e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703032 |

# Prediction of Probability of Disease Based on Symptoms

**Mr. Chougule V.V, Ms. Mamanabad A.A, Ms. Jadhav S.V, Ms. Shivagunde S.B, Ms. Mane N.D, Ms. Yadav S.N.**

**Lecturer,** Department of Computer Engineering, Shree Vatavruksha Swami Maharaj Devasthan's Kai. Kalyanrao(Balasaheb) Ingale Polytechnic College, Akkalkot, India

**Lecturer,** Department of Computer Engineering, Shree Vatavruksha Swami Maharaj Devasthan's Kai. Kalyanrao(Balasaheb) Ingale Polytechnic College, Akkalkot, India

Department of Computer Engineering, Shree Vatavruksha Swami Maharaj Devasthan's Kai. Kalyanrao(Balasaheb) Ingale Polytechnic College, Akkalkot, India

**ABSTRACT:** Today's health information management systems collect enormous volumes of healthcare data and information, including complex personal and medical history information. In order to find utilization patterns for research, medical data mining techniques are being used more and more. Nowadays, the greatest cause of death for humans is sickness; a single person may be afflicted with several ailments. This method attempts to predict sickness by using symptoms that are related to a patient's condition and behavior. Based on the user-provided health information, the system assesses symptoms and predicts illnesses with the use of a data mining algorithm. The symptoms of the condition may change the number of symptoms displayed by the search engine.

**KEYWORDS**: *Prediction, Symptoms, disease, algorithm, clustering*

## I. INTRODUCTION

Large amounts of information are produced and gathered everyday by healthcare companies nowadays. The suggested method will be put into place with the needs of the patients in mind in order to save them time. The primary cause of mortality in humans is sickness. The suggested system uses data analysis to incorporate symptoms associated with patients' conditions and behaviors, making it capable of predicting illness. The varied raw data from healthcare must be gathered and stored in orderly formats in order to be used for the early identification of human disease. The suggested method asks for symptoms, which it then analyzes and forecasts illnesses based on the patient's health. In addition to predicting diseases, this suggested system will suggest the right doctors based on a given ailment. The physician dataset list will be utilized for illness prediction as well as symptom assessment. The illnesses including polio, dengue, lung disease, and blood cancer would be predicted by this suggested approach. This method is easy to use. The nontrivial extraction of implicit, previously unknown, and possibly beneficial information from data recorded in databases is referred to as data mining, also known as knowledge discovery in databases. In order to obtain different and more relevant findings, KDD is an iterative process that allows assessment measures to be improved, mining to be optimized, and new data to be merged and changed. Although data mining and knowledge discovery in databases are sometimes used interchangeably, data mining is a step in the process of knowledge discovery.

## II. LITERATURE SURVEY

A literature review is a crucial component of project development. It is an analysis of the content related to the report's theme, the previously in place system. These could contain current system theories regarding the subject, completed research, difficulties encountered, and continuing projects. Understanding the project's risk and viability as well as adhering to best practices in project development are further benefits of conducting a literature review. The literature review also sheds light on a variety of platforms, operating systems, and tools that are appropriate for study and project development.

1. Prashanta Kumar Patra and Subasish Mohapatra The author of "Smart Health Care System using Data Mining" demonstrates how to extract datasets' hidden information. In addition to providing end-user support, the smart health

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703032 |

care management system enables users to navigate health challenges via an online platform. Because assessing, predicting, and detecting diseases requires well-organized procedures, this research is being conducted in the field of medical sciences. Applications for data mining are used in the administration of smart healthcare to identify and forecast diseases.

2. Yinglai Lin, Yuying He, Linyuan L¨ u, Xiaoping Min_, Alfonso Rodrñguez-Pat ´on, Xiangxiang Zeng, Senior Member, IEEE "In-depth cooperative filtering for illness gene prediction" The Inductive Matrix Completion (IMC) model is considered highly dependable due to its robust structure and exceptional ability to forecast gene-disease correlations. The experimental findings demonstrate that DCF is still useful for mining undiscovered associations and ranking unique illness characteristics.

3. Here, temporal mining, LVAD implants, electronic health records (EHRs), and EHR data can be exploited. Temporal mining techniques can be used to extract temporal patterns, such as changes between clinical occurrences across time. This has the advantage of turning extensive temporal data records into something that is simple to read and comprehend. Although difficult, temporal pattern extraction from massive clinical data sets is quite beneficial.

4. Three blood cancer classifiers have been identified for this study: support vector machines (SVM), decision trees (DS), and k-nearest neighbor (kNN). Leukemia has an impact on blood status and can be identified in the medical field utilizing the Blood Cell Counter (CBC). By employing data mining techniques to ascertain the associations between leukemia and blood characteristics and the variables of gender, age, and patient health condition, this study seeks to predict the existence of leukemia. According to the specified disease condition, the given data set is converted into binary format in earlier research. Another approach now in use gathers characteristics from the database to indicate risk level.

