Original Article

CLASSIFICATION OF CHIEF-COMPLAINTS FOR PATIENTS REPORTING AT EMERGENCY DEPARTMENTS

Muhammad Asif Ali[*], Shagufta Habib[**], Zahra Ahsan[**], Muhammad Naeem Ayyaz[*]
[*]C-SVAR, University of Engineering and Technology, Lahore, Pakistan
[**]MBBS Students, University Medical and Dental College, Faisalabad, Pakistan

ABSTRACT:

Timely disease detection in case of Infectious Diseases (IDs) can provide sufficient time for disease control activities that might help in controlling the spread of disease to a great extent. Generally, medical diagnostic procedures are time consuming and may take about one to two weeks in case of diagnosis by lab reports; in the meantime, IDs spread may go epidemic. Therefore, it is required to process the complaints presented by patients' well in time for early warning of ID spread. In this study, we have designed and developed a syndromic classifier, for automated processing of chief-complaints data for patients reporting at emergency departments to corresponding syndromes.

METHODS:

We have employed Artificial Neural-Networks (ANN) for classification of patients' chief-complaints data into syndromic categories. Network is trained and tested using diagnosed chief-complaints data from leading hospitals of Lahore city.

RESULTS:

Trained network was able to assign respiratory, gastro-intestinal, hemorrhagic, rash, fever, neurological, shock syndrome with sensitivity *98.1%*, *98.9%*, *85.4%*, *97.7%*, *99.1%*, *99%* and *77.3%* respectively. No case was reported for botulinic syndrome. Comparable results were achieved by using International Classification of Disease (ICD-10) codes for classification.

CONCLUSION:

Using automated syndromic surveillance IDs may be detected two to three weeks in advance as compared to diagnosis based reporting systems. High classification accuracy of the syndromic classifier provides us with the ability to timely detect imminent ID outbreak by processing chief-complaints information. Classification results allow monitoring of ID spread and optimum time for disease control activities to mitigate the emerging ID epidemics.

KEY WORDS: Syndromic classification, Artificial neural networks

INTRODUCTION:

Every year, millions of Pakistanis are exposed, and infected with deadly diseases such as hepatitis, tuberculosis, malaria, and nowadays dengue fever. Lack of advanced surveillance infrastructure for the timely collection, reporting and analysis of epidemic spread undermines awareness for ID epidemics and poses serious challenges to public health in Pakistan. In fact, automated monitoring of the epidemic and the response to naturally or man-caused IDs does not exist in the country due to scarcity of resources, poorly trained staff and inadequate implementation of health policies. For patients reporting in emergency departments detailed diagnosis and lab-test procedures are time consuming and cause a delay in detection of ID outbreaks. ID outbreaks can be detected two to three weeks beforehand as compared to traditional methods, by grouping chief-complaints/

Corresponding Author:
Muhammad Asif Ali, C-SVAR, UET Lahore
Email: asif.ali@kics.edu.pk

symptoms into relevant syndromes and timely determining when the numbers of cases for any syndrome surpasses a certain threshold.

Globally many chief-complaints/symptoms based classifiers have been designed and developed for automated classification and surveillance of ID data, a review is given by Conway et-al[1]. For example, the notable statistical classifiers include N-gram chief-complaint classifier[2], Complaint Coder (CoCo)[3], Symptom Coder (SyCo)[4,5], Queen's University Emergency Syndromic Surveillance Team (QUESST) [6], Medical Probabilistic Language Understanding System (MPLUS-CC) [7] and Bayesian networks for disease case detection [8]. Statistical classifiers are based on machine-learning techniques, and in most cases, these classifiers require initial training data to generate the test case classification results. Dominant keyword-based classifiers include BioPortal [9], Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) [10], North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC-Detect) [11],BioSense [12], Boston Public Health Commission's Communicable Disease Division Chief-Complaint Classification System [13] and Coded Chief-Complaints for Emergency Department Systems (CC-EDS) [14]. These classifiers need a stored dictionary of medical-terms and different text variants required for text-processing routines. Finally, the linguistic classifiers include MPLUS-CC[7] , Mayo Clinic Vocabulary Server (MCVS)[15,16], and Emergency Medical Text Processor (EMT-P)[17]. MPLUS-CC falls in both statistical/ linguistic categories. These classifiers also require stored collection of medical-terms as required by key-word based classifiers. All of the above-mentioned syndromic classifiers are trained and tested using medical records for their local regions, thus these systems are trained with respect to disease seasonality of the regions under study.

