IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Disease Prediction using Machine Learning

Suresh Singh Rajpurohit

ABSTRACT

The wide adaptation of computer-based technology in the health care industry resulted in the accumulation of electronic data. Due to the substantial amounts of data, medical doctors are facing challenges to analyze symptoms accurately and identify diseases at an early stage. However, supervised machine learning (ML) algorithms have showcased significant potential in surpassing standard systems for disease diagnosis and aiding medical experts in the early detection of high-risk diseases. In this literature, the aim is to recognize trends across various types of supervised ML models in disease detection through the examination of performance metrics. The most prominently discussed supervised ML algorithms were Na¨ıve Bayes (NB), Decision Trees (DT), K-Nearest Neighbor (KNN). As per findings, Support Vector Machine (SVM) is the most adequate at detecting kidney diseases and Parkinson's disease. The Logistic Regression (LR) performed highly at the prediction of heart diseases. Finally, Random Forest (RF), and Convolutional Neural Networks (CNN) predicted in precision breast diseases and common diseases, respectively.

Key Words: Health Care, Supervised Machine Learning, Diseases Prediction.

1. INTRODUCTION

Machine learning is programming computers to optimize a performance using example data or past data. Machine learning is the study of computer systems that learn from data and experience. Machine learning algorithm has two tracks: Training, Testing. Prediction of a disease by using patient's symptoms and history machine learning technology is striving from past decades. Machine Learning technology gives an immeasurable platform in the medical field so that healthcare issues can be resolved efficiently.

We are applying machine learning to maintained complete hospital data Machine learning technology which allows building models to get quickly analysed data and deliver results faster, with the use of machine learning technology doctors can make a good decision for patient diagnoses and treatment options, which leads to improvement of patient healthcare services. Healthcare is the most prime example of how machine learning is used in the medical field.

To improve the accuracy from massive data, the existing work will be done on unstructured and textual data. For the prediction of diseases, the existing will be done on linear, KNN, Decision Tree algorithm. The order of reference in the

running text should match with the list of references at the end of the paper.

2. OBJECTIVE

Health information needs are also changing the information seeking behaviour and can be observed around the globe. Challenges faced by many people are looking online for health information regarding diseases, diagnoses and different treatments. If a recommendation system can be made for doctors and medicine while using review mining will save a lot of time. In this type of system, the user face problem in understanding the heterogeneous medical vocabulary as the users are laymen. User is confused because a large amount of medical information on different mediums are available. The idea behind recommender system is to adapt to cope with the special requirements of the health domain related with users.

3. EXISTING SYSTEM

Since the arrival of advanced computing, the doctors still require the technology in various possible ways like surgical representation process and x-ray photography, but the technology perceptually stayed behind. The method still requires the doctor's knowledge and experience due to alternative factors starting from medical records to weather conditions, atmosphere, blood pressure and numerous alternative factors. The huge numbers of variables are granted as entire variables that are required to understand the complete working process itself, nevertheless, no model has analysed successfully. To tackle this drawback, medical decision support systems must be used. This system can assist the doctors to make the correct decision.

We are applying machine learning to maintained complete hospital data Machine learning technology which allows building models to get quickly analysed data and deliver results faster, with the use of machine learning technology doctors can make a big decision for patient diagnoses and treatment choices, which leads to enhancement of patient

healthcare services. Healthcare is the most prime example of how machine learning is used in the medical field.

4. PROPOSED SYSTEM

This system is used to predict disease according to symptoms. This system uses decision tree classifier for evaluating the model. This system is used by end-users. The system will predict disease based on symptoms. This system uses Machine Learning Technology. For predicting diseases, the decision tree classifier algorithm is used.

We have named this system as 'AI THERAPIST'. This system is for those people who are always fretting about their health, for this reason, we provide some features which acknowledge them and enhance their mood too. So, there is a feature for the awareness of health 'Disease Predictor', which recognize disease according to symptoms.

5. DATASET AND MODEL DESCRIPTION

This dataset is a knowledge database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patients at New York Presbyterian Hospital admitted during 2004. The first column shows the disease, the second the number of discharge summaries containing a positive and current mention of the disease, and the associated symptom. Associations for the 150 most frequent diseases based on these notes were computed and the symptoms are shown ranked based on the strength of association. The method used the Med LEE natural language processing system to obtain UMLS codes for diseases and symptoms from the notes; then statistical methods based on frequencies and co-occurrences were used to obtain the associations. A more detailed description of the automated method can be found in Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical reports. AMIA Annu Symp Proc. 2008. p. 783-7. PMCID: PMC2656103.

