International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Disease Prediction Using Machine Learning

Revati Shinde[1], Snehal Sawant[2], Tushar Maskar[3], Prof. Kirti Randhe[4]

[1-3]*Student, Dept. of Computer Engineering, ISB&M College of Engineering, Pune, India*
[4]*Professor, Dept. of Computer Engineering, ISB&M College of Engineering, Pune, India*

Abstract - *Nowadays people face different diseases because of the state of the climate and their living habits. While earlier prediction of disease becomes an essential task. But it becomes too difficult for the doc-tor to determine precisely on the basis of symptoms. Data mining plays an important part in predicting the disease in overcoming this problem. Medical science has a substantial increase in data per annum. We proposed general pre-diction of the disease based on the patient 's symptoms. For the prediction of disease, we use machine learning algorithm Support Vector Machine (SVM) for accurate prediction of disease. Dataset of disease symptoms needed for prediction of disease.*

***Key Words*: Support Vector Machine (SVM), Machine learning, Disease Prediction.**

1. INTRODUCTION

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. There are numerous kinds of Machine Learning Techniques like Regression, Semi Supervised, Supervised, Reinforcement, Evolutionary Learning, Classification and Deep Learning. These techniques are used to classify huge data very fastly. So, we use Support Vector Machine (SVM) machine learning algorithm for fast classification of big data and accurate prediction of disease. Due to the fact that medical data is growing day by day, the use of this is critical to predict correct dis-ease, but processing big data is very difficult in general, so data mining plays a very important role and classification of large datasets using machine learning becomes so simple.

The primary focus is on to use machine learning in healthcare to enhance the quality of care for improved outcomes. Machine learning has made it easier to accurately classify and diagnose the various diseases. With the aid of effective multiple machine learning algorithms, predictive analysis helps predict the disease more accurately and helps treat patients.

The healthcare industry generates regular vast volumes of healthcare data that can be used to collect information for disease prediction that will happen to a patient in the future by using the history of treatment and health data. This secret knowledge in the healthcare data would later be used to make affective decisions about the health of patients. These areas also need enhancement by using the healthcare in-formative data.

2. LITRATURE REVIEW

1. Paper Name: Disease phenol type similarity improves the prediction of novel disease associated micro RNAs
Author: Duc-Hau Le
Abstract: Several studies have shown the role of miRNAs (microRNAs) in human disease, and a range of computational methods have been proposed by rating candidate microRNAs ac-cording to their relevance to a disease to predict such associations. Network-based approaches are becoming dominant among them as they make effective use of the "disease module" concept in miRNA networks of functional similarity. Of which, the algorithm-based Random Walk with Restart (RWR) method on a functional similarity network miRNA, namely RWRMDA, is state-of-the-art. The use of this algorithm was motivated by its performance in predicting disease genes since the concept of "disease module" also exists in networks of protein interaction. In addition, for the prediction of disease genes, several other algorithms have also been de-signed. Nevertheless, they have not been used yet for disease microRNA prediction. In this research, we suggested a method for prediction of dis-ease-associated miRNAs, namely RWRHMDA. This approach was based on the RWRH algorithm, which was successfully proposed in a heterogeneous network of genes and disease phenotypes for disease gene prediction. In particular, we used this algorithm to rank candidate miRNAs for disease on a heterogeneous network of phenotypes and miRNAs, which was developed by integrating a functional similarity network of shared target gene-based microRNA and a similarity net-work of disease phenotypes. We found that RWRHMDA significantly

outperformed RWRMDA regardless of parameter settings by comparing the pre-diction performance of RWRHMDA with that of RWRMDA on a set of 35 disease phenotypes, because it better exploited the "disease module" concept. In addition, eight novel Alzheimer's disease-associated miRNAs were identified using the RWRHMDA system.

2. Paper Name: Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis
Author: Shang-Ming Zhou
Abstract: i) selection of variables by comparing relative frequencies of Read codes in the primary care dataset associated with disease case compared to non-disease control (disease/non-disease based on the secondary care diagnosis); ii) reduction of predictors/associated variables using a Random Forest method, iii) induction of decision rules from decision tree mod-el. The proposed method was then extensively validated on an independent dataset, and compared for performance with two existing deterministic algorithms for RA which had been developed using expert clinical knowledge.

