HEART DISEASES PREDICTION USING DEEP LEARNING

MAJOR PROJECT SYNOPSIS

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

by

Manak	Manaswi	Pankaj Kumar
06114802718	06214802718	07414802718

Guided by

Garima Gupta

Ms. Garima Gupta

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

MAHARAJA AGRASEN INSTITUTE OF TECHNOLOGY

(AFFILIATED TO GURU GOBIND SINGH INDERPRASTHA UNIVERSITY, DELHI)

(SESSION 2021-22)

INTRODUCTION

As we all know heart diseases or cardiovascular diseases (CVD) are increasing since past few years. CVD constituted for a third of all deaths in 2019, according to research published in the Journal of the American College of Cardiology. Cardiovascular Diseases are a leading cause of deaths in all over the world and it in turn contributes to disability and steep increase in costs of healthcare. CVD does not just happen to older adults; it may also target the younger adults more and more often. This is partly because the conditions that lead to heart disease are happening at younger ages. The main reason behind that is a higher rate of obesity and high blood pressure among young demographic (ages 35-64) are putting them all at risk for serious heart diseases earlier in life.

CVD can occur when arteries that supply blood and oxygen to our heart muscle and other organs become clogged when fatty material called plaque or atheroma. This process is called atherosclerosis. If our arteries become too narrow, less blood can reach our heart muscle or our brain. When this happens in arteries of the heart, it can lead to symptoms such as angina. If a blood clot forms in the narrowed artery and blocks the blood supply to part of your heart, it can cause a heart attack. If this happens in the arteries supplying blood to the brain, this can cause a stroke. Over the last decades, although the age-standardized mortality rates of CVD declined by 27.3%, the number of deaths increased by 42.4% from 1990 to 2015[1]. CVD led to over 17 million deaths, 330 million years of life lost and 35.6 million years lived with disability in 2017 worldwide [2,3]. Hence in the present endeavour, we have attempted to build something related to it, that is deciding whether a person has a possibility of having a heart disease in future based on the various factors and numbers.

Technical abbreviations

CP Chest Pain

Test BPS resting blood pressure (in mm Hg on admission to the hospital)

CHOL serum cholesterol in mg/dl

FBS fasting blood sugar > 120 mg/dl)

RESTECG resting electrocardiographic results

THALACH maximum heart rate achieved

EXANG exercise induced angina

OLDPEAK ST depression induced by exercise relative to rest

TP True Positives

TN True Negatives

FP False positives

FN False Negatives

OBJECTIVE

- To study the data entered by the user and use the training details to work on the numbers
- To predict whether the person is susceptible to heart diseases or not based on the past trends and make them aware of their health
- To improve the accuracy and produce better results than the existing material

FEASIBILITY STUDY

Financial feasibility: Being an application, this project will have an associated cost. Since the system requires the processing, it will require a better processor and specifications for the operation of this application.

The system will follow the freeware software standards. No cost will be charged from the potential user. Bug fixing and maintaining tasks may have an associated cost.

Beside the associated cost, there will be many benefits for the user. Especially the effort that is associated with going through specific diagnosis will significantly reduce.

Technical Feasibility: The main technologies and tools that are associated are:

- Python
- Machine learning
- Python libraries
- Google colab
- Jupyter notebook

Each of the technologies are freely available and skills are easily manageable. Time limitations of the product development and the ease of implementing using these technologies are synchronized.

Resource feasibility:

- Programming device (laptop)
- Hosting space (freely available)
- Programming tools (freely available)
- Programming individuals

Need and significance: Cardiovascular diseases (CVD) have been on the rise since past few years and constituted for a third of all deaths in 2019. Cardiovascular Diseases are a leading cause of deaths in all over the world and it in turn contributes to disability and steep increase in costs of healthcare. The main reasons behind the occurrence of CVD are Coronary artery disease, diabetes, drug abuse, excessive use of alcohol or caffeine, heart defects you're born

with (congenital heart defects), high blood pressure, smoking and some over-the-counter medications, prescription medications, dietary supplements and herbal remedies.

And the preventions for such serious illness are anyone's guess. Get moving: Aim for at least 30 to 60 minutes of activity daily, eating a heart-healthy diet, maintaining a healthy weight, getting good quality sleep, managing stress, getting regular health screenings.

Therefore, in the present endeavour, we have attempted to build something related to it, that is predicting whether a person is susceptible to heart diseases or not. A healthy population is any nation's most valuable asset. So, for that, a tool was needed to make detection easier and accessible and hence saving lives.

METHODOLOGY OF WORK

- Identification of all the target values necessary for the prediction process
- Study of correlation of different factors
- Normalising data
- Testing and training the data
- Using various algorithms for final prediction

Evaluating parameters

Accuracy = (TP + TN)/(TP+TN+FP+FN)
Recall = TP/(TP+FN)
Precision = TP/(TP+FP)

RESEARCH GAP IDENTIFIED

The main inspiration behind this project was this research paper [4] and we believe that with right dataset and with the right algorithms, we could even increase the accuracy of the model. We will be using 7 different algorithms so that we can identify the best working and the most efficient one.

Following is the list of algorithms we plan to use:

- Logistic Regression
- Naïve Bayes
- Support vector Machine (Linear)
- K-Nearest Neighbour
- Decision Tree
- Random Forest
- Artificial Neural Network with 1 hidden layer

BIBLIOGRAPHY

[1] Mensah GA, Sampson UK, Roth GA, Forouzanfar MH, Naghavi M, Murray CJ, et al. Mortality from cardiovascular diseases in sub-Saharan Africa, 1990–2013: a systematic analysis of data from the global burden of disease study 2013. Cardiovasc J Afr. 2015;26(2

Suppl 1):S6–10. https://doi.org/10.5830/CVJA-2015-036.

[2] Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. Lancet. 2018;392(10159):1736–88. https://doi.org/10.1016/S0140-6736(18)32203-7.

[3] Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet. 2018;392(10159):1859–922. https://doi.org/10.1016/S0140-6736(18)32335-3.

[4] Chandra Reddy, N. S., Shue Nee, S., Zhi Min, L., & Xin Ying, C. (2019). Classification and Feature Selection Approaches by Machine Learning Techniques: Heart Disease Prediction. *International Journal of Innovative Computing*, *9*(1). https://doi.org/10.11113/ijic.v9n1.210