

A Prediction Engine for Influenza Pandemic using Healthcare Analysis

Mr. Shubham Shende¹ Mrs. S. R. Hiray²

¹Student ²Professor

^{1,2}Department of Computer Engineering

^{1,2}SCOE Vadgaon, Pune, Maharashtra, India

Abstract— Influenza is one of the frequent viral infections in the world, and puts majority of people at risk every year. Recent systems for detecting influenza are not adequate due to the lack of efficient consultant and deficiency in recognizing the nature of serious medical conditions. Such systems are need to be developed for a quick and proficient detection system. The present technology has given rise interest in the field of influenza awareness, especially on the topic of prediction. The current technology can give reply only to the straightforward queries such as “Determine the common age of patients who have caused by flu?”, “How many patients have to continue the hospital stays extended than 10 days”. But still, present system are not able to respond to the complicated queries like predict the span of hospital stay before any surgical operation. The proposed prediction engine collects all related parameter & analysis is performed on collected data by giving a Viterbi prediction model for the flu forecasting of the patients resting on their current health condition.

Key words: Influenza Awareness, Healthcare Analysis, Viterbi Algorithm, Hemoglobin Concentration, Platelets

I. INTRODUCTION

In Asia countries, the infection increases in the monsoons. This has arisen many times in recent years and it just as constant trend. Due to Absence of professional doctors and incorrect diagnosed cases there is need to build up a speedy and efficient disease forecasting scheme. The system takes input parameters of the patient. We have analyzed old data in order to learn, train and test the model. Representing the flu prediction system. We also know that privacy is important factor. None of the queries in our project’s database can be associated with a particular individual. Our project’s database retains no information about the identity, IP address, or specific physical location of any user.

Influenza is an infectious disease which have an effect on the people in seasonal parts that would be a grave difficulty like bacterial pneumonia which will affect the aged or patients having an impaired immune system. Influenza can be serious and noted that normally 20,000 people lose their life due to influenza in the U.S. every year, and many of them admitted in the hospital as a effect of influenza.

Because early revealing of influenza could help to save the number of lives, great accent has to be take place on influenza inspection by monitoring indicators which are responsible for spreading of contagious disease [6].The plagues in 20th century caused tens of millions of people to loss their lives. Over and over again, these new virus species result from the spread of the existing flu virus in humans from other animals. At the first time death of humans in Asia in the 1990s, occurs due to deadly virus named H5N1, pretended as great threat for a new influenza disease.

II. LITERATURE REVIEW

The present actions and future prospects associated with big data for health is explained in [1].

A novel deep learning scheme used to guess the diseases by giving the queries of health seekers is proposed by L. Nie et al.[2]. The two major parts of the system are distinguishing medical signatures attribute from raw features and next estimates the raw attributes and their signatures like input nodes in one layer and unknown nodes in the consequent layer, respectively, and gives the mutual relation among two layers by pre-training with pre-labeled data. In general, it gives precise tasks with enhancement. These experiments are performed on real-world dataset marked by online doctors. A stochastic prediction model [3] to forecast the upcoming health situation of the most associated patients depending on their current health condition is giving 98% accuracy.

A hybrid K-means and Support Vector Machine algorithm [4] is also used for disease prediction. The hybrid K-means algorithm gives dimensionality reduction of the dataset which is an input to Support Vector Machine classifier. The simulation is performed in MATLAB and the results are more accurate in detection of the disease.

The customized forecasting models uses convolution neural network (CNN) [5] for disease risk calculation with structured and unstructured data. The real-time hospital data gathered from central China in 2013-2015.

Multilayer perceptron algorithm in Weka [7] is used to discover the similarity in pattern between the trained data set which includes health attribute of patients with heart problems and test data which contains the health data of the application users which gives prediction of the heart problems while the patient is movable.

III. PROPOSED SYSTEM

The system takes input parameters of the patient. We have analyzed previous data to learn, train and verify the model representing the flu prediction system. The system gets Get historical data of patient form dataset. After studying on the influenza affected patient’s data and then decides which factors are common in all patient.

