A Pair-Pattern Vector Space Model of Semantics based Prototype Model for Spreading Medical Awareness in Rural Areas

Sukanya Ray
Student, M.Tech(CSE)
Amity School Of Engineering
& Technology, Amity University,
Noida (U.P.), India

Nidhi Chandra
Asst. Professor (CSE)
Amity School Of Engineering
& Technology, Amity University,
Noida (U.P.), India

ABSTRACT

In rural areas there is always a lack of medical awareness. People are not aware of many diseases and even if they know the name of the disease or even the symptoms of the disease they are not aware about the prevention methods of the disease. In this paper a prototype model is designed for the social workers or leaders to spread awareness of prevention of various diseases and also the pre-cautions that must be taken so that those diseases do not occur. This prototype model will help social workers to know the reason of the spread or occurrences of various diseases, and what measures can be taken to prevent these diseases from occurring. This will help in spreading a general awareness about various diseases in rural areas where there is no proper medical treatment centre and preventing diseases from occurring is the best solution in those areas.

Keywords

Pair-Pattern, Vector Space Model, Medical Awareness

1. INTRODUCTION

Various surveys are showing that various fatal diseases occur in the rural areas as the people are not aware of such diseases and do not have any idea of how to prevent those diseases. Moreover in the rural areas people are generally are economically backward and they do not have proper hygienic environment in their place of living. So every year many people lose their life in such areas due to lack of general awareness. If general awareness about various diseases and sense of hygiene can be spread among these people, then this kind of mishap can be prevented from happening.

Moreover there are no proper medical centers in rural areas. In some areas each village has got a medical center and at some places two villages have only one medical centre. The doctors or medical staffs present in those medical centers are not well trained and inexperienced. [1] Often they lack the knowledge of what should be done to treat the patients suffering from any disease and they also almost have no idea of what should be done to prevent various diseases from occurring.

For developing countries this is a major area of concern and any negligence here is not acceptable as World Health Organization (WHO) have said in the year 2000, "Health for all". [2] So meet these standard proper care must be taken especially in rural areas. As this problem is mainly due to lack of general awareness about various diseases so if general awareness can be spread among the people and also among the doctors and medical staffs present in the different health centers then these mishap can be partially solved.

Now a day almost all hospitals, nursing homes and medical centers keep a record of the various diseases, their cause of occurrence, treatment provided for it and preventive steps that must be taken to prevent it from occurring. Social workers from various NGOs who have set up their offices in various villages or people from the municipal authority of the villages can take help from these medical records of the hospitals, nursing homes to spread general awareness about various diseases.

Various diseases like malaria, dengue, jaundice, bird flu, etc occurs every year due to lack of knowledge of prevention of these diseases and take away many lives. The main aim of this paper is to spread general awareness among the common people in rural areas with the help of the data and various case studies available in various hospitals and nursing homes, so that they can prevent the life loss to various fatal diseases.

This paper is divided into 4 sections. Second section focuses on the Current Medical Scenario. Third section focus on the proposed method and fourth section is for further scope of research in this domain.

2. CURRENT SCENARIO

Polio usually occurs in people who are not vaccinated at an early age. It is still at large in many African and Asian countries especially in the rural areas where people are not well aware of this disease. Polio can cause up to 79% paralysis if it affects the spinal cord and can also be very fatal in some cases. [3]

Malaria is treated as the most fatal disease in the world that is at large in about 106 countries. According to World Health Organization's (WHO) 2011 report malaria is very much at large in almost 106 countries around the world and accounts for approximately 2.23% of the death worldwide. [4] Reports say that 225 million cases of malaria were reported in 2009 and as many as 781,000 people died in 2009 due to this disease and in 2010 about 655,000 people lost their life to this fatal disease. 90% of the death in sub Saharan African countries is due to malaria and 60% of that 90% of the death are accounted by children of age five and below. This means now also people there are not aware about the cause of this disease, even though it is causing havoc every year. Malaria

mostly breaks out in areas having poverty and in rural areas. [5]

