



Hospital Recommendation System using Machine Learning

Mani V¹, GokulPrasath R², Jegatha S³, Kannathal K⁴, MadhuMitha K⁵

¹Department of CSE, M. Kumarasamy College of Engineering, India, vmanimkce663@gmail.com

²Department of CSE, M. Kumarasamy College of Engineering, India, gokulvcr007@gmail.com

³Department of CSE, M. Kumarasamy College of Engineering, India, srijegatha19@gmail.com

⁴Department of CSE, M. Kumarasamy College of Engineering, India, kjaya5952@gmail.com

⁵Department of CSE, M. Kumarasamy College of Engineering, India, madhukathir1906@gmail.com

ABSTRACT

At present, human beings face diverse diseases because of the environmental condition, lifestyle and dwelling habits. The prediction of diseases at a very early stage will become a critical task. Machine learning (ML) has been proven to be effective in making decisions and predictions from a large number of records produced through the healthcare industry. We propose a system for early prediction of diseases from symptoms. Additionally, this system suggests nearby hospitals based on their location and ratings of the users and it makes an appointment concerning to doctor availability. This system is an attempt to reduce the efforts and time put in by the doctors and patients by automating the disease prediction. For accurate disease prediction, the datasets are processed in python programming language using Random Forest machine learning algorithms.

Key words: Disease Prediction, Machine Learning, Random Forest, Rating.

1. INTRODUCTION

The flourishing society is built when its entire people are conscious of their health. In their busy schedule, People are careless and are not aware of their health status. Unless the symptoms get severe, People don't visit hospitals.

To consume time and to use in emergencies, we propose a system that predicts the diseases from symptoms specified by the user using machine learning and suggest nearby hospitals based on their location and ratings. It not only saves lives but also money.

Machine learning is a subfield of artificial intelligence that helps the computer to think automatically without human intervention. A machine learns from past Experiences (data feed by the user) and its performance improves with experience. Machine learning is classified into unsupervised,

supervised, semi-supervised, reinforcement learning. The computer systems make decisions by using these learning techniques. Here, we use a supervised learning technique to train and test our labeled data sets (which have both input and output parameters). Many practical machine learning applications use supervised learning.

The datasets are processed using Random forest classification algorithm. Classification comprises of learning step and prediction step. The learning step is developed using the provided training data set and the prediction step predicts the response for provided data. The hospitals are suggested based on the ratings provided by the user.

2. LITERATURE REVIEW

This project mainly focuses on the speech articulation problem by using symptoms of Parkinson's Disease affected persons and using various data mining approaches. statistical classifier, support vector machine classifier and Tree classifier are the data mining techniques involved in the project. The classifiers are calculated with respect to three performance matrices. There are lot of symptoms which leads to Parkinson's disease but it specifically focuses on the speech articulation difficulties. The consequences are based on the Logistic Regression (Statistical classifier), Sequential Minimization Optimization (Support Vector Machine) and Decision Stump (Tree classifier). It also uses k-fold cross validation technique and the performance matrices [3].

Reinforcement learning is used in this proposed paper. This algorithm gives highly reward for desired outcomes and low rewards for undesired outcomes. They assume predicted disease as Di and severity as Sri for cure time prediction. Each diagnosis has an individual effect on recovery cure time and severity. They assumed the ratings are denoted as [DSR1, DSR2..DSRk]. Moreover, in this paper, they showed the importance to find the corresponding recovery time in each case concerning the disease encountered. They assumed the recovery time be denoted as [DC1, DC2, DCk]. Formula to calculate the recovery time prediction of the disease is $NS_j = 1$

$-\frac{|SR_i - DSR_j|}{(|SR_i| + |DSR_j|)}$ by using the severity of the disease fed by the user[4].

This paper uses the Support Vector machine technique to calculate accuracy rate. Accuracy, specificity and sensitivity are the performance measures used for evaluating the system. The successfully and incorrectly classified instances displays the percentage of sample instances. SVM, Random Forest, Naïve Bayes Classifier and Classification Tree are the models used in this paper. A reliable multi-process method which combines decision tree algorithm and clustering was used in this system. This helps to build a prediction system for cardiac arrest. We can conclude that the Random Forest classifier algorithm tends to be more accurate and efficient than the others[7].

They used the analytical approach and classification algorithms such as Decision Tree, KNN, and Random Forest for prediction. There a number of algorithms used in machine learning like Naive Bayes, Convolutional Neural Network (CNN) etc., But they are less accurate than their experimental analysis.

