



Health Disease Prediction using Logistic Regression

Kishore Bezawada¹, Unnam Gopi²

¹ Assistant Professor of Computer Science, Badruka College of Commerce & Arts, Kachiguda, Hyderabad.

² Assistant Professor, Badruka College of Commerce & Arts, Kachiguda, Hyderabad.

ABSTRACT :

Health diseases are rising challenge in India. These have become a burden for the society and prevention and control of these Health diseases is still a challenge for the government. Large portion of population in India is infected from this disease every year. Due to the diversity in geographical and living standard of people, it becomes difficult to control these diseases at early stages in the present system. The goal of this study is to investigate the symptoms and study the influence of clinical test parameters that belong to Health disease. The main aim of this study is to develop a prediction model using machine learning techniques for Health diseases. The prediction model has been prepared with the help of Logistic Regression. At the end, study proposes a prediction model for Health disease capable of diagnosing Health disease at each stage of Health Disease. Logistic Regression gives accuracy of around 91% for prediction of Health Disease.

Keywords: Health Disease, Logistic Regression, Machine Learning, Artificial Intelligence, Prediction.

INTRODUCTION

Health diseases affect a large part of world population. Every country in the world is suffering from these diseases annually. These diseases have become extensive burden for the developing countries like India, Bangladesh, Nepal etc. Common Health diseases namely Health Disease, malaria, chikungunya and kala azar are seasonal diseases spreading due to environmental changes and geographical situations.

In India, Health diseases are spreading in different parts of the country. Due to diversity in environment, climate and geographical situations, there is a big challenge for controlling and preventing these kinds of diseases. There are five major Health diseases namely Malaria, Health Disease, Chikungunya, Kala Azar, which are very common diseases in India. There is a big challenge to prevent and control these diseases in India. In rural area, it has become more complex job to control these vector diseases due to the lack of proper medical facilities and infrastructure. So, it becomes necessary to find

solution to these outbreaks using latest 2 technologies such as Information Technology and Machine Learning. There is also a dire need of technology-based platforms which can reduce the gap between urban and rural by providing them medical facilities with the help of Machine learning. This chapter provides an overview of machine learning and its application on Health diseases.

EXISTING SYSTEM

This paper introduces a novel approach for Health Disease fever classification based on online learning paradigms. The proposed approach is suitable for practical implementation as it enables learning using only a few training samples. With time, the proposed approach is capable of learning incrementally from the data collected without need for retraining the model or redeployment of the prediction engine. Additionally, we also provide a comprehensive evaluation of machine learning methods for prediction of Health Disease fever.

The input to the proposed pipeline comprises of recorded patient symptoms and diagnostic investigations. Offline classifier models have been employed to obtain baseline scores to establish that the feature set is optimal for classification of Health Disease. The primary benefit of the online detection model presented in the paper is that it has been established to effectively identify patients with high likelihood of Health Disease.

EXISTING TECHNIQUE:

SVM & Random Forest

The first step in developing the model was to ascertain that the feature representation for the data samples was adequate for Health Disease detection. For this task, we evaluated the classification accuracy on the dataset using SVM classifier for multiple train-test split ratios. It can be seen that though the initial classification accuracy is low. The results using random forest classifier with varying the number of decision trees. It may be noted that the train-test ratio for these results is kept at 80-20. The highest classification accuracy obtained using random forests for our dataset is maximum.

DRAWBACKS:

- It does not execute very well when the data set has more sound i.e., target classes are overlapping.
- It doesn't perform well when we have large data set because the required training time is higher.

- A large number of trees can make the algorithm too slow and ineffective for real-time predictions.

PROPOSED SYSTEM

World Health Organization recognizes that more than 17% of all infectious diseases is accounted by Vector-borne diseases. Around the world, in 97 countries malaria transmission occurs which puts about 3.4 billion people at risk. World's population of over 40% is at risk of Health Disease. In 2006, the outbreak spread of Chikungunya was in several countries, including India, where 1,400,000 cases were reported. Countries like the United States of America, reported 51,258 vector-borne disease cases in 2013. The Global Vector Control Response recognises vector control as elementary approach to prevent the Health diseases and as a response in order to outbreak.

This paper focused on effective prediction of the Health disease outbreak of three diseases (Chikungunya, Malaria, Health Disease....) across the Indian-subcontinent. We have examined and refined our model over data collected across India in 2013-2017. We have put forward a Logistic Regression outbreak risk prediction algorithm using contrasting data.

PROPOSED TECHNIQUE:

Logistic Regression model

Logistic Regression is a parametric classification model which is a supervised classification algorithm where the model is trained with target labels. According to Eureka Definition, "Logistic Regression is a method used to predict dependent variable out of given a set of independent variables such that the dependent variable is a categorical". In logistic regression, linear relation is not required and prediction values fall under 0 and 1. The application of logistic regression is used to solve the classification problems like medical diagnosis, cyber security, image processing etc.

ADVANTAGES:

It is easier to implement, interpret, and very efficient to train.

