International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Machine Learning and Deep Learning Techniques to Prognosticate Heart Attack in Due of Abnormal Sleep Style

K R Sumana[1], Dr. H D Phaneendra[2], K S Kalpashree[3]

[1]*Assistant Professor, The National Institute of Engineering, Mysuru, Karnataka, India*
[2]*Professor, The National Institute of Engineering, Mysuru, Karnataka, India*
[3]*PG Student, The National Institute of Engineering, Mysuru, Karnataka, India*

Abstract - *Sleeping disorders, such as sleep deprivation and fragmented sleep have been shown to have a harmful impact on heart health. Sleep is critical for the body's recuperation. A heart attack also known as myocardial infarction, it happens when the flow of the blood to the heart gets blocked. Eating junk foods and being inactive can cause fatty material building up in arteries, if the arteries get damaged it can lead to a heart attack. Sleep deprivation, smoking, heavy drinking, sudden shock and a high stress level are all risky lifestyle choices that lead to heart disease. Shortened sleep and sleep disruption have also been related to an increased risk of heart attacks. Because both heart rate and blood pressure can rise suddenly when you wake up, it may cause cardiac stress and heart attack. Various medical data mining and machine learning techniques are being implemented to extract valuable information regarding heart disease prediction. In our work, we proposed a heart attack prediction system using deep learning and machine learning techniques to predict the likely possibilities of heart-related diseases of the patient. We found specifically Artificial Neural Network (76.30%), KNN (74.34%), Decision Tree (97.07 %), Random Forest (98.05%), Support Vector Machine (94.80%), and Naïve Bayes (84.09 %) classifiers for prediction and examined their performance and accuracy of the proposed system.*

Key Words: Artificial Neural Network(ANN), Support Vector Machine(SVM), Naïve Bayes(NB), K-Nearest Neighbor(KNN), Decision Tree(DT), Random Forest(RF).

1. INTRODUCTION

Heart disease is one of the most common ailments in today's globe. According to a report, more than 17.7 million people die each year due to heart disease over the world. A quarter of a million died from coronary heart disease, and 6.7 million died from stroke. Asymptomatic heart attacks are something that most doctors are unable to predict. Heart attacks are one of the worst diseases that can strike at any time and without warning. Chronic sleep deprivation has been related to a variety of cardiac issues, including hypertension, high cholesterol, heart attack, obesity, diabetes, and stroke. Sleeping disorders, such as sleep deprivation and fragmented sleep, have been shown to have a harmful impact on heart health. Multiple studies have found that sleep deprivation causes elevated daytime blood pressure, but it does not affect everyone similarly. In middle-aged people, the association between lack of sleep and high blood pressure is strongest. People who are working long hours in high-stress occupations or who are having other hypertension risk factors are more likely to have high blood pressure following a period of chronic inadequate sleep. Heart attacks are increased when people don't get enough sleep. People who slept less than six hours per night had a 20% increased probability of having a heart attack, according to one study. This project will address these issues and suggest the development of novel features in order to create a more comprehensive system.

2. LITERATURE REVIEW

Heart disease can be caused by a variety of factors, including changes in lifestyle, genetics, and smoking. The risk of heart disease is increased by a variety of genetic abnormalities. Cardiac arrest, arrhythmia, stroke, congestive heart failure, coronary artery disease, and congenital heart disease are some of the most prevalent heart diseases. The goal of this study is to examine several research studies on heart disease prediction and classification using various machine learning and deep learning approaches, in order to determine which strategies are the most successful and accurate[1]. A dataset were taken from the UCI machine learning repository. Rapid Miner was used to find the best fitting method for the provided dataset, and four algorithms were compared by constructing their processes in Rapid Miner, including Nave Bayes, Decision Trees, K-Nearest Neighbor, and Random Forest [2]. For the prediction of heart diseases, data mining techniques such as SVM, Random forest, KNN, and ANN classification algorithms are used. Data mining approaches may be useful as a solution [3]. To predict the likelihood of a patient developing heart disease, attributes such as age, sex, blood pressure, and blood sugar are extracted. These characteristics are input into classification algorithms such as SVM, Random forest, KNN, and ANN, with ANN providing the best results and greatest accuracy [4]. Random Forest was used to training the model, and the accuracy was compared to Logistic Regression and Naïve Bayes Classifier. It was compared with Logistic Regression, and Naïve Bayes Classifier. Random Forest performed better than Logistic Regression and Naïve Bayes Classifier [5].

