

# Prediction of Epidemic Outbreaks Using Social Media Data

## Problem Statement

The epidemic disease had become the most dangerous disease for the 21<sup>st</sup> Century. The infectious disease had still gone through the outbreaks despite of the modern medical treatment. Model of modern medical treatment has had turned far side due to the epidemic disease dissemination factors such as the increase of population density and the speedy outbreaks of new infectious diseases. In this work, we propose a predictive model to predict the status of the epidemic in the specific region. For the experimental purpose, we collect the data from EHR (Electronic Health Record) as well as from the social media such as Twitter and Facebook.

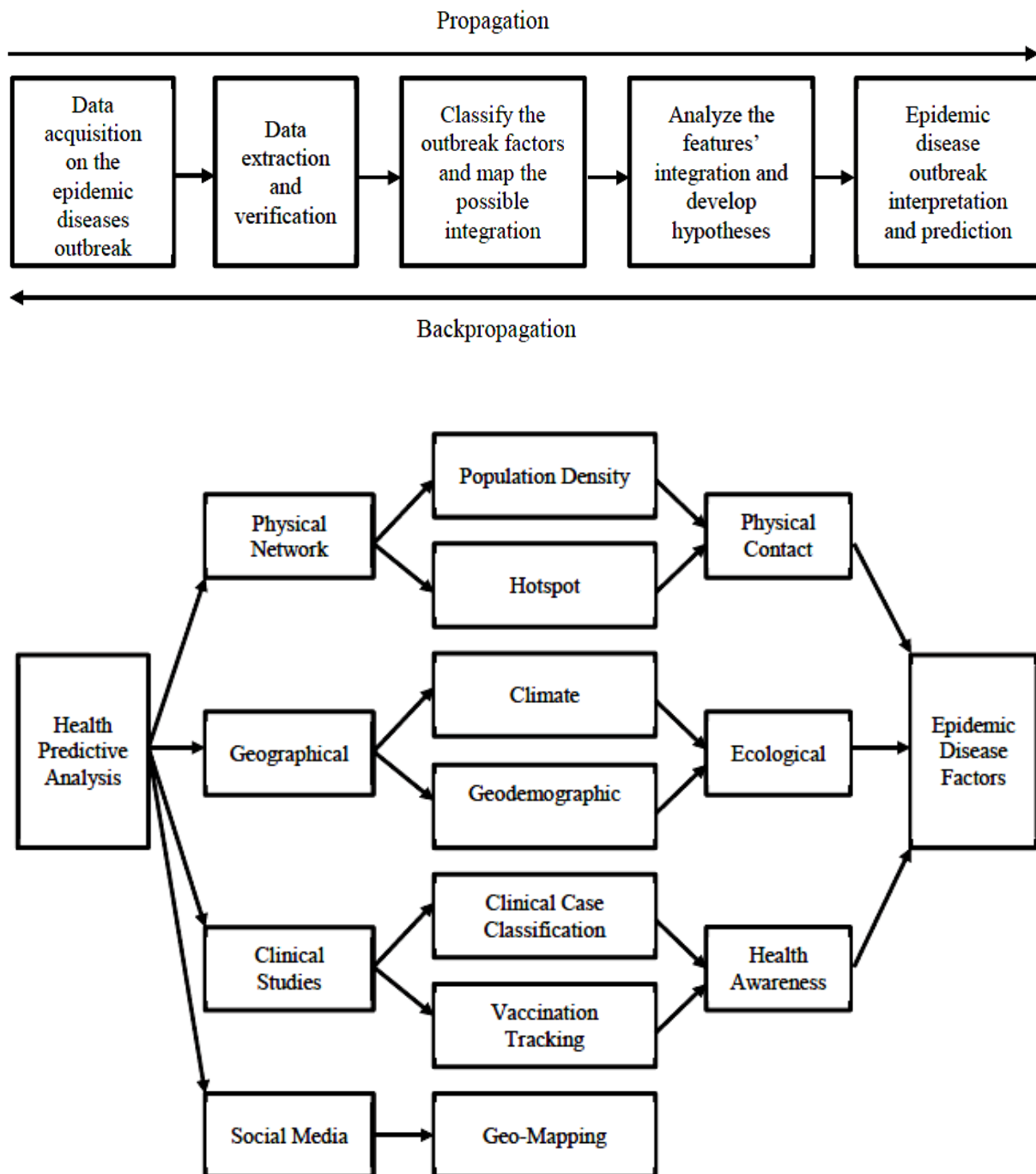
## Background

Epidemic diseases are the contagious diseases that are possible to be spread into the entire nation if the contagion measurement had reached the outbreak level and manage to wipe out the entire population [1-3]. There are some famous epidemic outbreaks that were occurred in the entire world e.g. dengue, yellow fever, cholera, diphtheria, influenza, bird flu and several others. This contagious disease had caused critical world health issues and was assumed to be one of the main factors that had caused the 43% of life lost globally. India also had experienced some of the epidemic diseases outbreak such as dengue, hepatitis, chikungunya and many more. Diphtheria may cause the inflammation of the nerves, paralysis and bleeding problem to the host [4]. The transmission usually happened via direct contact and contaminated air. However, this disease also was believed had been causing by the irregular of vaccination. These factors may contribute towards disease's identification, detection, prediction and controlling via features classification. However, due to the involvement of integration between the features, the features' clustering method was introduced and had caused the weight adjustment on the input. This features' clustering method may result in the less accurate and approximate detection and prediction on the epidemic disease. Hence, the study of finding the optimal result of the integration of clustered features was introduced via the backpropagation method that able to identify and correlate the factors that have an impact on the epidemic disease dissemination [5-6].

## Methodology

The classification can be made by propagation method or backpropagation method. The propagation method is the normal methods which classification can be done from the root to the tips. However, the backpropagation is the backward process which the classification can be done