- It can easily extend to multiple classes (multinomial regression) and a natural probabilistic view of class predictions.
- It is very fast at classifying unknown records

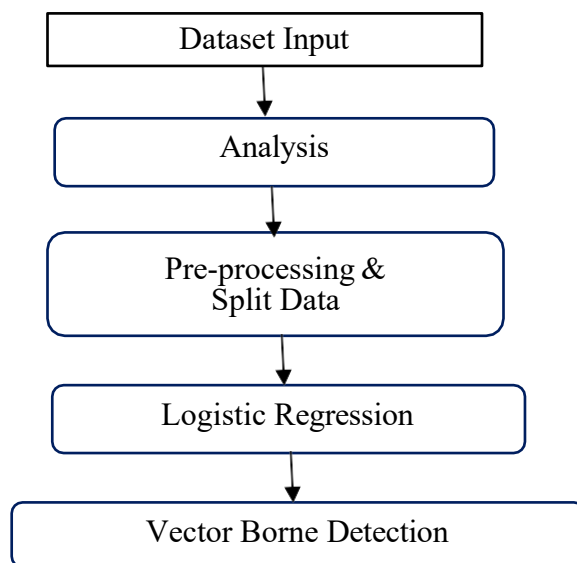


Figure 1: Proposed Model for Health Detection

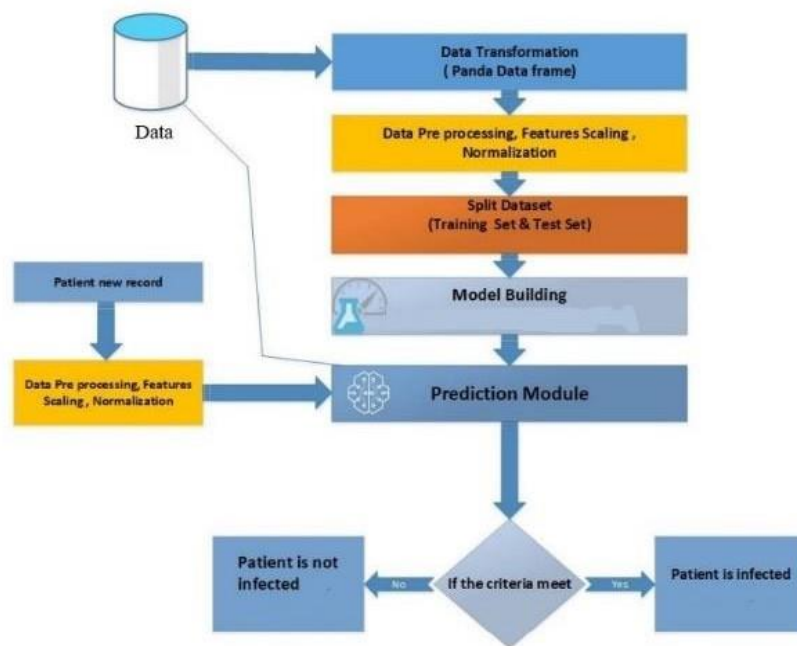


Figure 2: Proposed Methodology of Health Disease

RESULTS AND ANALYSIS

The investigator has studied twenty research papers related to Health Disease symptoms and parameters. From this study, the investigator selects important symptoms as features from each research paper. From here, the investigator prepares a table of symptoms related to Health Disease diseases. In the table, the researcher has calculated frequency of occurrence of each symptom. This table shows higher frequency symptoms that are found to be significant for detecting the Health Disease at early stage using machine learning techniques and methods.

The table 1 shows important factors for detection of Health Disease. This table also shows frequency of each symptom pertaining to Health Disease. The important factors pertaining to Health Disease detection are as shown in table 1. In this figure fever, Muscle Pain and Vomiting etc. are shown through bars. An investigator calculates the rank of each symptom as shown in table 1.

Serial Number	Reference	Fever	Headache	Skin Rash	Joint Pain	Muscle Pain	Pain Behind Eyes	Nausea	Abdominal Pain	Itching	Fatigue	Diarrhoea	Yellow Skin	Chills	Cough	Myalgia	Vomiting	Bleeding from Nose and Gum	Rhinorrhoea	Fatigue	Hematemesis	Enlarge Liver	Absence of Cough	Arthralgia	Respiratory	Shock	Bone Pain	Sore Throat	Gastric Bleeding	Swelling	Respiratory System	Petechiae
1	[66]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
2	[67]	✓	✓	✓					✓	✓						✓	✓	✓		✓	✓	✓										
3	[68]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
4	[69]	✓	✓	✓		✓			✓							✓	✓	✓	✓	✓	✓	✓										
5	[70]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
6	[71]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
7	[72]	✓	✓	✓	✓	✓			✓							✓	✓	✓		✓	✓	✓										
8	[73]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
9	[74]	✓	✓	✓		✓	✓	✓	✓							✓	✓	✓		✓	✓	✓		✓	✓	✓						
10	[75]	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	✓		✓	✓	✓					✓	✓				
11	[76]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓					✓	✓				
12	[77]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
13	[78]	✓	✓	✓		✓	✓	✓	✓							✓	✓	✓		✓	✓	✓										
14	[79]	✓	✓	✓		✓	✓	✓	✓							✓	✓	✓		✓	✓	✓					✓	✓				
15	[80]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
16	[81]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
17	[82]	✓	✓	✓		✓	✓	✓	✓							✓	✓	✓		✓	✓	✓										
18	[83]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓					✓	✓				
19	[84]	✓	✓	✓	✓	✓	✓	✓	✓							✓	✓	✓		✓	✓	✓						✓	✓			
20	[85]	✓	✓	✓					✓							✓	✓	✓		✓	✓	✓										
Frequency		19	19	13	4	1	10	8	11	2	11	8	1	19	3	10	10	10	1	2	1	1	3	5	1	1	2	4	3	1	1	2

