

# A Prediction Engine for Influenza Pandemic using Healthcare Analysis

Mr. Shubham Shende<sup>1</sup>, Mrs. S.R Hiray<sup>2</sup>

<sup>1</sup>Student, Dept. of Computer Engineering, SCOE Vadgaon Pune, Maharashtra, India

<sup>2</sup>Professor, Dept. of Computer Engineering, SCOE Vadgaon Pune, Maharashtra, India

\*\*\*

**Abstract** - The shortage of specialists and high wrongly diagnosed cases has necessitated the need to develop a fast and efficient detection system. Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have caused by flu?", "How many patients have resulted in hospital stays longer than 10 days". However, they cannot answer complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Give region wise or disease wise or on both analysis report of the patients taking treatments on the same disease?", and "Given patient records, predict the probability of patients getting infected with the disease."

The proposed system is a Flu prediction engine. The system collects probabilistic data and analysis of those collected data is performed. Finally, a Viterbi prediction model is designed to foresee the Flu prediction by the most correlated patients based on their current health status. And also the location-wise, gender-wise, age-wise analysis of the data is performed.

**Key Words:** Viterbi Algorithm, Flu Prediction, Analysis, White Blood Cell Count, Red Blood Cell Count, Platelet Count.

## 1. INTRODUCTION

Flu synonym for influenza is a viral disease that spreads in world in seasonal parts. Three influenza plagues have faced in the 20th century and it caused in tens of millions of people to death, with each of these pandemics is caused by the arrival of a new type of the virus in humans. Over and over again, these new virus species result from the spread of the existing flu virus in humans from other animals. When it first caused the death of humans in Asia in the 1990s, a deadly virus named H5N1 pretended as great risk for a new influenza pandemic.

In Asia countries such as India, influenza spreads during the monsoons. This has happened a lot of time in recent years and it looks like a stable trend. But in last few years since 2012, influenza has spotted in the summer months and raising buzz about a viral mutation. Two of last summer bursts have been particularly unadorned: if the number of influenza patients recounted by India in 2015 was more than the pandemic happened in 2009, 2017 has been following close behind.

By the lack of specialist doctors and incorrectly diagnosed cases has necessitated that the need to develop a fast and effective disease prediction system.

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.

## 2. RELATED WORK

[1] Paper describes some of the existing activities and future opportunities related to big data for health, outlining some of the key underlying issues that need to be tackled.

[2] The paper clarifies the nascent field of big data analytics in healthcare, discusses the benefits, outlines an architectural framework and methodology, describes examples reported and briefly discusses the challenges, and offers conclusions on it.

[3] This article highlights the security requirements in BSN based modern health-care system. Subsequently, it proposes a secure IoT based healthcare system using BSN, called BSN-Care, which can efficiently accomplish the requirements.

[4] In propose system novel deep learning scheme is used to infer the possible diseases by giving the questions of health seekers. The system is comprised of two key components. First are globally mines the discriminate medical signatures from raw features. The second deems the raw features and their signatures as input nodes in one layer and hidden nodes in the subsequent layer, respectively, and it learns the inter-relations between these two layers via pre-training with pre-labeled data. Following that, the hidden protuberances serve as raw features for the more abstract signature mining. With incremental and alternative repeating of these two components, the scheme builds a sparsely connected deep architecture with three hidden layers. Overall, it well fits specific tasks with fine-tuning. These experiments are performed on real-world dataset labeled by online doctors.

The paper [5] gives a probabilistic data collection mechanism and correlation analysis of collected data is performed. A stochastic prediction model is designed to foresee the future health condition of the most correlated patients based on their current health status. Performance evaluation is done through extensive simulations in the cloud environment, which gives about 98% accuracy of prediction.