from the tips to the root. Machine learning process was introduced in predicting the epidemic disease dissemination by cross-validate the impact of the epidemic diseases' training data using backpropagation. The backpropagation method then will identify several main factors that caused the dissemination of epidemic disease. After then these factors will be clustered and being analyzed which the correlation of the classifiers had been detected.



Nastaran Jafarpour Khameneh, 2014

## Experimental Design

***Dataset:***

EHR Electronic Health Records is the patient's clinical records and already implemented by almost all of the health organization in the developed country for tracking, data storage and clustering purposes.

***Evaluation Measures:***

The percentage of the infected people per area can be obtained by setting the area as a grid and setting the targeted area as the cells of the grid.

T is the number of day and i is the selected area to be analyze.

From this prediction of contagious disease, the contagious percentage of the disease can be predicted if the percentage value of the infectious disease outbreak are constant or increase for over 7 days monitoring process. If there are some continuations on the infections, the selected area can be quarantined and the countermeasure can be implemented based on this early prediction.

***Software & Hardware Requirements:***

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as Tensorflow, and Keras will be utilized for this process.

**References**

- [1] S. Alasaad et al., "War diseases revealed by the social media: massive leishmaniasis outbreak in the syrian spring," *Parasites & vectors*, vol. 6, no. 1, p. 94, 2013.
- [2] C. M. Bishop, *Pattern recognition and machine learning*. springer New York, Inc., 2006, vol.4.
- [3] J. P. Bradford, C. Kunz, R. Kohavi, C. Brunk, et C. E. Brodley, "Pruning decision trees with misclassification costs," dans *Machine Learning: ECML-98*. Springer, 1998, pp. 131–136.
- [4] J. Brillman, T. Burr, D. Forslund, E. Joyce, R. Picard, et E. Umland, "Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance," *BMC medical informatics and decision making*, vol. 5, no. 1, 2005.
- [5] C. O. Buckee, A. Wesolowski, N. N. Eagle, E. Hansen, et R. W. Snow, "Mobile phones and malaria: modeling human and parasite travel," *Travel medicine and infectious disease*, vol. 11, no. 1, pp. 15–22, 2013.
- [6] D. L. Buckeridge, A. Okhmatovskaia, S. Tu, M. O'Connor, C. Nyulas, et M. A. Musen, "Predicting outbreak detection in public health surveillance: Quantitative analysis to enable

evidencebased method selection,” dans AMIA Annual Symposium Proceedings, vol. 2008. American Medical Informatics Association, 2008, p. 76.