I J C T A, 10(9), 2017, pp. 869-874
© International Science Press

ISSN: 0974-5572

Detection of an Epidemics through GPS and Clinical Data

Shailaja Pede*, Ritika Hirawane, Prayag Kadam, Suvarna Joshi and Shrutika Jadhav**

ABSTRACT

Epidemics is the most severe public health issue now a days. To detect these epidemics accurately is one of the crucial tasks in the field of healthcare. Many researchers are paying their attention towards the predictive analytics of an epidemics. This paper presents an overview of proposed method for, how data analytics on electronic health records (EHR) of patients along with geographical area can be used to predict outbreak of particular disease. This paper also includes proposing an android application which notify people about an infectious disease within the region.

Index Terms: Clinical data, data mining, epidemics, prediction, Weka.

1. INTRODUCTION

Epidemic is defined as the spread if the disease in particular area within certain amount of time. Having efficient system to decline the percentage of mortality and morbidity is the crucial factor to consider in the prediction of the epidemics. Prediction is the intersection of data collection, clustering, devices/systems etc. Web searches and social media sources such as Twitter and Facebook have emerged as surrogate data sources for monitoring and forecasting the rise of public health epidemics. [1]- [3]. Many researchers uses data mining concept to predict lung cancer, heart attacks, breast cancer [4], etc. As Data Mining can catch the intricate patterns about the patients genetical and clinical characteristics which plays an important role for the predictions and to shows the possible outcomes.

Many diseases such as Dengue, Malaria, Swine Flu, etc. if been predicted prior could have saved lots of lives and also due to the sudden increase in the number of patients the hospitals were not able to provide adequate medicines. According to the statistics compiled by the Central Epidemic Command Centre (CECC) for dengue outbreak, indigenous dengue cases were confirmed with acumulative total of 41,317 imported and infrequent cases [5]. Hence, taking consideration into these points it is necessary to have efficient prediction system.

We can acquire the data from electronic health records and use Hadoop concepts [6], [7] and data mining algorithms to find the particular area affected the most and could notify the nearby people to take preventive measures. In this paper, weelucidated the data mining and predictivemodelling for the epidemic prediction in particular geographical area. We will acquire the sample data which consists certain attributes like address and phone number of each patient, then applying data mining techniques we will form a cluster of disease with specific region. If more number of patients have equivalent symptoms and belongs to same area then we can predict that it might be the start of certain epidemic. Hence we can take preventive measures by notifying the registered people before the disease gets plethoric.

*,** Assistant Professor, PCCOE, Akurdi, *Emails: pede.shailaja@gmail.com, prayag.kadam.17@gmail.com, suvarnajoshi95@gmail.com*

2. LITERATURE REVIEW

As collecting Electronic Health Records (EHR) is not a critical issue the major problem occurs while converting the data into useful information. So taking this objective into consideration two main processes are handled i.e. Data Extraction and Data Modelling.

[8] The power of the prediction with minimum relative risk of the mental health care data was evaluated. Decision support system for predicting treatment of patients was done using data mining techniques. In Data Extraction section the process of data extraction and necessary calculation are undertaken using CARLA (Centerstone Assessment of Recovery Level –Adult).

The Information extracted from realistic social contact networks can be processed by using Map Reduce concept. Initially large amount of data is divided into smaller chunks and grouped into clusters by mapper. Further considering smart mobile data as input and by checking necessary constraints reducer generates infected nodes [9].

Influenza related data collection is done with the help of web data which included web access logs, web articles, network information and search engine query [10]. Further neural network methodology is established in order to obtain relationship between influenza activities and query data thus detecting influenza epidemics [11].

A system in [12] describes about how relevant information can be extracted from massive dataset of twitter messages. Then algorithms like EARS (Early aberration reporting system) and Farrington can be used for generating predictive outputs.

EHR without modelling is like information of what occurred in past rather than prediction of future. Thus in our research paper we have elaborated methodologies used to detect and predict outbreak of epidemics.

