International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 02 | Feb 2019 www.irjet.net p-ISSN: 2395-0072

Epidemic Prediction

Yuvaana Sundarakrishnan[1], Akshay George Koshy[2], K.P. Vijayakumar[3]

[1,2]Student, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India
[3]Asst. Professor, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

---

Abstract - *The aim of the proposed system is to predict the spread of an epidemic by analyzing the conditions in the areas where people are affected. This project is focused on one disease, Influenza, which is an infectious disease, caused by exposure to the influenza virus. The prediction will be done by analyzing the spread based on the movement of the disease through the population. It will be implemented using Artificial Neural Networks and other machine learning techniques to predict the spread in particular geographical regions.*

Key Words: **Prediction algorithm, Epidemic spread, Machine Learning, Artificial neural networks, Public health.**

1. INTRODUCTION

Influenza, which is commonly known as a Flu is an infectious disease caused by the influenza virus. There are four types of influenza viruses, namely, A, B, C, and D. Types A and B cause seasonal epidemics during winter almost every year in the United States of America. Type C virus causes mild respiratory illness, rather than epidemics. Type D mostly affects cattle and are not known to be harmful to humans.

Influenza is caused by contracting the disease from the air, due to an infected person coughing or sneezing, or through contact by touching surfaces that are contaminated by the virus.

The spread of a disease differs based on the geographical location I.e around the equator, an outbreak can happen throughout the year while around the northern and southern parts of the world, outbreaks happen during cold climates (winter). A major outbreak is known as a pandemic. Three pandemics occurred in the 20th century - Spanish Influenza with about fifty million deaths, Asian Influenza with about two million deaths and Hong Kong Influenza with about one million deaths. The outbreak of Influenza A/H1N1 was also classified as a pandemic by World Health Organization in 2009.

1.1 Scope of the Project

In order to predict the spread of this disease, which is responsible for causing about half a million deaths per year, we are developing this model. Predicting the spread of Influenza can enable the residents of at risk areas to take precautions against contracting the disease and be a step towards reducing the number of cases per year.

The rest of the survey paper is organized as follows. Section 2 describes the literature review. Section 3 presents the summary of the proposed system to be implemented. Finally, Section 4 concludes this paper.

2. LITERATURE SURVEY

Most of the prediction algorithms reviewed so far implement Support Vector Machine, Social media, Artificial Neural Networks, etc to predict the outbreak/ spread of a disease. The systems that are specific to a particular disease are more accurate in their predictions than systems that are non-specific to a particular disease. This has something do with the data set that it uses to make said predictions. Most of the current systems are used to predict outbreak rather than to predict the spread of the disease through the population.

Some of the existing systems were surveyed and the following was observed.

In [1] it describes a system which is implemented using Support Vector Machines, and Artificial Neural Networks. Certain environmental conditions of a particular area are taken as the data sets and using that testing data, predictions were made as to whether or not the an area will be affected. This approach would be very beneficial for the proposed model as it is has a similar outline, except that since they are different diseases, they would require different testing data to make predictions.

The focus of [2] is the usage of social media as a data set to make predictions using text analysis, sentiment analysis and classifications using a supervised learning model. The author used social networking platforms as their data source and got data from real-time sources. Usage of sentiment analysis means doing predictions using the severity of the epidemic based on the social media spread.

The author of [3] dealt with using spatial and covariate information from scattered sources to improve the efficiency of outbreak detection. This was a more subdued approach as there were limitations associated with using large volumes of data to perform the predictions. Hence it

was discovered that the data size has to be kept in check to approach the maximum efficiency of the algorithm.

In [4], the author proposed a convolutional neural network (CNN) that uses both unstructured and structured data types. Using actual patient medical records and filling in the missing data using latent data reconstruction to make sure that data insufficiency is not the cause of inefficiency of prediction. This implementation provided a glimpse into a custom made algorithm to predict epidemics.

The system of [5] uses a form of multivariate hidden Markov model(MHMM) to detect the spread of influenza using data compiled from heterogeneous data collected from sources such as Google query data to predict outbreaks. It emphasized the need to also consider geographical condition when considering the spread of the disease.