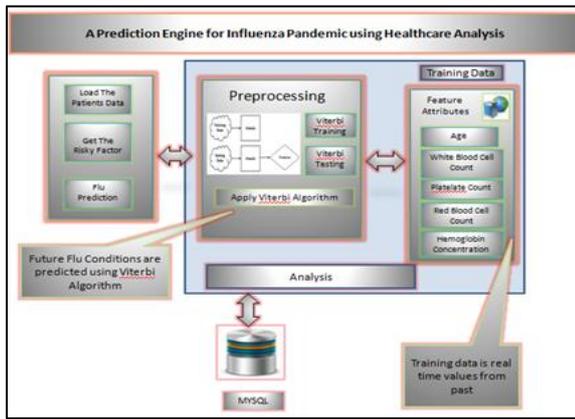


Fig. 1: Architecture Diagram

Data set contains 1000 of records from which we are using 8 attributes such as Age, Sex, White Blood Cell, CountRed Blood Cell Count, Plate late Count, Mean Plate late Count, Haemoglobin Concentration, Plate late Distribution Width. The results of our experimental analysis in finding significant patterns for Flu prediction parameters .Age wise and Gender Wise disease Analysis can be done to display how many patients have Flu, which patient age in between 10 -20 or 20 -30. Some of the diseases have an effect equally women and men but some diseases effects only women at a higher frequency. Such that 18% of women go through migraine headaches weigh against 6% of men in the U.S. thus, the advice regarding preventative measure can be given to the particular gender people.

Various diseases possibly will spread very quickly and the whole region gets affected by that disease. So these statistics would help to take precautions for the people leaving in a particular region to avoid the spread of that disease and reduce death ratio.

IV. ALGORITHM USED

A. Viterbi Algorithm

It is algorithms which discover the path throughout a trellis, i.e. shortest path. The trellis here is corresponding to a graph of a restricted set of status from a Finite States Machine (FSM). Every joint in the graph indicates a state and every edging is a possible changeover among two states at successive distinct time intervals.

Following is the algorithm description:

Initialization.

$t = 0;$

For all n, where $1 \leq n \leq N$

$\Gamma_{n0} = \ln \pi_n;$

$sp_{n0} = [x_n];$

End For;

Calculation.

For all t, where $1 \leq t \leq T,$

For all n, where $1 \leq n \leq N$

For all m, where $1 \leq m \leq N$

$\Gamma_{nt} = \text{Max} [\Gamma_{mt-1} + \ln a_{nm} + \ln b_n];$

End For;

$sp_{nt} = \text{Append}[x_n, sp_{mt}] \text{ such that } \Gamma_{mt-1} + \ln a_{nm} + \ln b_n = \Gamma_{nt};$

End For;

End For;

Decision:

If $t = T,$

For all n, where $1 \leq n \leq N$

$\Gamma T = \text{Max} [\Gamma_{nt}];$

End For;

$spT = \text{sp}_{nt} \text{ such that } \Gamma_{nt} = \Gamma T;$

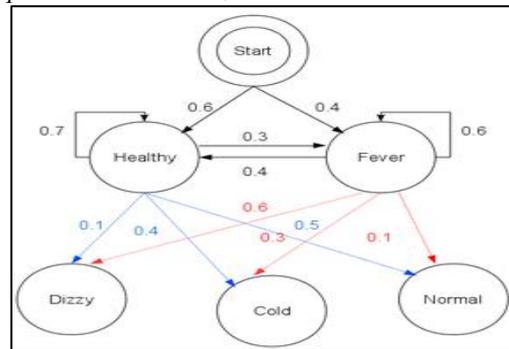


Fig. 2: Working Of Viterbi Algorithm [8]

In the VA looks at each state at time t, and for all the transitions that lead into that state, it decides which of them was the most likely to occur, i.e. the alteration through the greatest metric. If large number of an additional alteration is found and their metrics are similar, then one of the alterations is selected as the most possible transition. This highest metric is then allocated to the state's survivor path metric, Γ_{nt} . The VA then rejects the extra transitions into that state, and add on this state to the remaining path of the state at t - 1, from where the alteration begins. It turn out to be the survivor path of the state being observed at time t. The same operation is carried out on all the states at time t, at which point the VA moves onto the states at t + 1 and carries out the same operations on the states there. When we reach time t = T (the truncation length), the survivor trail is find out by the VA as earlier and it also has to decide which of the survivor paths is the most probable one. This is performed by resolving the survivor with the highest metric; and for the second time more than one survivor is the greatest, subsequently the most likely path is selected at random. The VA gives outcome to this survivor path, spT, through its survivor metric, ΓT .