5. The categorization techniques of data mining employed in data discovery were the topic of Monika Gandhi and Dr. Shailendra Narayan Singh's work, "Predictions in Heart Disease Using Techniques of Data Mining." The goal of the current endeavor is to employ data mining techniques to uncover the features of using healthcare data to assist individuals. The primary goal is to propose an automated method that considers past data and information to diagnose cardiac problems.

6. The author employed data mining, genetic algorithms, BP neural networks, and heart disease prediction. For the purpose of predicting cardiac disease, this research suggests an effective hybrid genetic algorithm and back propagation technique approach.

7. "Process Mining for Project Management" by Jeni Joe, Yasha Ballal, and Tanya Emmatty We have introduced the idea of process mining in this study. Process mining and software development allow us to optimize the Software Development Lifecycle of IT organizations' projects.

8. The author employed bootstrap aggregating, exploratory under-sampling, logistic regression, and the exponential moving average (EMA) early warning system (EWS). The EMA EWS is intended to give patients on general hospital wards (GHWs) dependable early alerts. Based on the patient's current electronic medical record, EWS automatically identifies patients who are at risk of a clinical decline. EWS's primary objective is to solve a difficult classification issue on high-dimensional stream data that includes measurement errors, outliers, irregular, multi-scale data gaps, and class imbalance.

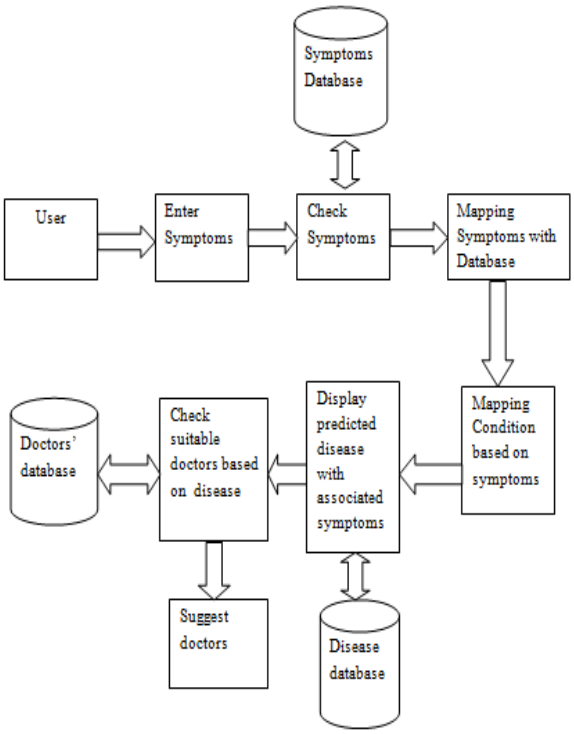
| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703032 |

## III. METHODOLOGY

We can forecast the disease based on symptoms by using the Prediction of Probability of Disease based on Symptoms tool. In addition to predicting the diseases, this suggested system would suggest the right doctors from the doctor list depending on a specific disease.




Initially, a lot of data is kept in an unorganized, noisy state. Enter symptoms according to the patient's state or behavior in the recommended work. The illness dataset was used in this to eliminate noisy data and missing values. This symptom's entry has been mapped with the database and is kept in the symptoms database. mapping the existing state of affairs and assessing the patient's status based on the patient's symptoms. The suggested system shows the anticipated disease along with its corresponding symptoms based on this. We can enter multiple symptoms into this proposed system, and the system will display the condition associated with each symptom as it increases in number. Following this procedure, the suggested system will also search the doctor database for an appropriate physician for a specific ailment, and the system will then present that physician.

In our system there are mainly 3 modules :
1. User module
2. Doctor module
3. Admin module

The user only needs to input the symptom into the user module to receive an accurate diagnosis and the name of the right physician. The user must first register in order to schedule an appointment with a recommended physician. He can schedule an appointment and obtain the doctor's details, including location, date, and time, after registering.
Physicians are able to self-register in the Doctor Module.