Disease	Count of Disease Occurrence	Symptom
UMLS:C0020538_hypertensive disease	3363	UMLS:C0008031_pain chest
		UMLS:C0012833_dizziness
		UMLS:C0004093_asthenia
		UMLS:C0085639_fall

[Fig 5.1. Dataset]

6. SYSTEM ARCHITECTURE

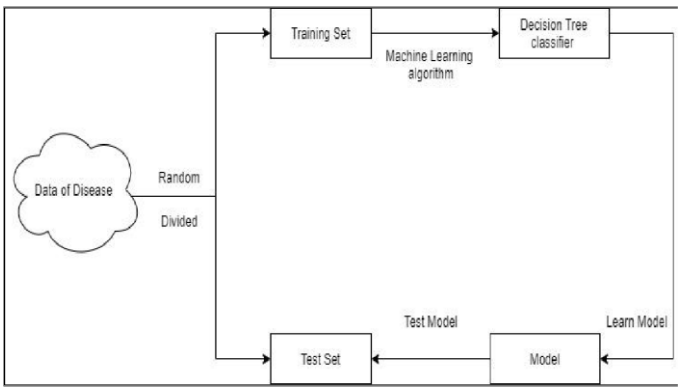
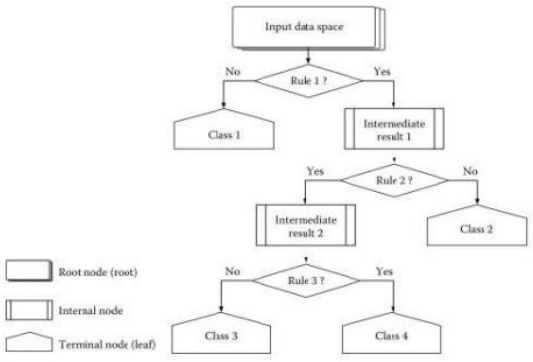


Fig -1: System Architecture

7. ALGORITHM

7.1 DECISION TREE

Decision tree is classified as a very effective and versatile classification technique. It is used in pattern recognition and classification for image. It is used for classification in very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts root, nodes and leaf. Roots consists of attribute which has most effect on the Roots consists of attribute which has most effect on the outcome, leaf tests for value of certain attribute and.



7.2 Random Forest Algorithm

Random Forest Algorithm is a supervised learning algorithm used for both classification and regression. This algorithm works on 4 basic steps – 1. It chooses random data samples from dataset. 2. It constructs decision trees for every sample dataset chosen. 3. At this step every predicted result will be compiled and voted on. 4. At last most voted prediction will be selected and be presented as result of classification.

7.3 Logistic Regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes 0 for failure and 1 for success.

7.4 K-Nearest Neighbour

K-Nearest Neighbour The K-Nearest Neighbour algorithm is a supervised classification algorithm method. It classifies objects dependant on nearest neighbour. It is a type of instance-based learning. The calculation of distance of an attribute from its neighbours is measured using Euclidean distance. It uses a group of named points and uses them on how to mark another point. The data are clustered based on similarity amongst them. K-NN algorithm is simple to carry out without creating a model or making other assumptions. This algorithm is versatile and is used for classification, regression, and search. Even though K-NN is the simplest algorithm, noisy and irrelevant features affect its accuracy.

7.5 Naïve Bayes

Naïve Bayes Naïve Bayes classifier is a supervised algorithm. It is a simple classification technique using Bayes theorem. It assumes independence among attributes. Bayes theorem is a mathematical concept that is used to obtain the probability. The predictors are neither related to each other nor have correlation to one another. All the attributes independently contribute to the probability to maximize it. Many complex real-world situations use Naive Bayes classifiers

$P(X/Y) = P(Y/X) \times P(X)/P(Y)$, $P(X/Y)$

is the posterior probability, $P(X)$ is the class prior probability, $P(Y)$ is the predictor prior probability, $P(Y/X)$ is the likelihood, probability of predictor

8. EVALUATING THE MODEL& RESULTS

The results obtained from our model are summarized in the following table:

Algorithm	Accuracy after pre processing
decision tree classification	**90.00**
K-Nearest Neighbour	**91.45**
Logistic Regression	**94.55**
Naïve Bayes classifiers	**92.47**
Random Forest	**95.32**

[Table 2: Accuracy comparison table of the five algorithms]

It can be observed that the pre-processing technique, discretization has improved the performance of all the five algorithms.

9. CONCLUSIONS

To conclude, our system is helpful to those people who are always worrying about their health and they need to know what happens with their body. Our main motto to develop this system is to know them for their health. Especially, people who are suffering from mental illness like depression, anxiety. They can come out of these problems and can live their daily lives easily.

Besides, our system provides better accuracy of disease prediction according to symptoms of the user, and also it will provide motivational thoughts and images. In the end, we can say that our system has no boundary of the user because everyone can use this system.

10.REFERENCES

[1] Pingale, Kedar, et al. "Disease Prediction using Machine Learning." (2019).Mr. Chala Beyene, Prof. Pooja Kamat, “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques”, International Journal of Pure and Applied Mathematics, 2018.

[2] Pingale, K., Surwase, S., Kulkarni, V., Sarage, S., & Karve, A. (2019). Disease Prediction using Machine Learning.

[3] Aiyesha Sadiya, Differential Diagnosis of Tuberculosis and Pneumonia using Machine Learning(2019)

[4] S. Patel and H. Patel, “Survey of data mining techniques used in healthcare domain,” Int. J. of Inform. Sci. and Tech., Vol. 6, pp. 53-60, March, 2016.

[5] Balasubramanian, Satyabhama, and Balaji Subramanian. "Symptom based disease prediction in medical system by using Kmeans algorithm." International Journal of Advances in Computer Science and Technology 3.

[6] Dhenakaran, K. Rajalakshmi Dr SS. "Analysis of Data mining Prediction Techniques in Healthcare Management System." International Journal of Advanced Research in Computer Science and Software Engineering 5.4 (2015).