3. PROPOSED SYSTEM

In existing system, earlier epidemic prediction systems were dependant on social data, internet search volumes and social contact networks. This makes the system unreliable. Hence, we are proposing system

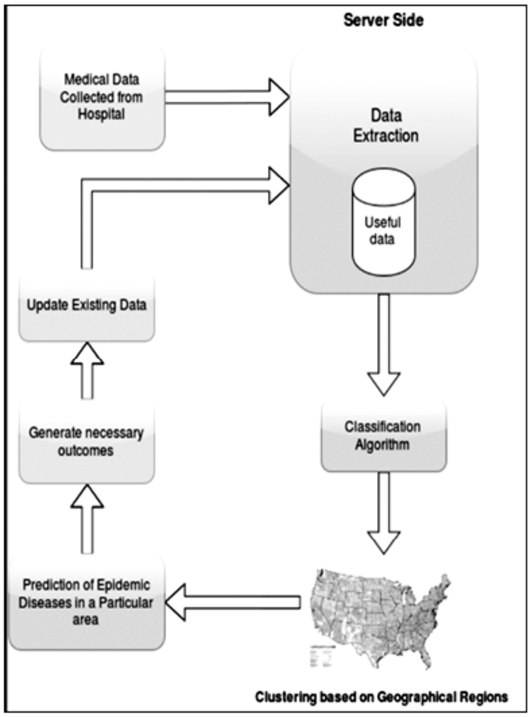

Server Side
Medical Data Collected from Hospital
Data Extraction
Useful data
Update Existing Data
Classification Algorithm
Generate necessary outcomes
Prediction of Epidemic Diseases in a Particular area
Clustering based on Geographical Regions

Figure 1: Architecture Diagram

to predict epidemic using the clinical data. We will be designing a notification application which will give the registered users information about the epidemics.

System architecture derives relationship between the components/modules. In this proposed system there are different modules such as:

- Data gathering-Initially, we will collect the medical data, focusing on attributes like Patient ID, address, phone number and symptoms.
- Data extraction - After collecting the necessary information we will extract the useful data i.e. the attributes for specific disease, data discretization is required for applying classification techniques.
- Classification - We have to apply the classification algorithm on the useful data to match with test cases. This classification can be done using data mining tools such as Weka, Knyme etc.
- Clustering – Using data mining tool, clustering algorithm, k-mean can be applied to form the clusters for respective disease based on regions.
- Prediction - After clustering is performed we will predict the epidemic in particular area and generate the necessary outcomes which helps society, hospitals to be prepared with preventive measures.
- Update –The existing data will be updated in the database.
- Notification system – The notification system will be used to give alerts to the registered user regarding the density of the area affected. This system will be as follows:

A new user will have to register to the system to get the notifications. He/she can register by submitting his mobile number and age. The location can be obtained by the GPS of that user. When the user enters a