The user's scheduled appointment can only be verified by the physician.
The admin module allows the admin to log in with their username and password. Once a doctor registers, only the administrator has the authority to accept the appropriate doctors or deny phony accounts. The administrator may be able to add new illnesses and symptoms. Admin is in charge of the entire project.
The K-means clustering method is used in this project to make predictions.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

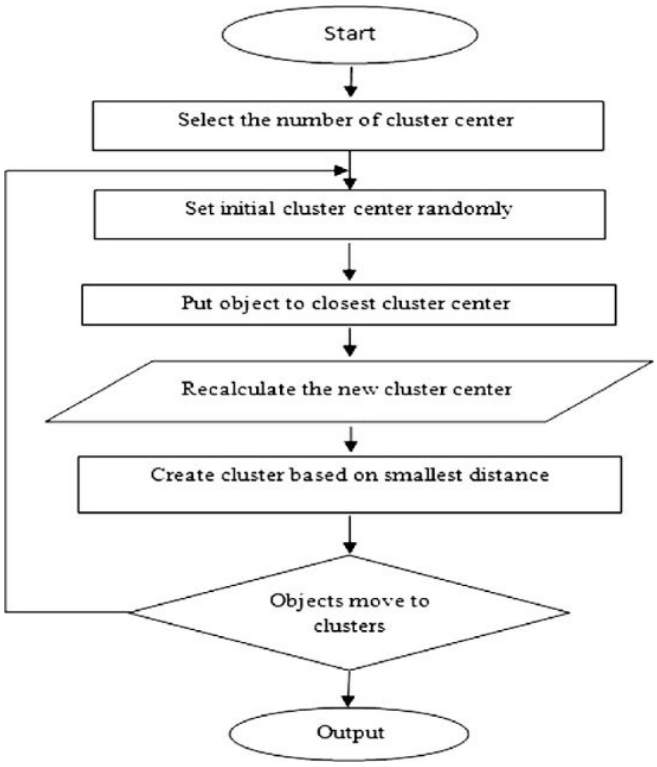
| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703032 |

The above diagram allows the user to insert each symptom individually. Our system is able to forecast the sickness by employing the k-means method.

**K-means algorithm**: The goal of the iterative K-means algorithm is to divide the dataset into K pre-specified, unique, non-overlapping subgroups, or clusters, with each data point belonging to a single group. The goal is to maintain as much distance between the clusters and as much similarity between the intra-cluster data points as feasible. It groups data points into clusters so that the total squared distance between the cluster's centroid—the arithmetic mean of all the data points in that cluster—is as small as possible. The homogeneity (similarity) of data points within a cluster increases with decreasing variation within the cluster.

The following is how the k-means algorithm operates:
1. Indicate the number K of clusters.
2. To initialize centroids, shuffle the dataset first, and then choose K data points at random, without replacement, for the centroids.
3. Continue iterating until the centroids remain unchanged. i.e., the distribution of data points among clusters remains constant.
4. Total the squared distances between each centroid and each data point.
5. Assign every data point to the centroid, or nearest cluster.
6. By averaging all of the data points that are part of each cluster, determine the centroids for each cluster.




When a patient or user enters a single disease, all diseases with the symptoms the patient entered will be shown. The user will be able to identify the precise illness if he provides precise symptoms. Additionally, based on a certain ailment, the disease system recommends the right doctors from the doctor list.

## IV. CONCLUSION

In this approach, we can use the K-means algorithm to forecast diseases based on their symptoms and recommend the right specialist for each individual disease.
According to the article, we should use Restful web services to create an online application that will enable doctors and patients to communicate more quickly.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703032 |

## V. HARDWARE AND SOFTWARE REQUIREMENT

Operating system : Windows 8/10
Language : PHP (5.5.6)
Server : Xampp (1.8.3)
Database : MySQL (5.6.14)
Web Technologies: HTML5, CSS3, JavaScript, Ajax, Query, PHP
Web Browser : Google Chrome

## REFERENCES

[1] "Smart Health Care System using Data Mining," by Prashanta Kumar Patra 2018 IEEE
[2] "Mining temporal patterns from sequential healthcare data" by Faezeh Movahedi, Yiye Zhang, Rema Padman, James F. Antaki, and others IEEE International Healthcare Informatics Conference, 2018
[3] Arulraj AM, Sugumar, R., Estimating social distance in public places for COVID-19 protocol using region CNN, Indonesian Journal of Electrical Engineering and Computer Science, 30(1), pp.414-424, April 2023.
[4] Meghna Sharma and Ankita Dewan "Hybrid Technique for Predicting Heart Disease in Data Mining Classification," IEEE, 2015.
[5] "Prediction and Diagnosis of Leukemia Using Classification Algorithms" by Khaled A. S. Abu Daqqa, Ashraf Y. A. Maghari, and Wael F. M. Al Sarraj The eighth International Conference on Information Technology (ICIT) was held in 2017.
[6] Monika Gandhi and Dr. Shailendra Narayan Singh "Data mining techniques for heart disease prediction," IEEE 2015, 978-1-4799-8433-6/15/$31.00
[7] P. Borra, Amazon Redshift, https://praveenborra.com/2020/11/17/amazon-redshift/, 2020.
[8] "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification," Ankita Dewan and Meghna Sharma, IEEE, 2015.
[9] "Process Mining for Project Management," by Jeni Joe, Yasha Ballal, Tanya Emmatty, and Snehal Kulkarni

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com