3. Paper Name: Design and Implementing Heart Disease Prediction Using Naives Bayesian
Author: Anjan Nikhil Repaka, Sai Deepak Ravikanti
Abstract: Datamining, a fantastic technique of creation that revolves around discovering and digging out valuable data from vast data collection, which can be further useful in analyzing and drawing up trends for making business-related decisions. Speaking of the medical field, the application of da-ta mining in this field will lead to the discovery and withdrawal of important trends and knowledge that can be useful in clinical diagnosis. The thesis focuses on the diagnosis of heart disease by considering previous evidence and knowledge. In order to predict risk fac-tors for heart disease, SHDP (Smart Heart Disease Prediction) is construct-ed via Navies Bayesian to achieve this. The rapid development of technology has contributed to a notable increase in the online use of mobile health technology. In a standardized fashion, the necessary data is assembled. The following characteristics are collected from medical profiles to predict the risks of heart disease in a patient, including: age, BP, cholesterol, sex, blood sugar, etc. For the Navies Bayesian classification for the prediction of heart disease, the collected attributes serve as an input. The used dataset is divided into two parts, 80% of the dataset is used for training and 20% of the remainder is used for research. The proposed method involves the following phases: selection of datasets, user registration and login (Application-based), Bayesian Navy classification, prediction and stable data transfer using AES (Advanced Encryption Standard). Results are then produced. By making use of data mining approaches that are adopted for heart disease prediction, the study elaborates and introduces several information abstraction techniques. The result shows that the diagnostic framework developed effectively helps predict risk fac-tors for heart disease.

4. Paper Name: Application of Machine Learning in Disease Prediction
Author: Pahulpreet Singh Kohli, Shriya Arora
Abstract: In the field of medical diagnosis, the application of machine learning is rapidly growing. This will mainly lead to improving the classification and identification systems used in disease detection, which are able to provide data that help medical experts identify deadly diseases early and thereby dramatically improve patients' survival rates. We apply different classification algorithms in this paper, each with its own advantage on three separate disease databases (Heart, Breast Cancer, Diabetes) available in the UCI disease prediction repository. By backward modelling using the p-value test, the feature selection for each dataset was achieved. The findings of the study support the principle of applying machine learning to early disease detection.

5. Paper Name: Diabetes Disease Prediction Using Data Mining
Author: Deeraj Shetty, Kishor Rit, Sohail Shaikh
Abstract: A subfield in the field of software engineering is data mining. It is the methodical process of discovering examples in large data sets, including manufactured intelligence crossing point methods, machine learning, observations, and database systems. The purpose of the technique for data mining is to think of data from a data set and turn it into a reasonable framework for further use. Our analysis focuses on this aspect of the Medical Conclusion Learning Design through the diabetes data collected and to build an emotionally supportive network of smart therapeutic choices to benefit doctors. The main purpose of this review is to assemble the Intelligent Diabetes Dis-ease Prediction Framework, which uses the database of diabetes patients to evaluate diabetes disease. In this meth-od, we propose the use of algorithms such as Bayesian and KNN (K-Nearest Neighbor) to apply and evaluate the database of diabetes patients by taking

different attributes of diabetes for diabetes disease prediction.

3. SYSTEM ARCHITECTURE

Initially we create disease dataset in the form of disease list with its required parameters and corresponding values. After that pre-processing is performed on that dataset for cleaning that is removing comma, punctuations and white places. And that is used as training dataset. After that feature extracted and selected. Then we classify that