V. EXPERIMENTAL RESULTS

The goal of our project is to predict flu and patient will be diagnosed relayed on the feature attributes, as Age ,Sex ,White Blood Cell Count ,Red Blood Cell Count , Plate late Count ,Mean Plate late Count Hemoglobin Concentration, Plate late i.e. the factors which are majorly responsible for flu. The system will give accurate results whether the patient will be having flu or not.

The predefined dataset which keeps historical records of a patient for matching the factors for experimental analysis and flu prediction is used. Dataset of 480 records with 4 attributes are used like a Cough, Fever, Headache, Normal

The analysis on mentioned parameters give a prediction that patient infected with influenza or not.

By using a viterbi algorithm the accuracy in the prediction can be increased to the great extent by training the large dataset with minimum time and it is less complex. Below table gives the comparison between different algorithms tested on 479 records.

Algorithm Comparison				
Sr.No	Algorithms	Tested On	Correctly Classified	Incorrectly % Accuracy
1	SVM Classifier	479	340	139 0.7098121
2	Random Forest	479	375	104 0.782881
3	Viterbi (Our Approach)	479	410	69 0.8559499

Table 1: Algorithm Comparison

The system gather probabilistic data and analysis of this gathered data is performed. The viterbi prediction representation is intended to foresee the Flu prediction with accuracy upto 85%.

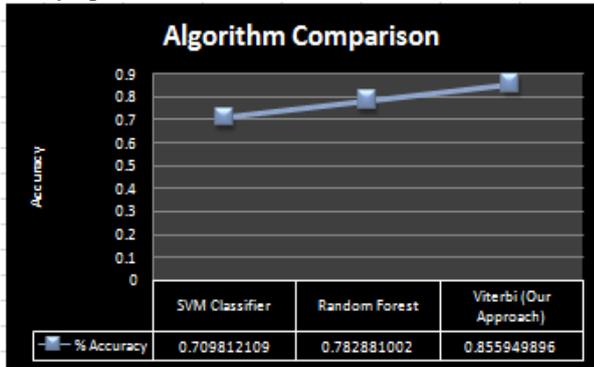


Fig. 3. Plot of Algorithm Comparison

VI. CONCLUSION

By analyzing the literature survey we come to know that the proposed system gives the accurate result compared with other methods. As we are using large dataset which will ensure the better performance compared as earlier.

Thus we build up a flu prediction engine which collects probabilistic data and analysis of those collected data gives the prediction of influenza with viterbi forecasting model this model is designed to predict the Flu by the most correlated patients depending on their current health status.

By taking the advantage of the extensibility of the system, use of other machine learning algorithm can increase the accuracy of the system in future. In addition working on other different feature attributes which make the system more efficient to give complete healthcare diagnosis system.

REFERENCES

- [1] J. Andreu-Perez, C. C. Y. Poon, R. D. Merrifield, S. T. C. Wong, and G. Z. Yang, "Big data for health," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1193–1208, July 2015.
- [2] L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T. S. Chua, "Disease inference from health-related questions via sparse deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2107–2119, Aug 2015.
- [3] I.S. Jacobs and C.P. Bean, "Fine particles, thin film and exchange anisotropy," in *Magnetism*, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [4] "Prasan Kumar Sahoo, Suwendu Kumar Mohapatra, Shih-Lin Wu", *Analyzing Healthcare Big Data with Prediction for Future Health Condition*, 2016 IEEE Transactions R. Nicole, Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.

- [5] Sandeep Kaur, Dr. Sheetal Kalra "Disease Prediction using Hybrid K-means and Support Vector Machine", 2016 IEEE.
- [6] MIN CHEN, YIXUE HAO, KAI HWANG, LU WANG1, AND LIN WANG, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities", *Digital Object Identifier 10.1109/ACCESS.2017.2694446*.
- [7] (2014, April) *The digital universe of opportunities: Richdata and the increasing value of the internet of things.* [Online]. Available: <https://www.emc.com/collateral/analyst-reports/idc-digital-Universe-united-states.pdf>.
- [8] Mr. Shubham Shende, Mrs. S.R Hiray, "A Prediction Engine for Influenza Pandemic using Healthcare Analysis", *International Research Journal of Engineering and Technology (IRJET)* e - ISSN: 2395 - 0056 Volume: 05 Issue: 03 | Mar – 2018.