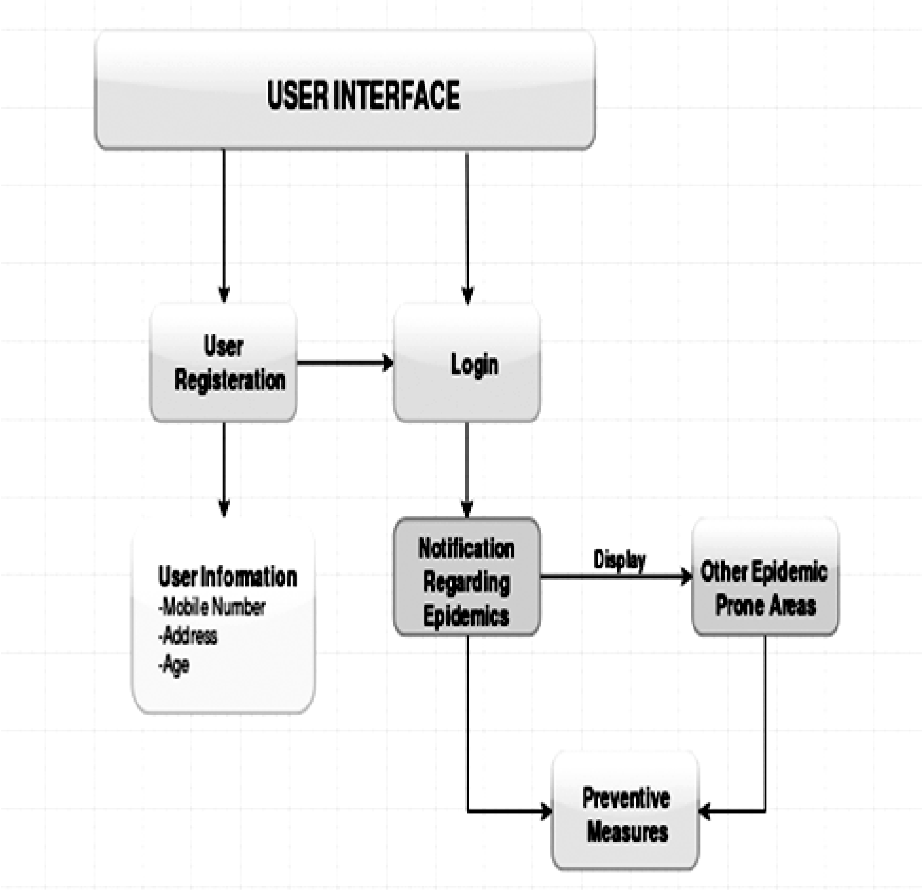


Figure 2: User notification flow diagram

region predicted as infected, he will soon get a notification describing the same. He can also check for other areas which are infected or prone to infection. After the notification is received users will be given preventive measures that should be undertaken depending on the disease of epidemic.

4. METHODS

4.1. Data Extraction

Data will be extracted from hospitals electronic health record (EHR) into a specialized layout in the data warehouse or the repository for data mining applications. Data extracted for the analysis includes variables like patients' ID, gender, age, all the measured vital characteristics i.e. related to symptoms, history of previous diagnosis, location, and region. This data will be stored in the form of comma separated values files (CSV). The target variables are the attributes defining the symptoms of the disease we are about to predict. These will form a training set. The target variables acquired will be validated for a particular disease by clinical experts. These will form the predictor variables which will be used a test set. The clinical experts will provide a range and systematic rating of each attributes of client symptoms.

4.2. Data Modelling

Classification allows us to use a set of pre-classified examples to create a model that can classify the population of records at large. It is the most commonly applied data mining technique. The reason to use classification is to predict the target class accurately for the data. Several classification algorithms like

