TO PREDICT HEART DISEASE USING MACHINE LEARNING ALGORITHM

Dr.M.Rajeswari[1]
Assistant Professor,
Department of B.Com (Business Analytics) ,
PSGR Krishnammal College for Women, Coimbatore, India.
rajeshwarim@psgrkc.ac.in

Krithika. p[2]
UG Scholar,
Department of B.Com (Business Analytics)
PSGR Krishnammal College for Women, Coimbatore, India.
ksharma101001@gmail.com

ABSTRACT

Heart is the most essential or crucial portion of our body. Heart is used to maintain and conjugate blood in our body. There are a lot of cases in the world related to heart diseases. People are leading to death due to heart disease. Various symptoms like chest pain, fasting of heartbeat and so on are mentioned. The health care industries found a large amount of data. This paper gives the idea of predicting heart disease using machine learning algorithms.

Keywords- Heart disease, Machine Learning, diabetes

I. INTRODUCTION

The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise. A major challenge faced by health care organizations, such as hospitals and medical centers, is the provision of quality services at affordable costs.1 The quality service implies diagnosing patients properly and administering effective treatments. The available heart disease database consists of both numerical and categorical data. Heart disease is one of the biggest causes of morbidity and mortality among the population of the worldThis makes heart disease a major concern to be dealt with. But it is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, and many other factors. Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease.

II. OBJECTIVE

To predict the reason for heart disease among different age groups using logistic regression algorithm using dataset collected from patients.

III. RELATED WORKS

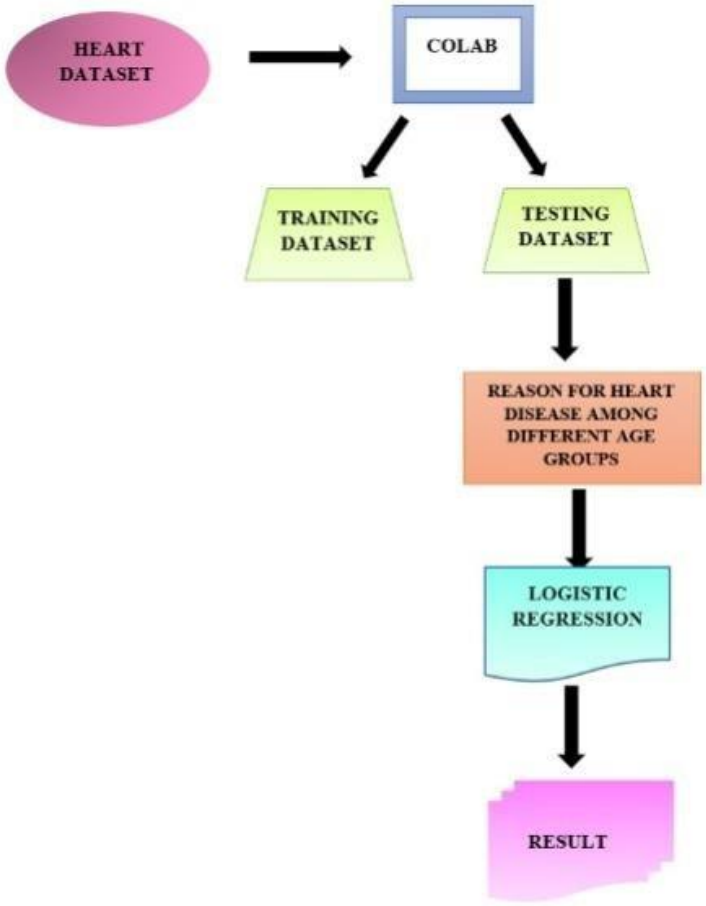
Several researchers and authors have studied, experimented with, and analyzed numerous techniques for heart disease predictions which includes the techniques for classification and feature selection.The authors proposed the hybrid HRFLM approach by combining the characteristics of the Linear Method (LM) and Random Forest (RF). They obtained a prediction accuracy of 88.4% [1].In 2018, the researchers used the Prediction models by using the different combinations of features, and seven classification techniques: k-NN, DT, NB, LR, SVM), NN, and VOTE (a hybrid technique with NaÃ¯ve Bayes and Logistic Regression). And their experiment results showed that the best-performing data mining technique, the VOTE technique with NB and LR achieved an accuracy of 87.4% in heart disease prediction [3].

The 10-fold cross-validation technique was used to validate the performance of the models[3]. The authors in [11] compared the use of several supervised machine learning (ML) algorithms for predicting clinical events in terms of their internal validity and accuracy and the results, which were obtained using two statistical software platforms that is R-Studio and Rapid Miner were then compared and showed that the decision tree algorithm gave better results.The authors in [15] performed the comparative study of heart disease diagnosis system using top ten data mining classification algorithms [27].

The data mining algorithms discussed were C4.5, SVM, Ada Boost, KNN, Naive Bayes, and CART, Random Forest, Bagging Algorithm, Logistic Regression, and Multilayer Perceptron (MLP). From their experimental study in terms of accuracy, the top three algorithms were Random Forest with 78.0%, kNN with 71.6%, and MLP with 63.8% and the top three based on speed were AdaBoost, kNN, and Naive Bayes. The authors in [16] carried the implementation of prediction algorithm and reached to the conclusion that the accuracy of the algorithms in machine learning depends upon the dataset that used for training and testing purpose[16].

