

Healthcare Portal System using Iterative Search, K-Nn, Differential Diagnosis, Baye's Theorem

Mathpati Dhondutatya¹, Memon Simran¹, Kesgire Akash¹, Karande Siddharth¹

Student, Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India¹

Abstract: Big data is now rapidly expanding in all science and engineering domains including physical, biological and biomedical sciences. Medical data are an ever growing source of information generated from hospitals consisting of patient records in the form of hard copies which can be made easier and convenient by using E-copies of the patient details. Our aim is to build a Health-care Portal system which will provide the features like clinical management, patient records and diagnosis of chronic diseases. Proposed system will be useful in emergency and also when the patient is relocating. Iterative search algorithm helps in symptom matching and providing possible diseases ranked according to number of symptoms matched in the database. The k-NN algorithm is applied to classify the diseases into subgroups and higher preference subgroups are displayed. Differential diagnosis is used to narrowing down the diseases to the root disease out of the list of diseases that show similar symptoms. Baye's theorem is used to find the probable occurrence of disease.

Keywords: Data mining, Differential diagnosis, k-NN, Baye's theorem.

I. INTRODUCTION

Since the advent of advanced computing, now data mining is making a really good impact in medical domain. We are using the technology in various possible ways, from surgical imagery to X-ray photography. But the technology has always stayed behind when it came to disease diagnosis. To overcome various problems like medical history to climatic conditions, blood pressure environment and various other factors medical decision support systems[1]-[2] are becoming more and more popular.

Medical system is highly specialized and challenging job due to various factors, especially in cases of diseases that shows the similar symptoms. Problems may vary from inexperience of doctors to stress and fatigue, due to this, ambiguities in symptom matching, misinterpretation may arise.

Normally the doctors treat the patients based on their previous knowledge about the same kind of symptoms and also by referring the patient's family history. But sometimes doctors not being able to come to a final decision about what kind of disease the particular patient has. Sometimes the process of treating the patient goes so longer leading to waste of money and time both and also it might happen that the patients health might come at risk.

The proposed system in this paper will be useful to patient as well as doctor, thus making the task of doctor easier. Here we are using the four different algorithms to achieve it. Proposed Healthcare Portal uses the Iterative search algorithm for symptom matching and providing the possible diseases ranked according to the number of symptoms matched in the database. The k-NN algorithm is applied to classify the diseases into subgroups and higher preference subgroups are displayed. Differential diagnosis is used to narrowing down the diseases to the root disease

out of the list of diseases that show similar symptoms. Baye's theorem is used to find the probable occurrence of the disease.

The Health Care portal takes care right from scheduling appointments to maintaining medical history and health records, we are hostage to redundant processes and repetitive investigations. The patients today need a cared network they can trust and be in control of their medical documents and history so that it is available to them whenever they need, to get the right treatment anywhere, any time.

Patients can maintain their complete medical history and data on healthcare portal. Since the same platform connects to Hospitals, Doctors, all their ongoing treatment information . with their Doctors, they can also search and make appointments with other specialist Doctors.

II. LITERATURE SURVEY

Medical diagnosis is an important domain of research which aids to identify the occurrence of a disease. There is a proposed novel guide path to knowledge discovery in medical systems by acquiring relevant information from the data set. This approach makes diagnosis easier. Using naive baye's, the overall speed and accuracy of the algorithm increased and extract high quality data set from an unstructured text. The primary advantage of the scheme is that it can be used to whatever sort of dataset whether it is a predefined dataset or not [3-4].

One of the well-springs of mathematical inspiration has been the continuing attempt to formalize human thought. From the syllogisms of the Greeks, through all of logic and probability theory, cognitive models have led to beautiful mathematics and wide ranging application. But



mental processes have proven to be more complex than any of the formal theories and the various idealizations have broken off to become separate fields of study and application. It now appears that the same thing is happening with the recent developments in connectionist and neural computation. Starting in the 1940s and with great acceleration since the 1980s, there has been an effort to model cognition using formalisms based on increasingly sophisticated models of the physiology of neurons. Some branches of this work continue to focus on biological and psychological theory, but as in the past, the formalisms are taking on a mathematical and application life of their own. Several varieties of adaptive networks have proven to be practical in large difficult applied problems and this has led to interest in their mathematical and computational properties [5].

Our capabilities of both generating and collecting data have been increasing rapidly in the last several decades. Contributing factors include the widespread use of bar codes for most commercial products, the computerization of many business, scientific and government transactions and managements, and advances in data collection tools ranging from scanned texture and image platforms, to on-line instrumentation in manufacturing and shopping, and to satellite remote sensing systems. In addition, popular use of the World Wide Web as a global information system has aided us with a tremendous amount of data and information. This explosive growth in stored data has generated an urgent need for new techniques and automated tools that can intelligently assist us in transforming the vast amounts of data into useful information and knowledge[6].