In this approach, Machine learning algorithms were used for analysis such as KNN, RF and DT and 10-Fold-croo-validation is done in these algorithms. The parameters segment describes various performance measures. They are measured and compared. To acquire a robust prediction model, the random forest serves the best. The health care domain can be efficiently managed using such models. We can note that the AUC and average accuracy of RF classifier is better than that of DT and KNN [9].

The brain's growth is linked with a disease named Autism Spectrum Disorder (ASD). Communication and behavior are affected because of ASD. This method predicts the ASD in adults using the Random Forest classifier. UCI machine learning repository contains the dataset for ASD prediction. This dataset was obtained as a result of the test which involves effective screening technique. It calculates Sensitivity, Accuracy, and specificity. Using the calculated values of FP, FN, TN and TP, the ROC curve was plotted. The model's performance has been analyzed using these measures.

This system evaluates the performance of both testing training tests. It can be stated that the performance of the classification model was good from the performance measures. From this, we can conclude, diagnosis for ASD patients can be done with this methodology. The classification algorithm has some limitations because it provides various results concerning time. This problem can be overcome by setting some parameters as constant [1].

3. SYSTEM ARCHITECTURE

3.1 ARCHITECTURE OVERVIEW

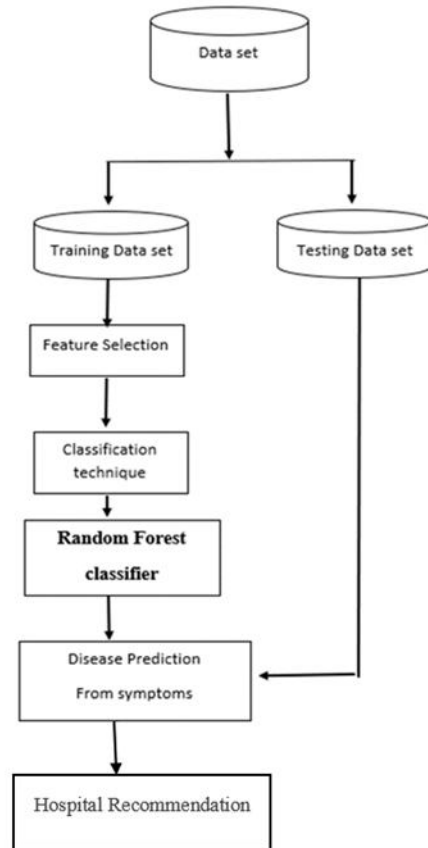


Figure 1: System Architecture of Hospital Recommendation system based on Diseases

Initially, the disease dataset is taken from machine learning websites and it will be displayed as a list of symptoms with diseases. This dataset is used as a training dataset. After that features are extracted and selected from the dataset and then classify that data using Random Forest classification techniques. With the help of machine learning techniques, we can predict accurate diseases from symptoms.

3.2 ALGORITHM USED

3.2.1 Random Forest Classifier

Ensemble Technique is a method in machine learning that integrates multiple learning algorithms to yield best predictive performance. The Random forest classifier is one of the ensemble algorithms.

Random Forest or Random Decision Forest is a supervised machine learning algorithm. The Random forest can be implemented in both Regression and Classification. The Random forest is one of the compatible and convenient algorithms. Random forests form Decision trees by selecting random data samples, obtains a prediction from every decision tree and best result is selected through voting.

Random forest algorithm functions as follows:

1. Selection of random samples from the provided dataset.
2. A decision tree is constructed for all samples and it obtains a prediction result from every decision tree.
3. Voting will be performed for each predicted outcome.
4. Most voted prediction result is selected as the final prediction.

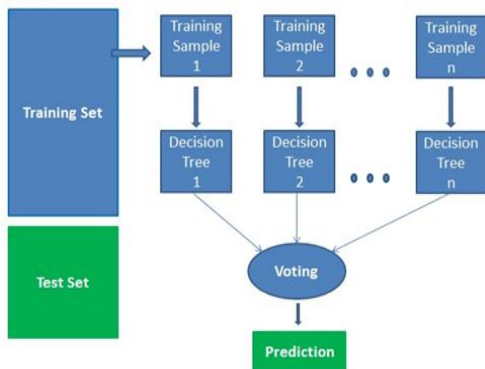


Figure 2 :Working of Random forest algorithm

Prediction of random forest for the classification problem:

$$f(x) = \text{majority vote of all predicted classes}$$

The advantage of Random forest is that it is one of the extremely accurate classifiers. As a several decision trees are involved in the process, it is robust. There is no overfitting problem, as it takes the average of all the predictions. It also handles missing values.