3. PROPOSED SYSTEM

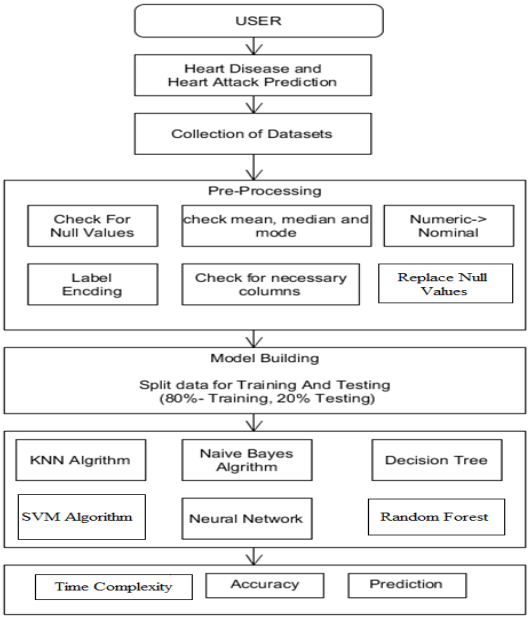

Fig -1: Work Flow of the Proposed System

The work flow of our work is as shown in above Fig-1. The dataset is collected to build a model for prediction. Firstly, the dataset is preprocessed using the preprocessing techniques. Found the mean, median and mode of the overall data of the each attributes then, check the null values, missing values and irrelevant data in the dataset. Then dataset is saved in .csv file format. To build the proposed model, we split the data into 80 % of training and 20% of testing dataset. The machine learning and deep learning techniques such as Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest and Artificial Neural Network are applied for the data to predict the heart attack. Through user interface we could displays the result and predict the heart attack with accuracy and time complexity of the algorithm.

4. IMPLEMENTATION

4.1 Data Preprocessing

As we know dataset is a collection of data objects which are having various numbers of attributes that deal with the basic characteristics of an object. We obtained 1025 records with 14 medical attributes (factors) from the UCI Machine Learning Data Repository. In the dataset, the data can have many irrelevant and missing parts. To handle this part, data preprocessing techniques is used i.e., data cleaning, data integration, data reduction, data transformation. After preprocessing, use of 1025 records with 14 medical attributes. Out of these 14 attributes, 8 have discrete values whereas 6 have continuous values. The heart attack dataset descriptions considered as shown Table-1. The dataset is saved in .csv format. CSV files, are a file type that allows us to save tabular data, such as spreadsheets.

Name	Type	Description
Age	Continuous	Age in years
Sex	Discrete	0: Female, 1: Male
CP(Chest Pain)	Discrete	Type of Chest Pain 1: typical angina, 2: atypical angina, 3: non-angina, 4: asymptomatic
Trestbps(Blood Pressure at Rest)	Continuous	Resting blood pressure(in mm/Hg)
Chol(Cholestrol)	Continuous	Serum Cholestrol in mg/dl
FBS(Fasting Blood Sugar)	Discrete	Blood sugar levels on fasting >120 mg/dl :1 else 0
RestECG(Rest Electro Cardio Graph)	Discrete	ECG at rest. 0: Normal, 1: Abnormality in ST-T wave, 2: Left ventricular hypertrophy
Thalch(Peak Heart Rate Achieved)	Continuous	Maximum heart rate achieved
Exang(Exercise Induced Angina)	Discrete	Angina induced by exercise 0: No, 1: Yes
Oldpeak	Continuous	Depression induced by exercise relative to rest
Slope	Discrete	The slope of the peak exercise segment 1: Upsloping, 2: Flat, 3: Downsloping
CA(Number of Blood Vessels)	Continuous	Number of blood vessels by fluoroscopy ranged between 0 to 3.
Thal	Discrete	3: Normal, 6: Fixed Defect, 7: Reversible Defect
Target	Discrete	1: patient is suffering from heart risk and 0: patient is normal

Table -1: Description of the Dataset

4.2 Classification

For classification, proposed model trained using classification algorithms such as SVM, KNN, RF, DC, NB and ANN. Then the model is tested for accuracy and time complexity. The obtained result is as showed in the Table-2. Their results are compared.

Algorithm Used	Accuracy Score	Time Taken(sec)
Naïve Bayes	84.09	0.00199
Support Vector Machine	94.80	3.98483
K- Nearest Neighbor	74.35	0.00200
Decision Tree	97.07	0.00500
Random Forest	98.05	0.01901
Artificial Neural Network	76.30	1.93014

Table -2: Accuracy Score and Time Complexity

4.3 Web Interface

The patient's risk factor is obtained in web interface. To access the classifier and check the risk factor, an interactive web interface was created. A Python-based CGI script serves as the backend. It accepts medical form data as input to a trained classification model, which forecasts the individual's risk factor. The below Fig-2 shows the user interface.

