IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-15, May 2019 E-ISSN: 2347-2693

Prediction of Heart Disease with Claims Data using Machine Learning Method

Lavanya L[1*], Megha V[2], Nagashree H[3] ,Pavithra S[4], Anusha K L[5]

[1,2,3,4,5]Department of Computer Science, East West Institute of Technology, Bengaluru, India

DOI: https://doi.org/10.26438/ijcse/v7si15.6568 | Available online at: www.ijcseonline.org

Abstract—Machine learning can be referred as discovery of relationships in larger datasets and in some cases it is used for predicting relationships based on the results discovered. Nowadays machine learning is achieving widespread in various fields such as healthcare industry, scientific and engineering. In healthcare industry, machine learning is mainly used for disease prediction. The main objective of our work is to predict heart disease using Naïve Bayes classifier. Naïve Bayes are the probabilistic classifiers used to classify the data using attributes. It retrieves the trained data and compares the attribute values with test data sets and predicts the result.

Keywords-Machine learning, prediction, healthcare industry, Naïve Bayes Classifier, Heartdisease.

I. INTRODUCTION

In this modern world people due to the changing lifestyle, unhealthy food habits, environmental impacts, they are tensed, they have blood pressure, diabetes at very young age and they don't even give enough rest for themselves this negligence leads to several disease.

One among them is heart disease. Heart disease is a general name for a variety of diseases, conditions and the blood vessels. Symptoms of heart disease vary depending on the specific type of heart disease. [1][2]

A problem with heart's structure and function due to abnormal heart development before birth refers to congenital heart disease. Congestive heart failure occurs when the heart fails to pump adequate blood to the other organs in the body.

Coronary heart disease (medical term is Ischemic heart disease) is the most frequent type of heart problem. Coronary heart disease is a term that refers to the damage to the heart. This happens because its blood supply is decreased; it leads to the fatty deposits build up on the linings of the blood vessels that provide the heart muscles with blood, results in narrowing. Most hospital today employs some sort of the hospital information systems to manage their healthcare or patient data. These systems typically generate large amounts of data which will be in the form of numbers, text, charts and images. Unfortunately, these data are used rarely to support in the clinical decision making. There is a huge amount of hidden information in these data that is largely unused. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?"

Heart disease prediction using machine learning is one of the most interesting and challenging task. The shortage of specialists and high wrongly predicted cases has necessitated the need to develop a fast and efficient detection system. [1] [2][3]

The main objective of this work is to identify features and key patterns from the medical data using the classifier model. The attributes that is more relevant to heart disease diagnosis can be observed. This will help the medical practitioners to understand the disease easily. [4]

A classifier will process mapping of feature space to a discrete set of labels. There are many applications of the classifiers which arewide ranging in the fields such as medicine, finance, face recognition, tracking of the targets, voice recognition, computer vision and in many other areas.[5]

II. RELATED WORK

A number of recent studies have been successful in predictingdisease via various methods, including Naïve Bayes Classifiers ,support vector machines [6], [10], [11], [12], logistic regression [9],[2], and time series modeling techniques [3].

We note that the following have been particularly salient inrecent literature:

• Embedding of Medical Concepts: Many papers have applied word embedding techniques from naturallanguage processing to obtain embedded representations of

medications, diagnoses, and procedures using adaptations of word2vec [14] and GloVe [15].

• Disease Prediction Models: A number of papers havebeen developed means of predicting future medical conditions using a number of classification models.

The vast majority of recent literature on disease predictionhas focused on Machine learning, to the exclusion of other viable models as baselines for prediction tasks.

To predict the Heart disease using Machine learning, which is one ofthe most interesting and challengingtasks, to identify the key patterns or features from the medical data using the classifier model, the attribute that are more relevant could be observed and for designing the applicationthat predicts the heart disease of the user by collecting his healthrecords.

III. METHODOLOGY

To enhance the prediction of classifiers, genetic search is incorporated; the genetic search resulted in 12 attributes which contributes more towards the prediction of the cardiac disease. The classifiers such as Naïve Bayes were used for prediction of heart disease. The classifiers were fed with reduced data set with 12 attributes.

Methods used in this paper are Naïve Bayes classifier, input to weight conversion, excel data uploading and MVC architecture.

a)Naïve Bayes Classifier:

A naïve Bayes classifier is simple probabilistic classifier where classification is done based on by applying Bayes theorem. Bayesian n classifier considers the feature of a class whether it is present are not unrelated to the presence of other feature which means it focuses on the independent existence of the features. For example a fruit maybe considered as Chikku if it is brown, round and 3cm in diameter, Naïve Bayes classifier will considers all of these features to contribute independently and it won't accept co-relation between the features.

The working of Naïve Bayes classifier is described below: i) converts the data sets into the frequency data ii) Creates the likelihood table for the frequency data iii) Applying Bayesian equation to calculate the posterior probability for the each class. The class with the highest posterior probability will be the outcome for the prediction.

Bayesian equation as follows

$$P(W|Q) = \frac{P(Q|W)P(W)}{\mathrm{P(Q)}}$$

By applying the equation posterior probability can be found.

