RESEARCH ARTICLE

K-NN CLASSIFIER FOR PREDICTION OF SWINE INFLUENZA VIRUS

Ms. Ankita R. Borkar[1], Dr. Prashant R. Deshmukh[2]

[1]Student, Computer Science and Engineering, Sipna C.O.E.T., Amravati, Maharashtra, India
[2]Professor, Computer Science and Engineering, Sipna C.O.E.T., Amravati, Maharashtra, India
[1] borkarankita972@gmail.com; [2] pr_deshmukh@yahoo.com

Abstract— Swine influenza disease causing mortality in modern society. In health care commercial enterprise contain some healthcare data which, unfortunately, are not "mined" to discover hidden information. The main motive of Data mining is to predicting diseases. Sometime the database report of patient is not more efficient, recently we are made an Endeavour to find the most globally spreading influenza virus in all over the world which we call Swine flu . Swine flu disease contain lots of tests which must be requisite from the patient for detecting a disease. So the solution for such situation is Advanced data mining techniques. This paper summarizes about a prototype using data mining techniques, namely k-nearest neighbored Algorithm. Which helps us in finding accurate results in many fields. The main goal of this paper is find out patient suffering from which type of swine flu virus by using k-NN algorithm

Keywords— Data mining, KDD process, Swine flu , K-nn-algorithm, Euclidean distance

I. INTRODUCTION

Swine flu is one of the most infectious and viral disease which contains different types of virus that causes thousands of deaths per year. Hence we have using a method of data mining to find various virus in this disease using KNN algorithm By which we can reduced the laboratory test cost as well as time too. "How can we turn important data into information that can enable healthcare practitioners to make an intelligent clinical decisions?" This is the main purpose for this paper

A. Swine flu

Swine flu is a respiratory disease caused through influenza viruses that infect the respiratory tract of pigs and result in a barking cough, decreased appetite, nasal secretions, and listless behavior; the virus can be transmitted to humans. Swine flu viruses are easily transmissible among humans beings. Swine influenza is also called pig influenza or swine flu, hog flu and pig flu, is an infection caused by any one of several types of swine influenza viruses. Swine influenza virus (SIV) or swine-origin influenza virus (S-OIV) is any strain of the influenza family of viruses that is endemic in pigs. In 2009, the known influenza C include in strain and there are some subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3.By researching worldwide, from pig populations the Swine influenza virus comes .The virus get transmit from pigs to humans is not common and does not always tend toward human flu, often resulting only in the productivity of antibodies in the blood of human. If transmission does cause human flu, it is known as zoonotic swine flu. People with regular come in contact to pigs are mostly lead to swine flu infection. Since then, only 50 such transmissions have been confirmed. Swine flu is transmitted from person to person by inhalation of droplets existing virus from people sneezing or coughing; it is not transmitted by eating cooked pork

products. The newest swine flu virus that has caused swine flu is influenza A H3N2v. Some common symptoms of Swine flu are caugh, fever, sore throat, runny nose, headache, chill, fatigue, nausea etc.

i. Statistics of Swine flu

In 2009 swine flu was an influenza A virus type designated as H1N1. In 2011, a new swine flu virus was detected with new strain was named influenza A (H3N2) Currently, more than 1500 people in india kill by swine flu virus and it will rapidly exploring all over the world so recently there are not notice large numbers of people affected with H3N2v.virus H3N2 caused flu, but this strain is different from H3N2v. In general, all of the influenza A viruses have a structure similar to the H1N1 virus; each type has a somewhat different H and/or N structure.

ii. Structure of swine flu

The influenza virion is roughly spherical and enveloped virus; the outer layer virus is multiple a lipid membrane which is taken from the host cell. Inserted into the lipid membrane are "spikes", which are proteins—actually glycoproteins, because they consist of protein linked to sugars name as HA (hemagglutinin) and NA (neuraminidase). These are the proteins that determine the subtype of influenza virus (A/H1N1, for example). The HA and NA are very important in the immune response against the virus; antibodies (proteins made to combat infection) may protect against infection against these spikes . The NA protein is the target of the antiviral drugs Relenza and Tamiflu. Also embedded in the lipid membrane is the M2 protein, which is the target of amantadine and rimantadine.[1,2].

II. LITERATURE REVIEW

In year 2012 Mai Shouman, Tim Turner, and Rob Stocker et. al. performed a work "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients". In this paper the author details work that applied KNN on a Cleveland Heart Disease dataset to investigate its efficiency in the prediction of heart disease. The author also investigated if the accuracy could be enhanced by integrating voting with KNN.[3]

Thakkar , Hasan and Desai were inspired to do this research by the study of mortality rate of swine flu. The paper focus on the aspect of medical diagnosis by learning predictions through the data collected for Swine flu. The authors proposed a method to identify swine flu by study 110 symptoms to decrease the cost incurred in the test of the disease .The author developed a prototype intelligent swine flu prediction software , they used Naïve Bayes classifier technique for classifying the patients of swine flu where we are using KNN algorithm . Based on the possibility of the diseases and guaranteed the accuracy of almost 63.3 percent.[4,5,6].