The paper [6] gives a simple measure which able to evaluate the trend similarity between physiological signals. It combines the Haar wavelet decomposition with the Karhunen-Loève transform, enabling to represent the signal by means of a reduced set of bases. The prediction process is based on reasoning process. By means of the similarity analysis procedure, a set of signals presenting a dynamics similar to the current condition is retrieved from the available historical data set. This set is then employed, through a nearest neighbor approach, in the prediction of the current condition.

[7] This paper proposes a hybrid K-means and Support Vector Machine algorithm for disease prediction. The hybrid K-means algorithm is helpful in choosing initial centroids, number of clusters and also to improve the efficiency of K-means algorithm. The hybrid K-means algorithm is used for dimensionality reduction of the dataset which is given as an input to Support Vector Machine classifier. The simulation is performed in MATLAB and the result are more accurate in prediction of the disease.

Paper [8] describes convolution neural network (CNN)-based multimodal disease risk prediction algorithm by using structured and unstructured data from hospital. Paper gives modified prediction models over real-life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, it gives latent factor model to reconstruct the missing data.

Padmashree T,Dr.N.K.Cauvery [9] gives prediction of the heart problems when the patient is mobile. The input to the Weka library is incorporated to the android application. Multilayer perceptron algorithm in Weka 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. It gives digital information of the patients stored centrally. This data can be used for designing an improved prediction of heart related disease.

### 3. PROPOSED SYSTEM

This system keeps historic records of a patient for matching the factors for experimental analysis and flu prediction. Firstly system gets the historical data of patient from Patients Dataset. After detail study on the influenza affected patients data and then decides which factors are common in all patient. We are using data set of 1000 records with 8 attributes. The parameter on which prediction of flu is depends are Age, Sex, White Blood Cell Count, Red Blood Cell Count, Platelet Count, Mean Platelet Count, Hemoglobin Concentration Platelet Distribution Width. and display how many patients have Flu, which patient age in between given condition.

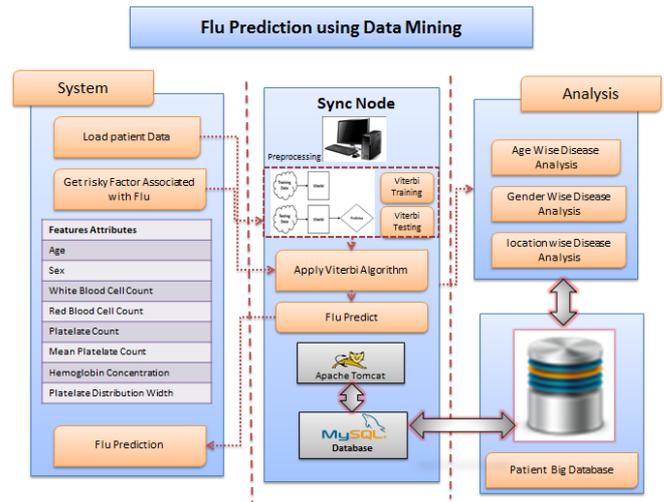


Fig.1 Architecture diagram Of Proposed System

Many diseases affect both women and men alike but some diseases occur in women at a higher frequency. For example, about 18% of women compared to 6% of men in the U.S. suffer migraine headaches. So, certain precautions can be suggested for the particular gender people. Some diseases may get spread very fast and the whole region gets influenced 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.

#### 3.1 Viterbi Algorithm:

We give the training set and the features as an input to the viterbi algorithm.

```

function VITERBI(O, S, Π, Y, A, B) : X
  for each state  $i \in \{1, 2, \dots, K\}$  do
     $T_1[i, 1] \leftarrow \pi_i B_{iy_1}$ 
     $T_2[i, 1] \leftarrow 0$ 
  end for
  for each observation  $i = 2, 3, \dots, T$  do
    for each state  $j \in \{1, 2, \dots, K\}$  do
       $T_1[j, i] \leftarrow \max_k (T_1[k, i-1] \cdot A_{kj} \cdot B_{jy_i})$ 
       $T_2[j, i] \leftarrow \arg \max_k (T_1[k, i-1] \cdot A_{kj} \cdot B_{jy_i})$ 
    end for
  end for
   $z_T \leftarrow \arg \max_k (T_1[k, T])$ 
   $X_T \leftarrow s_{z_T}$ 
  for  $i \leftarrow T, T-1, \dots, 2$  do
     $z_{i-1} \leftarrow T_2[z_i, i]$ 
     $X_{i-1} \leftarrow s_{z_{i-1}}$ 
  end for
  return X
end function