The advantages of the Naïve Bayes classifier are simple, easy and fast to implement. The main advantage of this classifier is considering the independent existence of the features and this classifier requires less training data to classify.

b) Input to weightage conversion

Input attributes are age, sex, chest pain, Rest BP, Cholesterol, fasting blood sugar, Rest ECG, Thalach(max rate achieved), Exang (Exercise induced angina), Old peak(ST depression), Slope of ST segment, Thal. For all these attributes weightage conversion method is applied.

1. Age
<40-0.3
41-50-0.5
51-60-0.7
61-70-0.85

2. Sex
Female – 0.45
Male – 0.75

3. Chest Pain
Typical Angina - 0.65
Atypical Angina – 0.6
Non – anginal – 0.1
Asymptomatic – 0.8

4. Rest BP
120/80 – 0.1
120-139/80-89 – 0.5
140-159/90-99 – 0.75

5. Cholesterol (mg/dl)
<200 – 0.1
200-240 – 0.75
>240 – 0.9

6. Fasting Blood sugar
70 – 100 – 0.25
101-126 – 0.6
>126 – 0.9

7. Rest ECG
Normal – 0.1
Left ventricular Hypertrophy - 0.85

8. Thalach (Max heart rate achieved)
<0.8(220-age) – 0.3
>0.8(220-age) – 0.6

9. Exang (Exercise induced Angina)
No – 0.1
Yes – 0.8

10. Old peak (ST depression)
<1.5 - 0.1
1.6 – 30 – 0.45
3.1-4.5 – 0.75
>4.5 - 0.9

11. Slope of ST Segment
Upslopping – 0.8
Flat – 0.4
Downsloping – 0.35

12. Thal (Thalium)
Nrmal – 0.1
Reversible defect – 0.5
Fixed defect – 0.85

All these are the attributes which are undergone weightage conversion method. User will enter all these attributes and for the entered values Weightage conversion method is been applied.

c) Excel data uploading
There are many EHR (Electronic Health Records) which are available. EHR's are used to store data and to maintain status of the patients. There is no need to keep track of previous patient records because the complete information is available in those EHR. EHR consists patient details like age, weight, medical history, immunization status, laboratory test results, vital signs, demography etc. The required features which are in EHR are used for training, which are known as training data sets. Trained data sets are uploaded in the form of CSV (comma separated values) files with the extension .CSV files.

d) MVC Architecture

MVC is an architectural pattern. It consists of three components model, view and controller. Model represents data and the business logic, view is a user interface and controller is an interface for model and view.

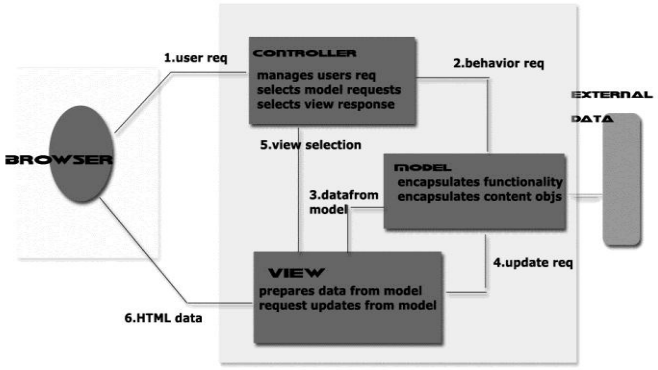

Figure 1: MVC architecture

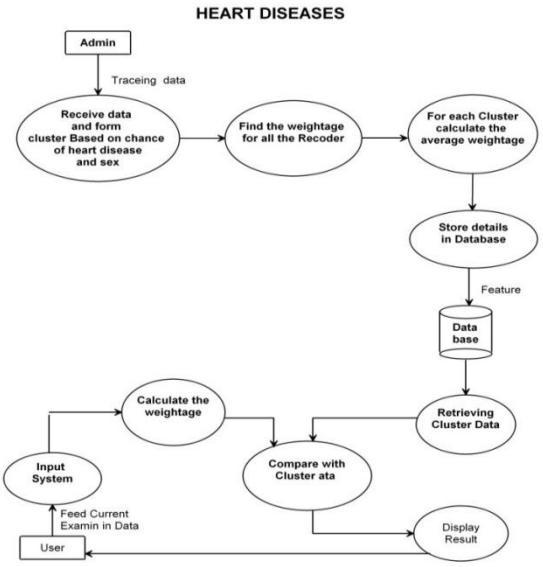

Figure 2: System architecture

IV. RESULTS

Our work results in predicting the heart disease in Human.That means the user or the patients should give his correct health inputs in the patient's history and he will be able to get the results in the form of report saying the prediction result in terms of percentage accuracy by predicting his/her heart disease.

V. CONCLUSION AND FUTURE WORKS

Our work gives the accurate prediction results using the above given attributes. More the number of attributes the

more accuracy is gained .Because the Naïve Bayes classifier is mainly a probabilistic.Classifier which is widely used for prediction of disease. We have chosen to work on it for getting better results out of it which can help the patients in the early stage itself. As it is saidthat "Prevention is better than cure".