III. DATA MINING

Data Mining came into existence in the middle of 1990's which consider as a powerful tool that fetching previously unknown pattern and useful information from huge database. Data Mining techniques help the data holder to analyze and discover unsuspected relationship among their data which in turn helpful for making decision. In general, It also allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among different of fields in large relational databases. The KDD process is an interdisciplinary subfield of computer science, in which discovering patterns in large data sets involving methods of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.[7,8]
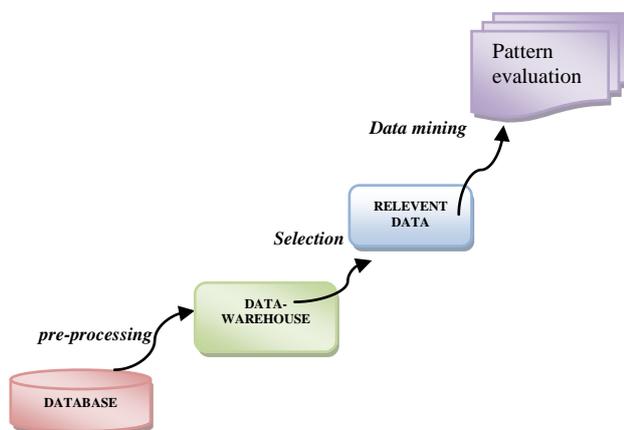

Fig.1:-Working of Data Mining

The above Fig.1 shows the working of Data Mining technique stages that initially we integrate data from database then we select our targeting attributes from data ware house then take relevant data use data mining techniques apply various algorithms , finally we get useful knowledge and using this approach we take action.[9,10].

IV. K-NEAREST NEIGHBOR

K-NN algorithm is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. It is called lazy because it does not contain any training phase or minimal training phase. All the training data is needed at the time of testing phase and it uses all the training data .The *k*-nearest neighbor algorithm is one of the simplest machine learning algorithms. The K-nearest neighbor classification, the output is class membership of an object is classified by majority vote of its neighbors, with the object being assigned to common among the k-nearest neighbor where k is an positive integer typically small. It is simple algorithm that stores all available causes and cases based on similarity measures .A case classified by a majority votes basis of its neighbor with case being assigned to most common amongst its knn measured by distance function.
Now see how knn works .
K-nearest neighbor has been used in statistical estimation and pattern recognition as consider an example –suppose we have to find how many patient are showing positive test report of having H1N1 and H3N2 virus in a particular state of India . For that first we use some method for Distance measures for continuous variable of distance function are as shown below.[11,12]

Distance function : -

- Euclidean $= \sqrt{\sum_{i=1}^{k} (xi - yi)^2}$

- Manhattan $= \sum_{i=1}^{k} | xi - yi |$

- Minkowski $= \left(\sum_{i=1}^{k} (| xi - yi |)^q \right)^{1/q}$

In knn algorithm there are some formulas are used to find the distance function which include Euclidean distance, Manhattan distance, Minkowski distance
But from all these distance we are using the most commonly distance called Euclidean distance.

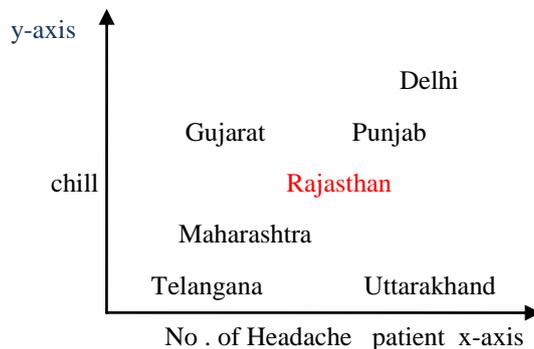$$\text{Euclidean} = \sqrt{(x1 - y1)^2 + (x2 - y2)^2}$$

Fig. 2 :- patients record structure using symptoms

The above Fig.2 shows the data concerning all the affected patient belongs from the various different state of India . And the symptoms chill and headache are two numerical variable or predictor. Here we find k closest training points to x-axis new with respect to Euclidean distance. Classify by   y-axis knn is equal to majority of vote among the   k-points .