Dengue which is treated as only second to Malaria has a mortality rate of 1%-5% without treatment and even less than 1% with adequate treatment. Dengue is at large in about 110 countries and every year approximately 50 million to 100 million people are hospitalized with this disease. In cities where people get proper treatment there also almost 25,000 people lose their life every year while in rural areas the number of deaths increases by leaps and bounds. World Health Organization (WHO) says Dengue is one of the sixteen tropical diseases that people are less aware of and this has lead to the growth of dengue by 30 folds between the year 1960 and 2010. 70% of the total of approximately 2.5 billion people residing in the Asian and Pacific region around the equator suffers from this disease. [6]

Almost two to three million people worldwide are said to be permanently disabled due to leprosy. India has the highest number of leprosy patients followed by Brazil and Myanmar. In 1999 640,000 people around the world were affected by this disease and by 2000 it increased to almost 738,284 people. World Health Organization (WHO) in 2000 has listed 91 countries where leprosy is endemic. India, Burma and Nepal have 70% of those cases and India alone has 50% of those 70% cases. In 2002, 763,917 cases were detected which gives an indication that people are not aware of the cause of the disease and every year it is increasing in leaps and bounds. USA were the first to have diagnosed and decreased the rate of spreading of this disease but in 2005 near about 166 new cases of leprosy were registered in the United States of America. Though the economically strong countries have managed to decrease the rate of leprosy but high prevalence of this disease is found in developing countries and mostly in the rural areas. [7]

Kala-azar is found to be endemic in 88 countries around the world. It affects as many as 12 million people worldwide and every year 1.5 million to 2 million new cases of kala-azar are registered and out of which 60,000 people die each year. Almost 90% of the total cases of kala-azar are in India, Bangladesh, Nepal, Sudan and Brazil. World Health Organization (WHO) reports that most of this cases are in rural areas, poverty stricken areas where the sense of hygiene is very less. Kala-azar has been identified as the disease of the developing countries. [8]

The reports of the year 2010 suggests that cholera affects approximately 3 million to 5 million people worldwide and causes 100,000 to 130,000 deaths each year. Cholera remains as both epidemic and endemic in many countries around the world and is identified as the disease of the developing countries. People are still not aware of what is the cause of cholera as it outbreaks for various reasons. But the main reason is lack of proper drinking water and proper sense of hygiene and that is the main reason why cholera is at large in areas of low poverty and rural areas and in developing countries. [9]

Trachoma is another infectious disease which is causing havoc worldwide. Nearly 41 million people suffer from this disease and out of that about 8 million people get visually challenged each year due to this disease. It is a major cause of infectious blindness around the world and is at large. This disease mainly occurs in poverty stricken areas and gradually becomes hereditary. [10]

3. EXISTING METHODOLOGY

Adjuvant! Online: Adjuvant! Online [11] is a web based tool that is developed to provide medical assistance by guiding users (doctors, medical staffs or even patients) to detect cancer at an early stage and discuss its risks and what benefits can be achieved in form of chemotherapy, hormone therapy or both the therapies after the surgery. The tool provides the estimates of reduction of risk after applying the adjuvant therapy. The predictions are given on printed sheets in simple textual or graphical form for consultation.

Cancer- Math.net: Cancer- Math.net [12] is another web based tool that helps doctors, medical staffs for predicting the clinical outcome of individual cancer patients with the help of web- based calculators. It also predicts the outcome of the treatment each patient will receive. The main drawback of this web based prediction tool is that it provides no flexibility in storing, retrieving and editing of patient's data in a secured environment.

Evidence Based Medicine (EBM): The data available here is the publicly available data on internet. The prediction is done by studying various data sets available. Here two algorithms are used and result achieved in each is compared with the other for reaching the conclusion. The main drawback of this approach is that it requires both internal clinical experience as well as external clinical experience. [13]

Predicting Breast Cancer Model: In this model the researchers used SEER public database to study [14] the survivability of breast cancer patients between 1973 and 2000. The patients were grouped into two categories (survival and non survival) based on the values present in the SEER database. The drawback of this model is that the database did not have survival rate of each case and also the cause of the death.