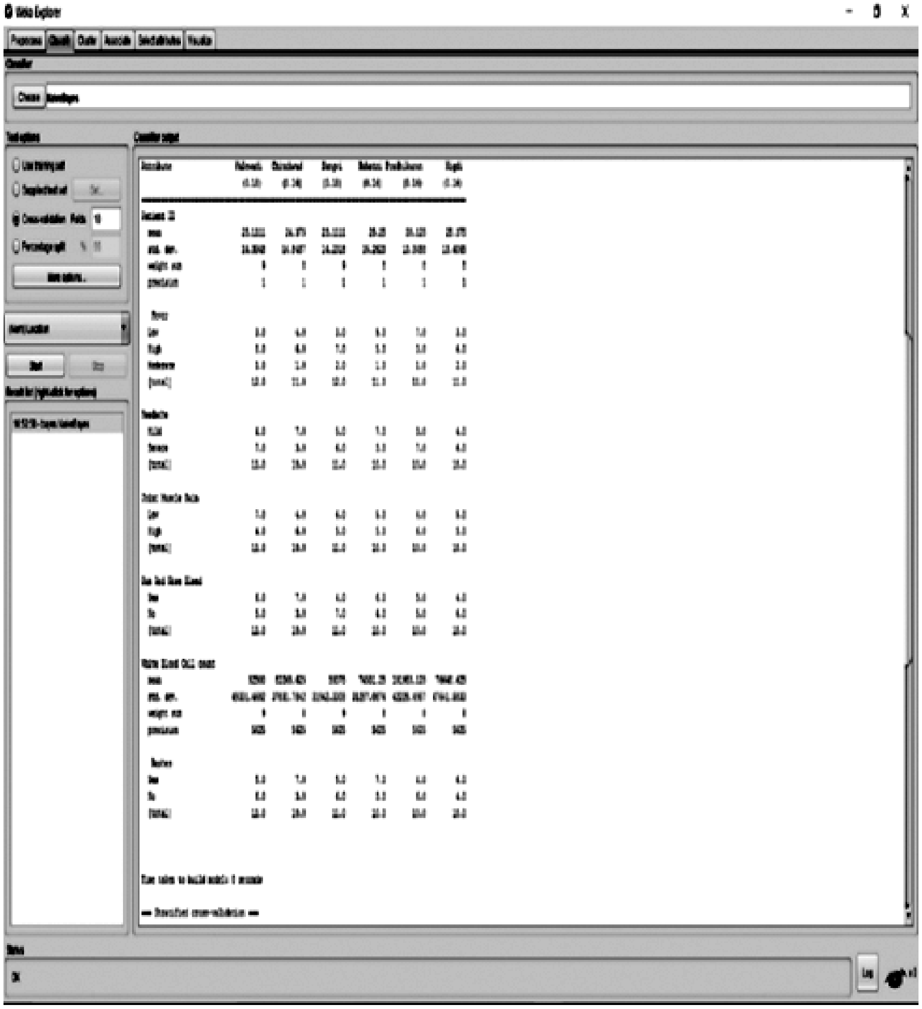
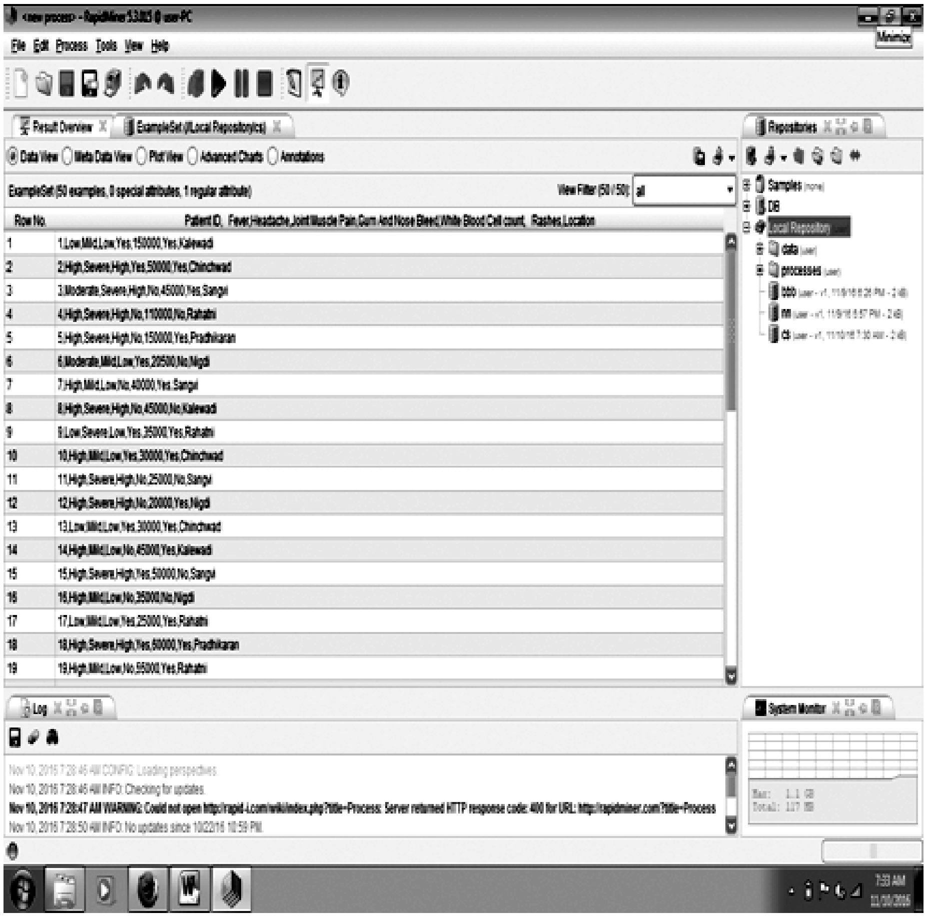