```

Suppose we are given a hidden Markov model (HMM) with state space  $S$ , initial probabilities of being in state  $i$  and transition probabilities  $a_{i,j}$  of transitioning from state  $i$  to state  $j$ . that produces the observations is given by the recurrence relations:

$$V_{1,k} = P(y_1 | k) \cdot \pi_k$$

$$V_{t,k} = \max_{x \in S} (P(y_t | k) \cdot a_{x,k} \cdot V_{t-1,x})$$

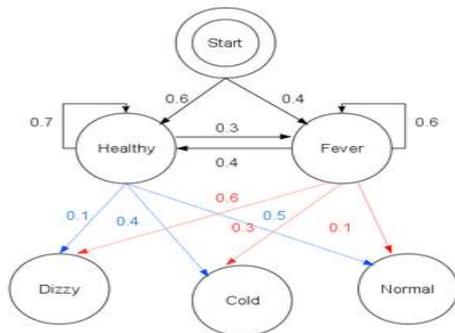
Here  $V_t, k$  is most probable state sequence responsible for the first  $t$  Observations that have  $k$  as final state.

$V_{t,k}$  if  $t > 1$ , or  $k$  if  $t = 1$ . Then:

$$x_T = \arg \max_{x \in S} (V_{T,x})$$

$$x_{t-1} = \text{Ptr}(x_t, t)$$

Here we're using the standard definition of  $\arg \max$ . The complexity of this algorithm is  $O(T \times |S|^2)$ .



Working Of Viterbi Algorithm

#### 4. CONCLUSION

The system collects probabilistic data and analysis of those collected data is performed. Finally, a viterbi prediction model is designed to foresee the Flu prediction by the most correlated patients based on their current health status. And also the location-wise, gender-wise, age-wise analysis on the data is performed.

#### 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] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," Health Information Science and Systems, vol. no. 1, pp. 1–10, 2014.

[3] P. Gope and T. Hwang, "Bsn-care: A secure iot-based modern healthcare system using body sensor network,"

IEEE Sensors Journal, vol. 16, no. 5, pp. 1368–1376, March 2016.

[4] 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.

[5] "Prasan Kumar Sahoo, Suwendu Kumar Mohapatra, Shih-Lin Wu", Analyzing Healthcare Big Data with Prediction for Future Health Condition, 2016 IEEE Translations.

[6] J. Henriques, P. Carvalho, S. Paredes, T. Rocha, J. Habetha, M. Antunes, J. Mora "Prediction of heart failure decompensation events by trend analysis of telemonitoring data" Citation information: DOI 10.1109/JBHI.2014.2358715, IEEE Journal of Biomedical and Health Informatics.

[7] Sandeep Kaur, Dr. Sheetal Kalra "Disease Prediction using Hybrid K-means and Support Vector Machine", 2016 IEEE.

[8] MIN CHEN, YIXUE HAO, KAI HWANG, LU WANG<sup>1</sup>, AND LIN WANG, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities", Digital Object Identifier 10.1109/ACCESS.2017.2694446.

[9] (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>.

[10] Padmashree T, Dr. N. K. Cauvery, "Patient Health Monitoring System and Prediction using Data Analytics", International Journal of Innovations in Engineering and Technology (IJJET) <http://dx.doi.org/10.21172/ijjet.81.022>.

[11] "Automated Detection of Influenza Epidemics with Hidden Markov Models", Toni M. Rath Maximo, Carreras and Paola Sebastiani.