Predicting Breast Cancer Survivability: In this model the above work is further extended and the survivability rate and cause of death in each case is added in the database and a formula has been developed to calculate it. The patients were again grouped into two categories (survival and non survival) and the survival category contained people who will survive for more than 5 years or 60 months. The drawback of this approach is that this model does not consider the variables of Estrogen and Progesterone hormone therapy. [15]

4. PROPOSED METHODOLOGY

Proposed methodology is motivated from the crisis of such system in the present scenario where technology in medical domain has made quite advancement. The proposed

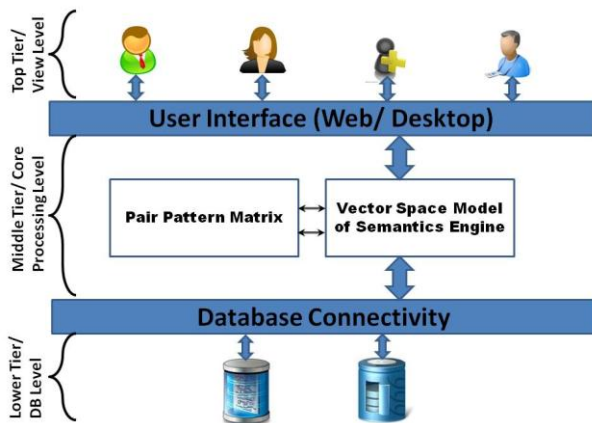methodology concept can be understood by the architecture shown in Fig 1.

Figure 1: 3-Tier Architecture of proposed system

The proposed architecture is a 3-tier architecture where the top tier of the architecture is user or view level, middle tier is the Vector Space Model of semantics engine and the lower tier is data set or corpus which will gather data from different medical domain to give accurate and efficient results. User can use this software either from the desktop or laptop like any other system software or by using an internet connection like an web application; the basic architecture is both the cases will remain the same as the logic behind will remain the same. As discussed in the introduction part the proposed architecture will help social workers, doctor or medical staffs, people from Municipal Corporation, and also the common people to spread awareness about various diseases. To understand this architecture flow let us take an example: Suppose malaria is spreading at large in a nearby village and the people of this village wants to take precaution from that but they have no idea of what the disease is or what are the exact causes of this disease, only thing they know is that this is a fatal disease. This only leads to panicky situation. In that case the concerned person for handling this or even the common people can know about this simply by entering words like "Give all reasons that causes Malaria". They will get the information of the causes of Malaria and will help in spreading a general awareness among people.

According to the prototype model in the first phase the people concerned will need to enter the keywords like "give all reasons that causes Malaria" in the system which can be either in desktop or laptop software or in web based application. There will be more options also on entering keywords like "how fatal is Malaria" will give about the fatality of the disease. On entering keywords like "cure for Malaria" will give the basic treatment that must be given to the patient so that he/ she can be moved out of danger immediately. More specific the relations between the keyword and disease name will give more accurate result.

In the second phase the keywords entered will go to the Vector Space Model of semantics engine in the middle tier.

The Vector Space Model of semantics engine is equipped with pair pattern matrix which will quickly find the similarity among the relationship for the relationship like causes Malaria or fatality Malaria or cure Malaria. The Vector Space Model of semantics will apply the pair-pattern matrix and then will return the result based on the similarity of the relationship. The pair-pattern matrix will not wait for the complete execution till the end of the dataset, but as soon as it gets a result it will start returning back the result and will keep on appending more results as it will be getting returned.

The database will have dataset of all causes, fatality and cure of various diseases. The dataset must be well prepared and contain large amount of data as more relevant data will give more accurate result. Suppose some disease can be kept away by the intake of certain vitamins. If only the name of the vitamin is written it can be a problem. Instead if the dataset is big enough then it will have definitely a list of different food items that give that vitamin and the range of the food items have to be such that it is easily available to people in rural areas as well as urban areas, in Asian countries as well in European and African countries. Though this prototype is mainly designed keeping the people of rural areas in mind but it is better if it is widely used and even by people from urban areas. Hence the pair-pattern matrix used in phase II will work on the data set and will return back accurate results to the user who can find a solution accordingly.