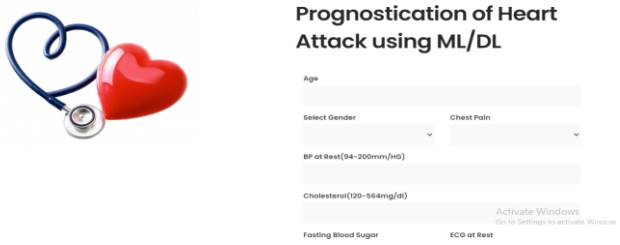

Fig 2: Web Interface

5. RESULT AND DISCUSSION

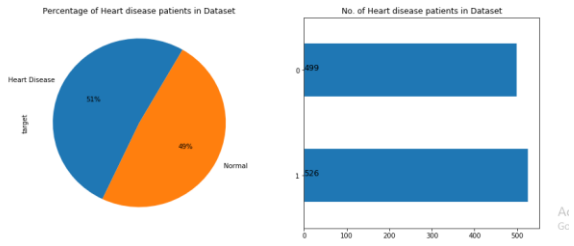

Fig-3: Heart Patient and Normal Patients Calculation

The Fig-3 represents the percentage and total number of heart disease patient and normal patient in dataset. There are 499(49%) of normal patient records and 526(51%) heart patient record.

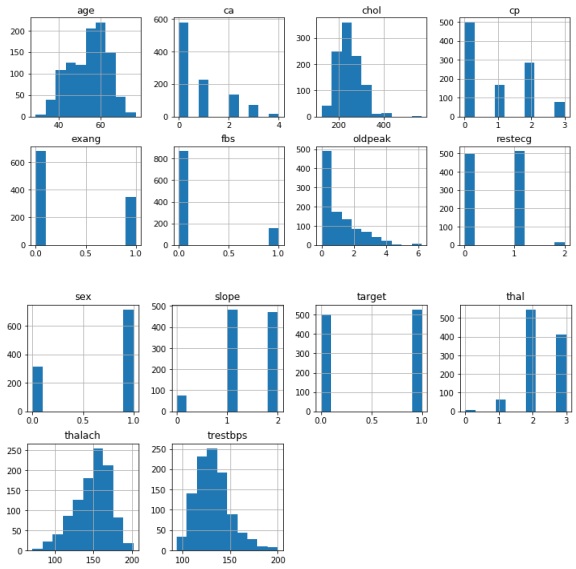

Fig-4: Histogram of Each Attribute

The Fig-4 represents the histogram graph of each attribute. A histogram is an approximate representation of numerical data distribution.

From the above results, we can conclude that Random Forest algorithm gives the highest accuracy with less time. SVM take more time to complete the process, compared to all other algorithms. The Naïve Bayes algorithm taken least time.

6. CONCLUSIONS

Our proposed work uses machine learning and deep learning to prognosticate the heart attack in due of abnormal sleep by using the Support Vector Machine, Random Forest, K-Nearest Neighbor, Decision Tree, Naïve Bayes and Artificial Neural Network techniques. The experimental findings are produced by taking into account 14 attributes. According to the findings, Random Forest is the most accurate method, also the second most accurate method is Decision Tree. Naïve Bayes, Support Vector Machine, K-Nearest Neighbor and Artificial Neural Network observed less accuracy.

A random forest classifier has an accuracy of 98.05% and time taken is 0.01901sec in this study. The accuracy of the decision tree was 97.07% and time taken is 0.00500sec. Support Vector Machine (SVM) achieved 94.8% and time taken is 3.98483sec, Naïve Bayes achieved 84.0% and time taken is 0.00199sec and K-Nearest Neighbor (KNN) achieved 74.35% and time taken is 0.00200sec, which is the lowest of all the classifiers in the analysis and ANN achieved 76.3% and time taken is 1.93014sec.

REFERENCES

[1] SP Rajamhoana C. Akalya Devi K.Umamaheswari R.Kiruba"Analysis of neural networks based heart disease prediction system" 978-1-538650250/18/ IEEE.

[2] Sushmita Manikandan "Heart Attack Prediction System" International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).

[3] N. Komal Kumar , G.Sarika Sindhu , D.Krishna Prashanthi , A.Shaeen Sulthana "Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers" 2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS).

[4] Mamatha Alex P and Shaicy P Shaji "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique" International Conference on Communication and Signal Processing, April 4-6, 2019, India.

[5] Obasi, M. Omair Shafiq, "Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases" 2019 IEEE International Conference on Big Data (Big Data).

[6] M. Saqlain, W. Hussain, N. A. Saqib and M. A. Khan, "Identification of heart failure by using unstructured data of cardiac patients," 2016 45th International Conference on Parallel Processing Workshops (ICPPW), Philadelphia, PA, 2016, pp. 426-431. doi: 10.1109/ICPPW.2016.66 .

[7] Dataset obtained from https://archive.ics.uci.edu/ml/machine-learning-databases/heartdisease/processed.cleveland.data