IV. METHODOLOGY

A. WORKFLOW



B. PROPOSED SYSTEM

Step 1-To collect Heart dataset among different age groups

Step 2- Data was implemented under training and testing modules

Step 3-Using Visualization and prediction the reason for heart disease among different age groups was found out

Step 4- Logistic Algorithm was used to find out the accuracy and result

C. ALGORITHM - LOGISTIC REGRESSION

Logistic regression is used to predict the class (or category) of individuals based on one or multiple predictor variables (x). It is used to model a binary outcome, that is a variable, which can have only two possible values: 0 or 1, yes or no, diseased or non-diseased.

V. IMPLEMENTATION

```
#CURRENT STATUS
#accuracy for HeartDisease
metrics.accuracy_score(test_Z,predicted_value_1)*2
```

```
2.0
```

```
[74] #prediction  HeartDisease using slope
pd.DataFrame({'predicted_value':predicted_value_1,'KnowO/P':test_Z})
```

	predicted_value	KnowO/P
559	0	0
897	0	0
74	0	0
587	0	0
470	0	0
...	...	...

Fig1.1

	predicted_value	KnowO/P
559	3	3
897	3	3
74	3	3
587	7	7
470	7	7
...	...	...
393	3	3
716	7	7
582	6	6
110	3	3
936	7	7

300 rows × 2 columns

Fig1.2

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 16 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            1000 non-null   object
 1   age               1000 non-null   float64
 2   hypertension      1000 non-null   int64
 3   BMI               1000 non-null   float64
 4   Cholestrol        1000 non-null   int64
 5   diaBP             1000 non-null   float64
 6   Max heart rate    1000 non-null   int64
 7   glucose           1000 non-null   int64
 8   prevalent Stroke  1000 non-null   int64
 9   diabetes          1000 non-null   int64
 10  cigsPerDay        1000 non-null   int64
 11  PhysicalHealth    1000 non-null   int64
 12  MentalHealth      1000 non-null   int64
 13  F blood sugar     1000 non-null   int64
 14  thallium          1000 non-null   int64
 15  HeartDisease      1000 non-null   int64
dtypes: float64(3), int64(12), object(1)
memory usage: 125.1+ KB
```

Fig1.3

Fig 1.1, 1.2, 1.3 are Various datasets like heart disease, BPM, Maxheartrare, glucose etc are collected in colab using logistic regression algoritgm it predicted an accuracy of 2.0 for people who are at a risk of heart disease among different age groups

VI. CONCLUSION

As my objective was to predict the reason for heart diseases among various age groups after applying the algorithm in the dataset it shows an average accuracy and therefore conclude that due to various reasons like hypertension ,diabetes and max heart rate persons above the age of 65 have a higher risk of developing heart disease.

REFRENCES

[1] R. S. Singh, B. S. Saini, and R. K. Sunkaria, "Detection of coronary artery disease by reduced features and extreme learning machine," *Medicine and Pharmacy Reports*, vol. 91, no. 2, pp. 166–175, 2018.

[2] F. Yaghouby, F. Yaghouby, A. Ayatollahi, and R. Soleimani, "Classification of cardiac abnormalities using reduced features of heart rate variability signal," *World Applied Sciences Journal*, vol. 6, no. 11, pp. 1547–1554, 2009.

[3] B. M. Asl, S. K. Setarehdan, and M. Mohebbi, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal," *Artificial Intelligence in Medicine*, vol. 44, no. 1, pp. 51–64, 2008.

[4] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, *Feature Extraction: Foundations and Applications*, Springer, Cham, Switzerland, 2008.

[5] R. Rajagopal and V. Ranganathan, "Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification," *Biomedical Signal Processing and Control*, vol. 34, pp. 1–8, 2017.

[6] D. Zhang, L. Zou, X. Zhou, and F. He, "Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer," *IEEE Access*, vol. 6, pp. 28936–28944, 2018.

[7] S. Negi, Y. Kumar, and V. M. Mishra, "Feature extraction and classification for EMG signals using linear discriminant analysis," in *Proceedings of the 2016 2nd International Conference on Advances in Computing, Communication, & Automation (ICACCA) (Fall)*, IEEE, Bareilly, India, September 2016.

[8] D. Avendaño-Valencia, F. Martinez-Tabares, D. Acosta-Medina, I. Godino-Llorente, and G. Castellanos-Dominguez, "TFR-based feature extraction using PCA approaches for discrimination of heart murmurs," in *Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5665–5668, IEEE, Minneapolis, MN, USA, September 2009.

[9] P. Kamencay, R. Hudec, M. Benco, and M. Zachariasova, "Feature extraction for object recognition using PCA-KNN with application to medical image analysis," in *Proceedings of the 2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 830–834, IEEE, Rome, Italy, July 2013.

[10] N. R. Ratnasari, A. Susanto, I. Soesanti, and Maesadji, "Thoracic X-ray features extraction using thresholding-based ROI template and PCA-based features selection for lung TB classification purposes," in *Proceedings of the 2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICIBME)*, pp. 65–69, IEEE, Bandung, Indonesia, November 2013.