III. ALGORITHMS

1) ITERATIVE SEARCH:

The first step of the algorithm involves selecting the symptoms mentioned by the patient. As an output, the algorithm gives the list of all possible diseases ranked according to the number of symptoms matched in the database. The list is generated after input of every symptom. After the first iteration, for the second iteration, the next list of symptoms will be shortlisted according to the disease list that was obtained in the previous iteration. The new symptom list will contain symptoms of only those diseases that were obtained in the previous list. These related symptoms will then be shown to the user who short lists another symptom from the new list[8].

2) k-NN:

This algorithm subcategories the disease. K-NN classification is applied to classify diseases into subgroups. If a group of symptoms match higher preference is given to that subgroup and searching in that new smaller subgroup thus reduces database access. In pattern recognition, the k-nearest neighbour algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of

instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. This feature has been identified as the most suitable for the present system. The k-NN algorithm takes input as the output of iterative search which is diseases ranked according to the symptoms matched in the database. After that the k-NN algorithm finds the nearest subgroup from symptoms matched in the database as like it finds the weather the particular disease is from endocrine, cardiovascular, pericarditis, neurological etc category of the disease.

3) DIFFERENTIAL DIAGNOSIS:

Differential diagnosis is used to narrowing down the disease to root disease. It takes input as a list of diseases from k-NN algorithm. It utilizes the database of a comprehensive list of medical history. With the help of this the probability of occurrence of a disease may be calculated. The algorithm will give the output as list of probable diseases from the symptoms entered by the patients.

Patient	Disease Diagnosed	Actual Disease
A	Diabetes	Diabetes
B	Diabetes	Diabetes
C	Diabetes	Diabetes
D	Diabetes	Diabetes
E	Diabetes	Diabetes
F	Diabetes	Hypertension
G	Diabetes	Hypertension
H	Diabetes	Hypertension
I	Diabetes	Arthritis
J	Diabetes	Arthritis

4) Baye's Theorem:

To find the occurrence of disease It is used to obtain the probability of disease in a group of people with some characteristic on the basis of the overall rate of that disease and of the likelihood of that characteristic in healthy and diseased individuals. The most familiar application is in clinical decision analysis where it is used for estimating the probability of a particular diagnosis given the appearance of some symptoms or test result.

Formula: $P(A/B) = P(B/A) P(A) / P(B)$

IV. SYSTEM ARCHITECTURE

The working of a Healthcare Portal is as shown in figure. At first the doctor as well as patient needs to register himself in the portal. Here we are assigning a unique id to patient as well as doctor. Hence the duplication of record will not take place. By using that unique patient id, doctor id the patient as well as doctor will get the access to the portal respectively.

Here we are using the client-server architecture which is implemented on a web server so that it can be accessed by anyone with an internet connection. The server improves the delivery of important instructions and make sharing of data possible across the whole community of healthcare professionals more



practical in cost, security and risk of deployment. For implementation of HealthCare portal here we are making the use of web services like pattern matching, recent trends, differential diagnosis and recent trend diagnosis.

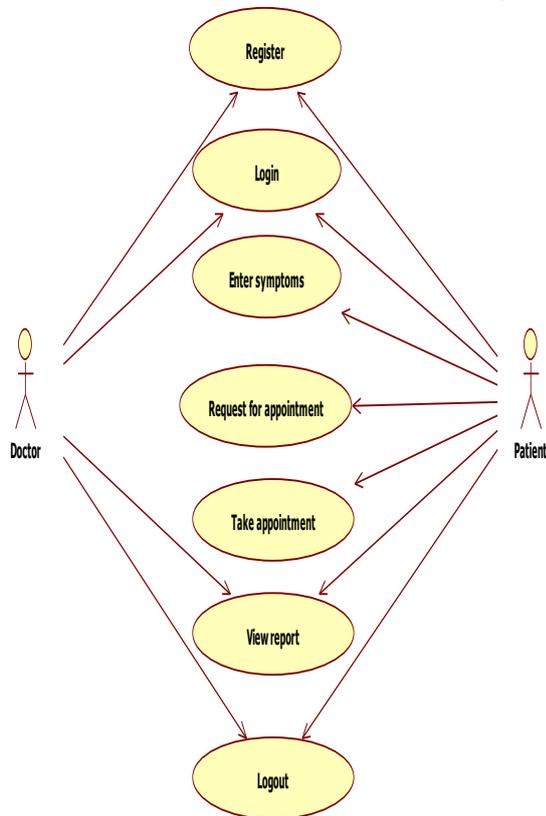


Fig: Log in System

The portal maintains the two databases first one is Disease/Symptoms database which is centralized one and it contains the list of the existing known diseases and their corresponding symptoms. The second database is named as the records database which maintains the records of the patients in the network from all hospitals.