Table 1: Identification of features pertaining to Dengue Disease

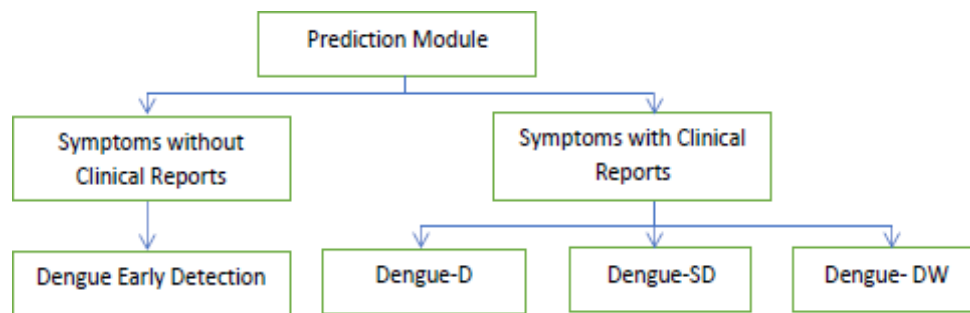


Figure 3: Prediction module of proposed prediction model for Health diseases (Dengue)

FUTURE ENHANCEMENTS

Each year, the controlling of Health diseases is a challenging task for India. Due to diversity of climate in different geographical locations of the countries, there is a probability of spreading Health diseases. This research contributes a prediction model which diagnoses Health Disease at early stages and then classifies the types of Health Disease. The present research is focused on preparing a prediction model to diagnose Health Disease at early stages and then classifying the same. A number of steps are followed to prepare a prediction model to achieve the objectives of this research. A research methodology has been adopted to achieve all the four objectives of this research. The research opens new avenues to prepare a prediction model for other Health diseases. There is a huge possibility to conduct research in classifying and handling critical patients of Health Disease as well as other Health diseases. The findings of this research provide a platform to detect and classify Health Disease. The research results show that the proposed prediction model has achieved 97.5% accuracy, which helps the medical team in taking timely decisions and saves the lives of Health Disease patients. This research can further be extended to other Health diseases namely chikungunya, Zika, Kala Azar etc. with the help of advanced machine learning algorithms.

CONCLUSION

This work, a new artificial neural network-based multimodal outbreak prediction (CNN-MDOP) algorithm is proposed with the use of contrasting data. To the best of our understanding, none of the existing works have focused on contrasting data in the area of analysis of medical data. The prediction accuracy of our suggested CNN algorithm is 88%.

REFERENCES

1. Md. Osman Goni Nayeem, Maung Ning Wan and Md. Kamrul Hasan on "Prediction of disease level using multilayer perceptron of Artificial Neural Network for patient monitoring" *International Journal of Soft Computing and Engineering (IJSCE)* Volume-5 Issue-4: Pgs 17-23, September 2015
2. Sivaranjani.R and Dr.N.Yuvaraj on "Artificial Intelligence model for earlier prediction of cardiac functionalities using multilayer perceptron" *International Conference on Physics and Photonics Processes in Nano Sciences*
3. Mahalakshmi and G. Suseendran on "Prediction of Zika Virus by multilayer perceptron neural network (MLPNN) using Cloud B" *International Journal of Recent Technology and Engineering (IJRTE)*, Volume-8 Issue-2S11: Pgs. 249- 254, September 2019
4. V. Janani , N. Maadhuryaa , D. Pavithra and S. Ramya Sree on "Health Disease Prediction using multilayer perceptron - a Machine Learning approach" *International Journal of Research in Engineering, Science and Management* Volume-3, Issue-3, March 2020
5. Agus Qomaruddin Munir and Edi Winarko on "Classification model disease risk areas endemicity Health Disease Fever outbreak-based prediction of Patients, death, IR and CFR using forecasting techniques" *International Journal of Computer Applications*, March 2015
6. Juan M. Scavuzzoa , Francisco Truccoa , Manuel Espinosac , Carolina B. Taurob , Marcelo Abrilc , Carlos M. Scavuzzob ,and Alejandro C. Freryd Modeling on "Health Disease vector population using remotely sensed data and Machine Learning" in press.
7. Zulkiflee, N. F., & Rusiman, M. S. (2021). Heart Disease Prediction Using Logistic Regression. *Enhanced Knowledge in Sciences and Technology*, 1(2), 177-184.