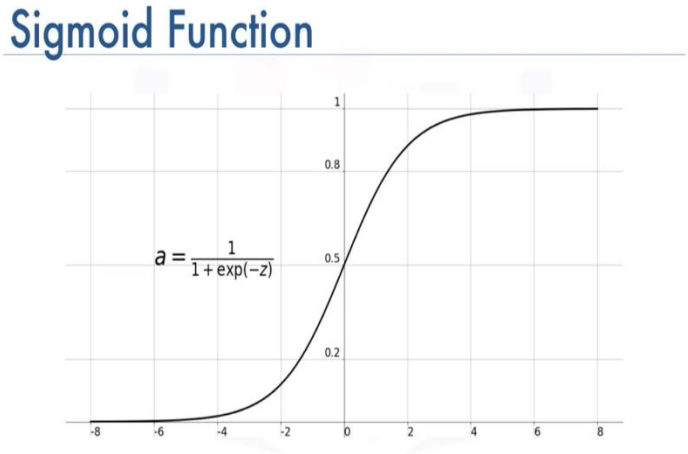

Figure 3.3 *Sigmoid Function Graph*

3.2.2 Adam Optimization Algorithm

Adam is learning rate optimization algorithm designed for training deep neural networks. Adam optimization is a stochastic gradient descent method that is based on adaptive learning. The learning rate of 0.001 was used with the optimizer. Method is efficient for the case when we have to work with large problem involving a lot of data and parameters.

Chapter 4

Result and Discussion

The dataset from the UCI repository was obtained from the Kaggle website and contained several attributes, listed in table 1, with a possible description combined by medical experts. Out of 76 attributes only 14 were selected. Before starting the experiment, the dataset needs to be pre-processed and prepared. The pre-processing step involves removing anomalies from the dataset by replacing missing values with mean values. Once the dataset is processed it will be split into training and testing data. The model classifies the data into two binary classes: 1 for the patient suffering from heart disease and 0 for the patient having no heart-related disease. The computed prediction accuracy by the model was around 86%.

4.1 Experimental Setup

The experimental setup of the model was established to discover the performance of the Deep Neural network with the dataset available .Table 1 shows the attributes used in the dataset and its description.

Sr No.	Attributes	Representation	Description	Type
1	Age	age	Age in years	Integer
2	Gender	sex	Male and female	Binary(1 for male and 0 for female)
3	Chest pain	cp	Four types of chest pain	Categorical
4	Cholesterol level	Chol	Measure of cholesterol in mg/dl	Integer
5	Resting blood pressure	trestbps	Blood pressure when the body is in a state of rest	Integer
6	Fasting blood sugar	fbs	Blood sugar level while fasting	Binary (1 for true and 0 for false)
7	MaxHR	thalach	Maximal heart rate	Integer
8	Rest ECG	restecg	Resting electrocardiograph	categorical
9	Exercise-induced angina	exang	Exercise-induced angina	Binary (1 for yes and 0 for no)
10	Old peak	oldpeak	ST depression brought by exercise comparative to rest	Continuous
11	Slope	slope	Slope of exercise peak	Discrete
12	Vessels	ca	No. of major vessels	Continuous
13	Thalassemia	thal	Normal, fixed, and reversible defects	discrete
14	Heart disease	target	Predicted attribute	Binary

Table 1 *Description of the attributes used in the dataset*

4.2 Performance Metrics

The performance of the proposed model could be evaluated by using some performance metrics. In deep learning, there are various standards for evaluating the performance of a system. Some performance metrics i.e., accuracy, precision, recall, and F1 score, are given in table 2.

Table 2.Performance metrics of the model

Performance Metrics	Accuracy(%)
Accuracy	85.55
Precision	86.88
Recall	86.75
F1 score	85.70

Chapter 5

Conclusion and Future Work

5.1 Conclusion

The early-stage prediction of heart or cardiovascular disease can prevent much life. The use of an efficient algorithm can help physicians or doctors in detecting the possible presence of heart disease before it can lead to serious health issues. This mini-project is based on a state-of-the-art UCI repository for the early detection of heart-related diseases. Initially, the dataset was collected-processed and then used with a Multilayer perceptron network algorithm for the prediction of heart diseases on jupyter notebook. The proposed system was evaluated regarding the performance metrics of accuracy, precision, recall, and F1 score, in which it achieved 85.55%, 86.88%, 86.75%, and 85.70% respectively.

5.2 Future Work

The current research mainly focused on the age group of 20 or above for heart disease prediction. The prediction models can be further improved to predict heart disease in children. Nowadays heart-related problems are common among children and hence there is a need for designing a model to predict heart diseases in children at an early stage so that they can get treatment as soon as possible this can decrease the death caused by heart-related problems among children.

References

[1] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng, and E. J. Lin, "HDPS: Heart disease prediction system," *2011 Computing in Cardiology*, 2011, pp. 557-560.

[2] Hana H. Alalawi , Manal S. Alsuwat, 2021, "Detection of Cardiovascular Disease using Machine Learning Classification Models", *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 10, Issue 07 (July 2021)*

[3] Sumit Sharma and Mahesh Parmar, "Heart Diseases Prediction using Deep Learning Neural Network Model" *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020*

[4] P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithyavishnupriya, "Heart Disease Prediction Using Deep Neural Network," *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 666-670, doi: 10.1109/ICICT48043.2020.9112443.

[5] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

[6] Heart Disease Dataset | Kaggle[online]

[8] Tensorflow [Online]: TensorFlow[online]

[9] Github repository akankshadubey-202/Heart-Disease-Prediction-Using-Deep-learning (github.com)