System architecture is as shown in figure below. At first patient enters the system with his log in credentials and then the patient enters the symptoms in the portal which are observed by him then symptoms are matched in the existing database with the possible diseases. According to that the possible diseases are then shown to the patient. At that point the iterative algorithm will be applied in the system and the possible diseases are shown to the patient arranged in decreasing order of the symptoms matched in the database. The rank will be given to each disease. Higher rank disease will be on top position.

This process will work iteratively and the pattern matching will happen. In second stage k-NN algorithm will be applied which categories the diseases into sub-categories. The searching will happen in disease and symptoms database. Then the results are forwarded to the differential diagnosis.

Finally to avoid ambiguity in decisions the portal will make use of differential diagnosis which makes use of the

Disease/symptoms Database and Record Database along with the results obtained from iterative and k-NN algorithms to get the results. Then the specialist doctors of that diseases are shown to the patient. With the reference of that the patient will be able to take the appointment of the doctor. After taking the appointment patient will meet the doctor on specified date [7]. With the help of his/her login credentials doctor will be getting the access of the portal then the Doctor will do his/her regular check-up and will be able to see the symptoms of the patient then the doctor will proceed for further treatment.

Baye's theorem is used to find the possible occurrence of the disease which makes use of the previous records to make decisions. It analyses the data in particular region within particular duration to make decisions. Doctors makes use of this algorithm to come to possible results.

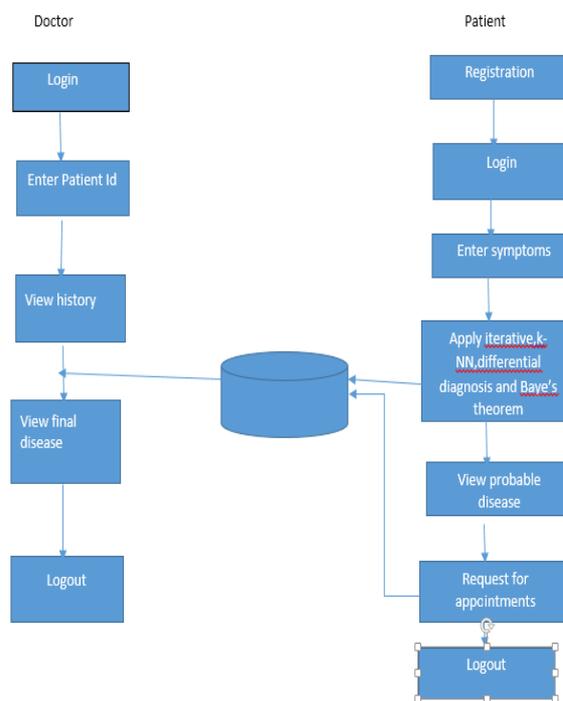


Fig: System Architecture

V. CONCLUSION

Thus we can conclude that by using proposed methodology we can predict the probable disease within minimum time. Here we are maintaining the E-copies of patients so that they will be useful while the patient is being relocated. We provide unique ID to the patients as well as doctors. So the duplication of records will not take place. This can be done using mentioned classification techniques hence, it will be easy to find the probable diseases. The approaches can be used to supplement and improve existing systems that provide differential diagnosis. It will also help the medical fraternity in the long run by helping them in getting accurate diagnosis and sharing of medical practices which will facilitate faster research and save many lives.

**REFERENCES**

1. R. Carvalho, R. Isola, and A. Tripathy, MediQueryAn automated decision support system, in Proc.24th Int. Symp. Comput.-Based Med. Syst., Jun. 2730, 2011, pp.16.
2. R. A. Miller, Medical diagnostic decision support systemsPast, present, and futureA threaded bibliography and brief commentary, J. Amer. Med. Inf. Assoc., vol. 1, pp. 827, 1994.
3. W. Siegenthaler, Differential Diagnosis in Internal Medicine: From Symptom to Diagnosis. New York: Thieme Medical Publishers, 2011
4. H. R. Warner and O. Bouhaddou, "Innovation review: Iliad—A medical diagnostic support program," Top Health Inf. Manage., vol. 14, no. 4, pp. 51–58, 1994.
5. R. Rojas, Neural networks: A Systematic Introduction. Berlin,Germany: Springer- Verlag, 1996, pp.337370.
6. J. Han and M. Kamber, Data Mining Concepts and Techniques. San Mateo, CA: Morgan Kaufmann, 2011
7. Taking Our Time: Chronic Illness and Time-Based Objects in Families by AndreaBarbarin, Tiffany c.veinot ,PredragKlasanja, March 14, ACM
8. Rahul Isola, Student Member, IEEE, RebeckCarvalho, Student Member, IEEE, and Amiya Kumar Tripathy, Member, IEEE Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and k-NN, IEEE transactions on information technology in biomedicine, vol. 16, no. 6, november 2012
9. Website: www.webhealthcentre.com
10. Website: www.practo.com