Figure 3: Output produced by WEKA Tool



Result Overview
ExampleSet (//Local Repository/cs)
○ Data View ○ Meta Data View ○ Plot View ○ Advanced Charts ○ Annotations
View Filter (50 / 50): all
Repositories
Samples
DB
Local Repository
data
processes
Log
System Monitor

ExampleSet (50 examples, 0 special attributes, 1 regular attribute)

Row No.	Patient ID, Fever,Headache,Joint Muscle Pain,Gum And Nose Bleed,White Blood Cell count, Rashes,Location
1	1,Low,Mild,Low,Yes,150000,Yes,Kalewadi
2	2,High,Severe,High,Yes,50000,Yes,Chinchwad
3	3,Moderate,Severe,High,No,45000,Yes,Sangvi
4	4,High,Severe,High,No,110000,No,Rahatni
5	5,High,Severe,High,No,150000,Yes,Pradhikaran
6	6,Moderate,Mild,Low,Yes,20500,No,Nigdi
7	7,High,Mild,Low,No,40000,Yes,Sangvi
8	8,High,Severe,High,No,45000,No,Kalewadi
9	9,Low,Severe,Low,Yes,35000,Yes,Rahatni
10	10,High,Mild,Low,Yes,30000,Yes,Chinchwad
11	11,High,Severe,High,No,25000,No,Sangvi
12	12,High,Severe,High,No,20000,Yes,Nigdi
13	13,Low,Mild,Low,Yes,30000,Yes,Chinchwad
14	14,High,Mild,Low,No,45000,Yes,Kalewadi
15	15,High,Severe,High,Yes,50000,No,Sangvi
16	16,High,Mild,Low,No,35000,No,Nigdi
17	17,Low,Mild,Low,Yes,25000,Yes,Rahatni
18	18,High,Severe,High,Yes,60000,Yes,Pradhikaran
19	19,High,Mild,Low,No,55000,Yes,Rahatni

```
Nov 10, 2016 7:28:46 AM CONFIG: Loading perspectives
Nov 10, 2016 7:28:46 AM INFO: Checking for updates.
Nov 10, 2016 7:28:47 AM WARNING: Could not open http://rapid-i.com/wiki/index.php?title=Process: Server returned HTTP response code: 400 for URL: http://rapidminer.com?title=Process
Nov 10, 2016 7:28:50 AM INFO: No updates since 10/22/16 10:59 PM.
```

Figure 4: Output produced by RapidMiner Tool

Decision trees, Fuzzy logic, Bayesian network, K-Nearest Neighbor (KNN), neural networks, Support vector machines, etc. can be used. These algorithms needs data to be operated on in a proper format so as to distinguish the target variables from other variables. Hence after the initial data gathering and calculations, it is better to convert data from .CSV to ARFF (Attribute-Relation File Format). A model is created from this obtained data and optimal performance of that model is determined using models incorporated from tools like WEKA (Waikato Environment for Knowledge Analysis), Rapid Miner, Knime (Konstanz Information Miner), etc. Weka includes several machine learning algorithms for data mining tasks. Users can use their own algorithm in the form of java code or use inbuilt algorithms to be applied on data sets. Weka provides general purpose environment tools for data pre-processing, classification, clustering, regression, association rules, visualization feature and selection. Weka has ARFF viewer which allows to view and modify the data attributes Further this data can be fed to any of the algorithms. This generates a detailed result along with summary of all the calculations. Another tool is Rapid Miner. It is a powerful visual programming environment. It is a client/server model with latest version 7. To import the data we have to use the read database option and have to create a SQL server connection. After the connection is established assign the connection to the database link. After linking it with the database and selecting .CSV file the output of the data will be displayed on the screen.