With the help of this prototype model doctors or medical staffs or social workers or even common people can be aware about various diseases which can be fatal if not prevented at early stage. For this prevention general awareness need to be spread and also if queries are for relation like cure of X disease then results for all the treatment done to that disease and their outcome will be shown and immediate medication can be given to patients even by not so experienced doctors in the rural areas before the patient can be moved to any big hospital or nursing home in the city. The same prototype model can actually serve three purposes, create awareness about various diseases, detect fatality of diseases and also provide remedies or cure to various diseases. A comparison of the Proposed Model with the previously discussed Existing Models is shown in Table 1 below:

Table 1: Comparison of Proposed Approach with the discussed Existing Approaches

Model Name	Pros	Cons
Existing Model	Prediction tools that can predict the fatality of the disease and also the effect of the treatment on that disease.	No prediction for the prevention of the disease from occurring.
Proposed Model	This model can guide a user how	A lot depends on the data set that

	to take precaution from a disease from occurring and also provides guidelines about the immediate medical treatment that has to be given if the disease has occurred to someone. This model can also tell about the fatality rate of the disease.	works in the lower level of the 3-tier architecture.

In Table 1, the proposed model is being compared with the existing models. One main advantage of this proposed model is that it can guide the user to take pre cautionary steps against many fatal diseases. The main aim of this paper was to spread medical awareness among the people in rural areas and in developing countries and by providing the guidelines of how one can prevent various diseases from occurring, this proposed model can work as a very beneficial model in the rural areas.

5. CONCLUSION

This paper focuses on the medical awareness needed in the rural areas and in developing countries so that common people can be cautious and prevent the spread of various fatal diseases. This paper further proposes a prototype model from which social workers, doctors or medical staffs, municipal leaders as well as common people can know about the cause of occurrence of various diseases and prevent them from occurring thus decreasing the death ratio. Proposed method consider pair-pattern matrix of vector space model of semantics and thus give very fast and accurate result.

6. REFERENCES

[1] M. Refaee, Prof. M. A. Fayed "HEALTH CARE TECHNOLOGY FOR DEVELOPING COUNTRIES" IEEE 2000

[2] Prof. J. K. Choudhury, "Technology and Status of Health Care Delivery in Third World With Special Reference to India", Proc. 14th Annual Int. Conf. of the IEEE Eng. in Medicine & Biology Society, Paris, pp2869-2870, 1992

[3] U.S. National Library of Medicine - The World's Largest Medical Library. [Online accessed 10-March-2012]. [Online]Available: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0002375/

[4] World Health Organization, "Malaria report". [Online accessed 10-March-2012]. [Online]Available: http://www.who.int/malaria/world_malaria_report_2011/en/index.html

[5] Wikipedia, "Malaria — wikipedia, the free encyclopedia," [Online; accessed 10-March-2012]. [Online]Available: http://en.wikipedia.org/wiki/Malaria

[6] Wikipedia, "Dengue Fever — wikipedia, the free encyclopedia," [Online; accessed 10-March-2012]. [Online]Available: http://en.wikipedia.org/wiki/Dengue_fever

[7] Wikipedia, "Leprosy — wikipedia, the free encyclopedia," [Online; accessed 10-March-2012]. [Online]Available: http://en.wikipedia.org/wiki/Leprosy

[8] Wikipedia, "Leishmania — wikipedia, the free encyclopedia," [Online; accessed 10-March-2012]. [Online]Available: http://en.wikipedia.org/wiki/Leishmaniasis

[9] Wikipedia, "Cholera — wikipedia, the free encyclopedia," [Online; accessed 10-March-2012]. [Online]Available: http://en.wikipedia.org/wiki/Cholera

[10] Wikipedia, "Trachoma — wikipedia, the free encyclopedia," [Online; accessed 10-March-2012]. [Online]Available: http://en.wikipedia.org/wiki/Trachoma

[11] Adjuvant! Online, Decision making tools for health care professionals. www.adjuvantonline.com.

[12] Cancer Math. Web based decision support system for oncologist. www.cancermath.net

[13] Ir CATH Tee, Ali H. Gazala "A Novel Breast Cancer Prediction System" IEEE 2011 978-1-61284-922-5/11.

[14] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Public-Use Data (1973-2005).

[15] Abdelghani Bellaachia, Erhan Guven. Predicting Breast Cancer Survivability Using Data Mining Techniques.