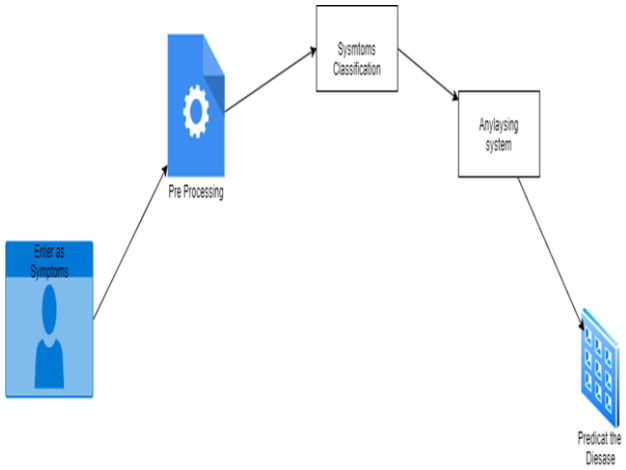

Fig -1: System Architecture

data using classification techniques such as Support Vector Machine (SVM). Based on machine learning we can predict accurate disease.

4. MATHEMATICAL MODEL

Let S be the Whole system

S = {D1, D2, D3}; Where, D1, D2, D3 are the types of Disease

S= {I, P, O}; I-input, P-procedure, O-output

Input (I),

I= {Personal Info of Patient, Categorical Feature, Value of Feature}

Where,

Checker -> Personal Info of Patient, Categorical Feature, Value of Feature,

Model ->Analyse Feature

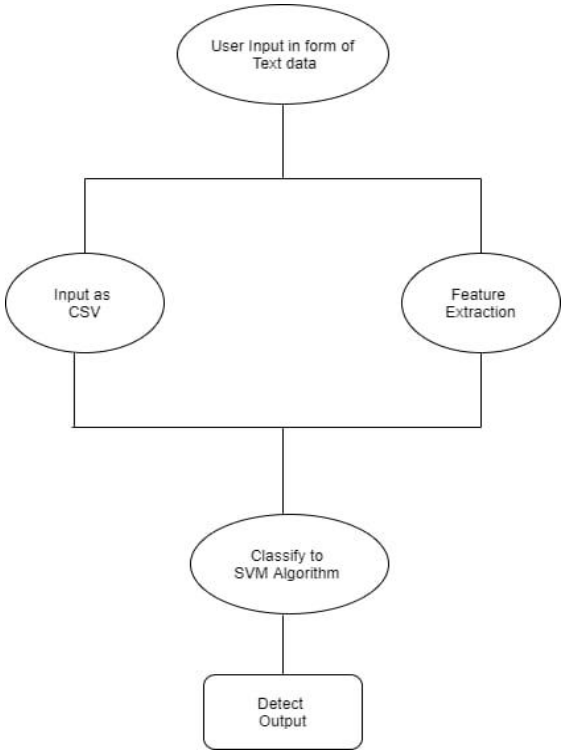

Fig -2: Mathematical Model

Procedure (P),

P= {I, discrete SVM Algorithm}

I – Patient's data

SVM Algorithm – For analysis of Features Value

Detect: Disease

Output(O),

O= {Disease Detected or Not}

5. CONCLUSION

Based on a machine learning algorithm, we proposed a general disease prediction system. We used Support Vector Machine (SVM) algorithms to classify patient data because medical data is increasing at an incredible rate, necessitating the processing of existing data in order to predict exact disease based on test results. By providing the input as patient records, we were able to obtain accurate general disease prediction as an output. Because of this system, disease prediction could be done in a short amount of time and at a low cost.

REFERENCES

[1] Duc-Hau Le, "Disease phenotype similarity improves the prediction of novel disease-associated microRNAs", "2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science"

[2] Shang-Ming Zhou, "Defining Disease Phenotypes in Primary Care Electronic Health Records by a Machine Learning Approach: A Case Study in Identifying Rheumatoid Arthritis", "May 2, 2016, https://doi.org/10.1371/journal.pone.0154515"

[3] Anjan Nikhil Repaka, "Design and Implementing Heart Disease Prediction Using Naives Bayesian", "Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8".

[4] Pahulpreet Singh Kohli, "Application of Machine Learning in Disease Prediction", "2018 4th International Conference on Computing Communication and Automation (ICCCA)"

[5] Deeraj Shetty, "Diabetes Disease Pre-diction Using Data Mining", "2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)"