As our work will only predict the disease, future work can be made on the diagnosis to improve the efficiency of our project.

ACKNOWLEDGMENT

Firstly, we express our sincere thanks to our guide Mrs.Anusha K L, Assistant Professor, Department of CSE, EWIT and Dr.ArunBiradar, Head of Department, Computer science and engineering for their moral support. We express our sincere gratitude to our principal Dr. K Chennakeshavalu for his constant support and encouragement; we also thank all the faculties of East West Institute of Technology for their co-operation and support.

REFERENCES

[1] E. Choi, M.T.Bahadori, L.Song, W.F. Stewart, and J.Sun, "Gram: Graph- based attention model for health acre representation learning," in proceedings of the *23*rd ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '17. New York,NY, USA:ACM, pp 787-795.

[2] D. Kartchner, T. Chirstensen, J. Humphreys, and S.Wade, "Code2vec: Embedding and clustering medical diagosis data," in *2017* IEEE International Conference on Healthcare Informatics (ICHI), Aug 2017, pp. 386-390.

[3] R. Miotto, L. Li, B. A. Kidd, and J.Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," in Scientific reports, 2016

[4] Z. Che, Y. Cheng, S. Zhai, Z.Sun and Y.Liu, "Boosting deep learning risk prediction with generative adversarial netwoks for electronic health records, " in *2017* IEEE International Conference on Data Mining (ICDM) , Nov 2017, pp. 787-792.

[5]. R.Snyderman, "Personalized health care: from theory to practice," Biotechnology Journal, vol. 7, no. 8, pp. 973-979, Aug. 2012

[6] N.Razavian and D. Sontag, "Temporal convolution neural networks for diagnosis from lab tests," CoRR, vol. abs/1511.07938, 2015.

[7] A. Perotte, R. Ranganath, J.S. Hirsch, D. Blei, and N.Elthadad, "Risk prediction forchronic kidney disease progression using heterogenous electronic helath record data and tie series analysis," Journal of the American Medical Informatics Association: JAMLA, vol. 22, no. 4, pp. 872-880, Jul. 2015.

[8] N. Tangri, L.A. Stevens, J, Griffith, H. Tghiouart, O.Djurdjev, D. Naimark, A. Levin, and A.S levey, "A Predictive model for progression of chronic kidney disease to kidney failure," Jama, vol.305, no.15, pp. 1553-1559, 2011.

[9] E. Choi, M.T. Bahadori, A Schuetz , W. F. Stewart, and J. Sun "Doctor ai: Predicting clinical events via recurrent neural networks ," in *Proceedings of the !st Machine Learning for Healthcare Conferences ser.* Proceedings of Machine Learning Research, F. Doshi-Velez, J. Fackler, D.Kale, B. Wallace, and J.Wiens, Eds., vol.56. Children's Hospital LA, Los Angeles, CA, USA: PMLR, 18-1, Aug 2016, pp.301-318

[10] M.Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," BMC medical informatics and decision making, vol.11, no. 1, p. 1,2011.

[11] X.Wang, D. Sontag, and F.Wang, "Unspervised learning of disease progression models," *in Proceedings of the 20th ACM SIGKDD International Conference on knowledge discovery and data mining.* ACM, 2014, pp. 85-94.

[12] Z.C. Lipton, D.C. Kale, C.Elkan, and R. Wetzell, "Learning to diagnose with lstm recurrent neural networks," *arXiv preprint arXiv: 1511. 03677, 2015*

[13]N.Barakat, A.P.Bradley, and M.N.H. Barakat, "Intelligible support vector Machine for Diagnosis of Diabetes Mellitus," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no, pp.1114- 1120, Jul. 2010.

[14] "Prediction Modelling Using HER data: Challenges, Strategies Medical Carehhtp://journals.lww.com/lwwmedcalcare/Fulltext/2010/06001/Prediction Modeling Using EHR data Challenges,.17.aspx.

[15] W. Yu, T. Liu, R. Valdez, M. Gwinn, and M.J. Khoury, "Application of support vector machine modelling for prediction of common disease: the case of diabetes and pre-diabetes," *BMc Medical Informatics and Decision Making,* vol. 10, p. 16, 2010.

[16] A. V. Lebedev, E, Westman, G.J. P. Van Western, M.G. Kramberger, A. Lundervold, D, Aarsland, H. Soininen, I. Kloszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone, And A. Simmons, "Random Forest ensembles for detection and prediction of Alzheimer's disease with good between- cohort robustness," NeuroImage: Clinical, vol.6, pp. 115-125, 2014

Authors Profile

Ms. Lavanya L is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Machine Learning.

Ms. Megha V is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Machine Learning.

Ms. Nagashree His pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Machine Learning.

Ms. Pavithra S is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Machine Learning.

Mrs.Anusha K L got her M.Tech degree in Computer Networks & Engineering, Bengaluru, India. She is currently working as Assistant Professor in the Department of CSE, EWIT. Her areas of interest includes Image Processing, Machine Learning. Cloud Computing, SDN.