4.3. Model Evaluation

Model performance is determined using exhaustive cross validation methods. In this method the system will learn by dividing the original sample into training set and test set. The k - folds cross validation method is generally used where the data is partitioned into k equal subsamples. 1 test set and remaining k-1 as training sets. Then the validation process is performed k times each time changing the test set. The model is evaluated based on multiple performance metrics like relative errors, absolute errors, variance and standard deviation of training set from test set, various graphs, etc. Based on these all the possible cases matching test case will be obtained. These patients will be mapped based on their location into specific regional clusters. Density of patients predicted as infected with a particular disease will be calculated in this regions. If this density is beyond a calculated threshold value we can say that it might be an epidemic outbreak for that region.

5. FUTURE SCOPE

In existing system it is found that data can be collected from trusted sources and encapsulated into clusters by using various methodologies. In advancement to this auto clustering can be adapted by using advance algorithms and tools. Weather reports can also be utilized to get prior knowledge about the relationship between particular epidemic and its favorable climatic condition. Moreover in order to make people aware in more effective way mobile applications can be built which can be used to notify people about outbreak of epidemics as well as measures needed to be taken.

6. CONCLUSION

This paper mainly focuses on detection of epidemic disease propagation and altering the people by providing notifications. Earlier attempts were made to extract data from social media, social contacts which was unreliable, so this proposed system deals with the hospital data. This system may predict the accurate results using clinical data and location of patient through GPS is used. Proposed systems will help people to take preventive measures and also this will help the hospitals to take the prior action to prevent epidemics

REFERENCES

[1] Prakash, B. Aditya. “Prediction Using Propagation: From Flu Trends to Cybersecurity.” *IEEE Intelligent Systems* 31.1 (2016): 84-88.

[2] E. Aramaki, S. Maskawa, and M. Morita, “Twitter catches the flu: Detecting influenza epidemics using twitter,” in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, ser. EMNLP ’11, 2011.

[3] A. Culotta, “Towards detecting influenza epidemics by analyzing twitter messages,” in *Proceedings of the First Workshop on Social Media Analytics (SOMA’2010)*, 2010.

[4] Salama, Gouda I., M. B. Abdelhalim, and Magdy Abd-elghany Zeid. “Experimental comparison of classifiers for breast cancer diagnosis.” *Computer Engineering & Systems (ICCES), 2012 Seventh International Conference on*. IEEE, 2012.

[5] Rahmawati, Dini, and Yo-Ping Huang. “Using C-support vector classification to forecast dengue fever epidemics in Taiwan.” *System Science and Engineering (ICSSE), 2016 International Conference on*. IEEE, 2016.

[6] Ni, Jun, et al. “Hadoop-Based Distributed Computing Algorithms for Healthcare and Clinic Data Processing.” 2015 Eighth International Conference on Internet Computing for Science and Engineering (ICICSE). IEEE, 2015.

[7] Rallapalli, Sreekanth, and R. R. Gondkar. “Map Reduce Programming for Electronic Medical Records Data Analysis on Cloud using Apache Hadoop, Hive and Sqoop.”

[8] Bennett, Casey, and Thomas Doub. “Data mining and electronic health records: selecting optimal clinical treatments in practice.” arXiv preprint arXiv: 1112.1668 (2011).

[9] Ranjan, Rakesh, and Rajiv Misra. “Epidemic disease propagation detection algorithm using MapReduce for realistic social contact networks.” High Performance Computing and Applications (ICHPCA), 2014 International Conference on. IEEE, 2014.

[10] Johnson, Heather A., et al. “Analysis of Web access logs for surveillance of influenza.” *Stud Health Technol Inform* 107.Pt 2 (2004): 1202-1206.

[11] Xu, Wei, Zhen-Wen Han, and Jian Ma. “A neural network based approach to detect influenza epidemics using search engine query data.” *2010 International Conference on Machine Learning and Cybernetics*. Vol. 3. IEEE, 2010.

[12] Romano, Sara, et al. “Challenges in Detecting Epidemic Outbreaks from Social Networks.” 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE, 2016.