4. FUTURE WORK

As previously discussed, this paper predicts diseases from symptoms and recommends hospitals based on location and user ratings. This work can be further extended by predicting the diseases by analyzing the medical history of the other patients with similar symptoms and cure time taken by the patients. It can also involve the results of other medical conditions to enhance reliability.

5. CONCLUSION

Our Primary objective is disease prediction from symptoms using a machine learning algorithm. We utilized the Random forest algorithm to classify the symptoms of a patient. For creating this system, a web interface is made with two functionalities.

The first one was to get accurate disease prediction as output. The second one was to suggest hospitals when a location is fed as input. The significance of this system is low time consumption, accurate prediction, robustness and

computation. The accuracy of Random forest algorithm which is more than all supervised machine learning algorithm and time required for classification for Random Forest is less than all. So, we can conclude that Random Forest is better in terms of accuracy and time.

This system is a basic step for predicting the disorders. It is not an alternative for the prevailing medicinal practices and assessments. It miles just a initial step in figuring out the affected persons problems.

REFERENCES

[1] E.T.Venkatesh, P.Tangaraj, S. Chitra, **Classification of cancer gene expressions from micro-array analysis**, International Conference Innovative Computing Technologies (ICICT), 2010. <https://doi.org/10.1109/ICINNOVCT.2010.5440095>

[2] E.T. Venkatesh , P. Thangaraj , and S. Chitra ,**An Improved Neural Approach for Malignant and Normal Colon Tissue Classification from Oligonucleotide Arrays** European Journal of Scientific Research , vol. 54 , pp. 159 – 164 , 2011.

[3] GeethaYadav, Yugal Kumar, G.Sahoo, **Predication of Parkinson’s disease using Data Mining Methods: a comparative analysis of tree, statistical and support vector machine classifiers**, National Conference on computing and communication systems,2015.

[4]Mani Shankar, Mayank Pahadia, Divyang Srivastava, Ashwin T S, G. Ram Mohana Reddy,**A Novel Method for Disease Recognition and Cure Time Prediction Based on Symptoms**, Second International Conference on Advances in Computing and Communication Engineering, pp.679-682, 2015. <https://doi.org/10.1109/ICACCE.2015.66>

[5]Prabakaran .N and Kannadasan. R, **Prediction of Cardiac Disease Based on Patient’s Symptoms**, Second International Conference on Inventive Communication and Computational Technologies, pp. 794-799, 2018. <https://doi.org/10.1109/ICICCT.2018.8473271>

[6]Shaon Bhatta Shuvo, Joyoshree Ghosh,AtiaSujanaOyshi, **A Data Mining Based Approach to Predict Autism Spectrum Disorder Considering Behavioral Attributes**, 10th International Conference on Computing, Communication and Networking Technologies, 2019.

[7]Divya Krishnani, Anjali Kumari, Akash Dewangan, Aditya Singh, NenavathSrinivasNaik, **“Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms”**, TENCON 2019- IEEE Region 10 Conference, pp.367-372, 2019.

[8]Sreelekha Panda,Satyasis Mishra, Manas Ranjan Senapati, **Detection and Classification Methods for EEG Epileptic Seizures**, International Journal of Advanced Trends in Computer Science and Engineering, pp.2925-2934, 2019. <https://doi.org/10.30534/ijatcse/2019/40862019>

[9]Adil Alharthi, Nouf Alzahrani, Ikram Moalla **Convolutional Neural Network based on Transfer Learning for Medical Forms Classification**, International

Journal of Advanced Trends in Computer Science and Engineering, pp.3405-3411, 2019.

[10] Vijayakumar, P, Pandiaraja, P, Balamurugan, B & Karupiah, M 2019, **A Novel Performance enhancing Task Scheduling Algorithm for Cloud based E-Health Environment**, International Journal of E-Health and Medical Communications, (IGI-Global), Vol 10, Issue 2, pp 102-117.

<https://doi.org/10.4018/IJEHMC.2019040106>

[11] P. Pandiaraja, N Deepa 2019, **Hybrid Context Aware Recommendation System for E-Health Care by merkle hash tree from cloud using evolutionary algorithm**, Journal of Soft Computing, Springer.

[12] P. Rajesh Kanna and P. Pandiaraja 2019, **An Efficient Sentiment Analysis Approach for Product Review using Turney Algorithm**, Journal of Procedia Computer Science, Elsevier, Vol 165, Issue 2019, Pages 356-362.

<https://doi.org/10.1016/j.procs.2020.01.038>