TABLE NO. I

Diff. state in india affected by swine flu virus	No. of Patients having symptom chill	No.of Patient having symptom Headache	Virus type
Gujarat	3	104	H1N1
Punjab	2	100	H1N1
Maharashtra	1	81	H1N1
Delhi	101	10	H3N2
Telangana	99	5	H3N2
Uttarakhand	98	2	H3N2
Rajasthan	18	90	?

 Above Table No.I contained the database report of swine flu virus affected   patient belong from different states.

So here we are applying Euclidean distance to our data base report of various state which are showing some basic symptoms of swine flu disease through which we are finding whether the patients from Rajasthan has been suffered by virus H1N1 or H2N3 by Euclidean formula-

Euclidean distance of Rajasthan to Gujarat Patients ,

$$D = \sqrt{(x1 - y1)^2 + (x2 - y2)^2}$$
$$= \sqrt{(18 - 3)^2 + (90 - 104)^2}$$
$$= \sqrt{(15)^2 + (14)^2}$$
$$= \sqrt{421}$$
$$D = 20.5$$

Euclidean distance of Rajasthan to Punjab Patients,

$$D = \sqrt{(x1 - y1)^2 + (x2 - y2)^2}$$

$$= \sqrt{(18 - 2)^2 + (90 - 100)^2}$$

$$= \sqrt{(16)^2 + (-10)^2}$$

$$= \sqrt{356}$$

D = 18.7

Like this way we compute all distance of state from Rajasthan we get following result -

TABLE NO. II

Diff. State in india affected by swine flu virus	Euclidean distance
Gujarat	20.5
Punjab	18.7
Maharashtra	19.2
Delhi	115.3
Telangana	117.4
Uttarakhand	118.9

The above Table No.II shows result of Euclidean distance of all state from the Rajasthan so here we can clearly see that nearest near neighbors means minimum distance of state from Rajasthan we can there are three values of the Euclidean distance are closely near are 20.5,18.5,19.2 from this we take minimum distance 18.5 which belong from Punjab state so here we see that from which virus type the patient in Punjab affected so we see it in ours database report and we found that they were suffered from H1N1 virus type so finally we predict that patient showing symptoms of swine flu who belong form Rajasthan has been affected by H1N1 virus type of the swine influenza by using the K-nn algorithm.[13,14].

V. CONCLUSIONS

The Data mining technique, can be used in collaboration With a K-NN classifiers algorithm which used in diagnosing Swine flu disease virus affected across in different state in India This proposed approach showed some Promising results which may lead to further attempts to utilize information technology for diagnosing virus from which patients is suffering from Swine flu . Here we used K-NN Classification rules which are easy to interpret. In future, we will try to get more the accuracy results for the swine flu disease which helps to find different parameters using different data mining techniques as per suggested by doctors.

REFERENCES

[1]   The Swine influenza [Online].Available:-   http://en.wikipedia.org/wiki/Swine_influenza
[2]   The Swine influenza [Online] .Available :-http://www.medicinenet.com/swine_flu/article.htm
[3]   Mai Shouman, Tim Turner, and Rob Stocker, June 2012 applying k-Nearest Neighbour in Diagnosing Heart disease Patients, International Journal of Information   and Education Technology, Vol. 2, No. 3.

[4] Thakkar. Hasan and Desai, 2010,"Health care decision support system for swine flu prediction using Naïve bayes classifier", International Conference on Advances in Recent Technologies in Communication and Computing.,IEEE.

[5] WorldHealthOrganization[Online].Available:-www.tbevidence.org/documents/dxres/models/tb_diagnostics.pdf   World Health Organization

[6] Swine Flu paper [online] Available http://www.ijarcsse.com/docs/papers/Volume_4/6_June2014/V4I6- 0235.pdf

[7] Data mining [Online] .Available:- http://en.m.wikipedia.org/wiki/Data_mining.

[8] D. Hand, H. Mannila and P. Smyth, 2001, "Principles of data mining", MIT.

[9] Kaur, H., Wasan, S. K., 2006 "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200.

[10] J. Han and M. Kamber, 2006"Data mining: concepts and techniques", 2nd ed. The Morgan Kaufmann Series.

[11] K-nearest neighbor classifire [Online].Available: http://en.wikipedia.org/wiki/K-NN_classifier

[12]   Machine Learning for Medical Diagnosis: History, State of the Art and Perspective by Igor Kononenko .

[13] D. Y. Liu, H. L. Chen, B. Yang, X. E. Lv, N. L. Li and J. Liu, 2012"Design of an Enhanced Fuzzy k-nearest Neighbor Classifier Based Computer Aided Diagnostic System for Thyroid Disease", Journal of Medical System, Springer .

[14] M. Shouman, T. Turner and R. Stocker, 2012 "Applying K-Nearest Neighbour in Diagnosing Heart Disease Patients", International Conference on Knowledge Discovery (ICKD-2012).