A protocol to consider when compiling data set for training due to the possibility of inaccuracies in data was established in [6]. The paper proposes a four step protocol to aggregate the data with a global consensus while maintaining a low fault rate.

The system implemented in [7] detects interactions of an infected node using smart phone based personal and community sensing to gauge the interactions of the infected node with other nodes and thus predicting the health of the node. This paper examines the spread of a disease with respect to a single node and predicts the number of infected nodes.

The algorithm in [8] predicts the probability of an outbreak and spread of dengue fever using data from multiple traditional and modern sources of info. It also considers a multitude of other characteristics such as the temperature and weather at the time to predict outbreaks due to dengue spread through a live vector.

From the above survey of related works, it is understood that many researchers use Support vector machine, artificial neural networks, social media, etc to have an effective prediction system. Many of the papers have a particular disease that is the focus of the study, and very few of them have a broad scope where t predict epidemics in general. Having a specific disease in focus will enable us to include the parameters that is specific to the disease, hence making the prediction more efficient.

3. PROPOSED SYSTEM

The models will be trained using existing data from previously observed incidents to predict the spread of disease during future incidents. The proposed model will take into consideration the movement of population with environmental factors which could indicate favourable conditions for the spread of the disease.

3.1 Implementation

It is proposed to use FluNet (WHO-based influenza surveillance) as a source of information of past outbreaks whose data will be used to train the model with data from time periods considered.

Implementation of the project will require the support of Anaconda (Python distribution) and Jupyter (open source document sharing web application) which will be used to store and access data while providing the libraries required to build and execute the machine learning algorithm. The data from FluNet will be used to recognize a trend in the spread of the influenza virus from person to person, which will then be used to predict the possible spread with future outbreaks (with up to date population and climate changes being taken into consideration).

4. CONCLUSION

There is a need to predict the spread of epidemics such as influenza as this will give rise to the implementation of a warning system so that at risk people can be prepared and hence we can overall reduce the number of cases in a given year. There are new diseases coming into existence almost everyday due to change in climate, evolution of the microorganisms that already exist, etc. It will be potent on our part to reduce the effects of this as much as possible.

REFERENCES

[1] Vijeta Sharma, Ajai Kumar, Lakshmi Panat, Dr. Ganesh Karajkhede, Anuradha lele, “Malaria Outbreak Prediction Model Using Machine Learning”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 4, No. 12, December 2015.

[2] Nimai Chand Das Adhikari, Arpana Alka, Vamshi Kumar Kurva, Suhas S, Hitesh Nayak, Sankalp Kumar Nayak, Vaisakh Shaj, Karthikeyan, “Epidemic Outbreak Prediction using Artificial Intelligence”, Internation Journal of Computer Science and Information Technology (IJCSIT), Vol. 10, No. 4, August 2018.

[3] David L. Buckeridge, Howard Burkom, Murray Campbell, William R. Hogan, Andrew W. Moore, “Algorithms for rapid outbreak detection: a research synthesis”, Journal of Biomedical Informatics, No.38, 2005.

[4] Min Chen (Senior Member, IEEE), Yixie Hao, Kai Hwang, (Life Fellow, IEEE), Lu Wang, Lin Wang, “Disease Prediction by Machine Learning Over Big

Data From Healthcare Communities", IEEE Access, Vol. 5, 2017

[5] Xichuan Zhou Member, IEEE, Fan Yang, Yujie Feng, Qin Li, Fang Tang, Shengdong Hu, Zhi Lin, Lei Zhang, "A Spatial-Temporal Method to Detect Global Influenza Epidemics Using Heterogeneous Data Collected from the Internet" IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 15, No. 3, May-June 1 2018.

[6] Mosab M. Ayiad and Giuseppe Di Fatta, "Agreement in Epidemic Data Aggregation", IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS), 2017.

[7] Rakesh Ranjan, Rajiv Misra, "Epidemic Disease Propagation Detection Algorithm using MapReduce for Realistic Social Contact Networks", International Conference on High Performance Computing and Applications (ICHPCA), 2014.

[8] P. Siriyasatiyan, S. Chadsuthi, K. Jampachasri, K. Kesorn, "Dengue Epidemics Prediction: A Survey of the State-of-the-Art Based on Data Science Processes", IEEE Access, Vol. 6, 19 September 2017.