OBJECTIVE OF THE STUDY:

Pakistan is a sub-tropical country experiencing huge weather changes throughout the year. Change in weather causes corresponding changes in ID epidemics. Keeping in view, impact of sub-tropical region, the drastic situation of IDs caused by floods and variation in IDs due to change in weather, we have designed and developed a syndromic classifier to process pre-diagnostic information of patients (chief-complaints/symptoms).

After detailed discussion with medical health officers and review of syndromic definitions by Chapman et-al[18] we have finalized nine syndromic categories namely: *respiratory, gastro-intestinal, hemorrhagic, rash, fever, neurological, botulinic, shock/coma, and other.* We have used ANN for classification of chief-complaints to appropriate syndromic category. This helps us to set-up a proactive paradigm to timely monitor IDs across the country, without waiting for time-consuming procedures, for example, lab reports; which may sometimes take about one to two weeks to declare confirmed diagnosis.

MATERIALS & METHODS:

This study, for the time being focusses on Lahore city, located at 31°32'59"N and 74°20'37"E of equator with an area of about 1761 km^2. This city is divided into 11 towns, which are further subdivided into 152 Union-Councils (smallest geo-graphical units of a city). Figure 1 shows town-level distribution of Lahore city, along with locations of leading public-sector hospitals.

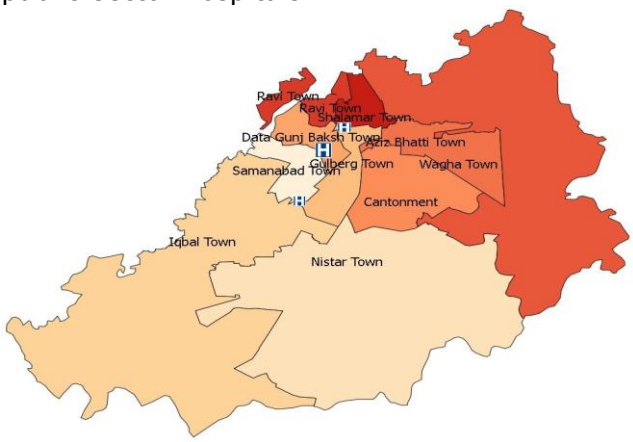

Figure 1: Map of Lahore city (town-level)

We have collected patients' data for year 2011 to 2013 from leading public-sector hospitals that share maximum patients' workload namely: *Institute of Public Health*, *Sir Ganga Ram hospital* and *Mayo hospital*. This data includes space-time stamped chief-complaints records mapped to provisional and/or

confirmed diagnosis as diagnosed by medical officers. This information is further processed for errors and missing values by text processing routines, followed by assigning ICD-10 diagnostic codes. Processed data is used to train a neural network, which automates the classification of chief-complaints into one of the nine given syndromes. Detail of each syndrome is given as under:

i) respiratory: cough, shortness of breath, upper and lower respiratory tract infections, influenza and pneumonia
ii) gastro-intestinal: vomiting, abdominal pain, nausea
iii) hemorrhagic: hemorrhagic fevers, bleeding
iv) rash: body rash, infections of skin and subcutaneous tissues
v) fever: fever, chills, shivering
vi) neurological: headache, nervous system and paroxysmal disorders
vii) botulinic: bacterial food-borne intoxications
viii) shock/coma: hypoglycaemic coma, disorders of pancreatic internal secretion, cardiogenic shock
ix) other: any case that does not lie in above-mentioned categories

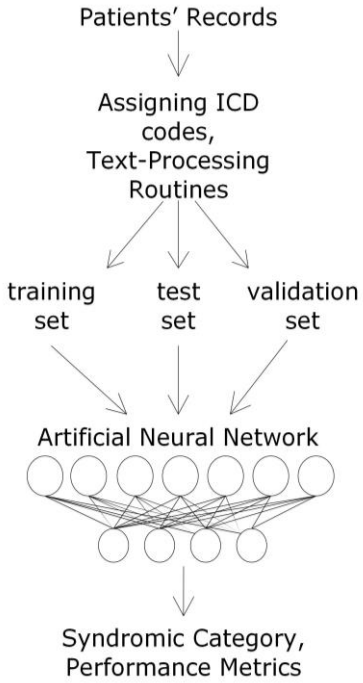

Figure 2: Syndromic classification workflow

Detailed process flow of syndromic classification is given in figure 2. Initially data is processed by text-processing routines for errors and missing values followed by assigning of ICD-10 diagnostic codes. Processed data is partitioned into *40%* training, *15%* validation and *45%* test sets. We have used two-layered feed-forward network with 40 neurons in hidden layer. Network is trained using scaled-conjugate gradient back propagation for fast learning [19] and its performance is evaluated for training, validation and test data sets. Different performance evaluation metrics are considered to evaluate classification performance of classifier including cross-entropy error, sensitivity, specificity, Receiver-Operating Characteristics (ROC), confusion matrix, classification accuracy and zero-error histogram. Detailed description of all the performance evaluation measures is given below:

Confusion matrix for case of binary classification is given in table 1; it is easily extendable for multiple classes.

Table 1: Confusion matrix for binary classification

		Actual		*Total*	
		p	*n*	*P*	*N*
Predicted	*p'*	*true positive(TP)*	*false positive(FP)*		
	n'	*false negative(FN)*	*true negative(TN)*		

Sensitivity gives an estimate of proportion of actual positives, which are predicted positive, it is calculated as,

$$TP/(TP+FN) \quad (1)$$

Specificity is a measure used to estimate proportion of actual negative which are predicted negative, it is calculated as,

$$TN/(TN+FP) \quad (2)$$

Overall classification accuracy is given as,

$$(TP+TN)/(TP+TN+FP+FN) \quad (3)$$

Finally, ROC is used to measure the percentage of true positive predictions as a function of false positive classifications.

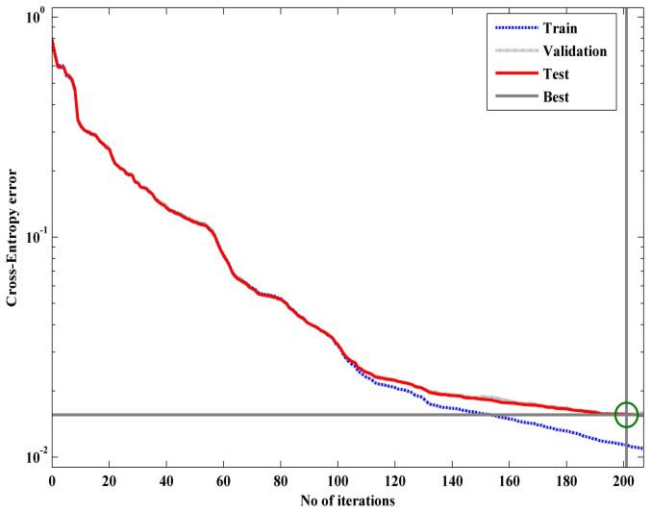

Figure 3: Performance of neural-network for training, test and validation data

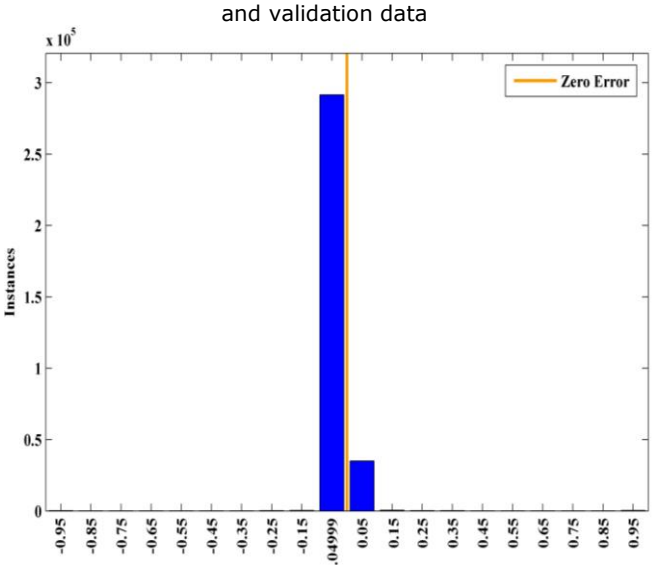

Figure 4: zero-error histogram for test data

RESULTS:

During the study period *i.e.* from 2011 to 2013, 36602 cases given with chief-complaints and syndrome information were considered for classification. Given data was partitioned into *40%*, *15%* and *45%* for training, validation and test respectively. Figure 3 shows training, test and validation set cross-entropy error for neural-network trained for classification. Minimum validation error observed is *0.015* at *200th* iteration. As another measure of classification performance of network figure 4 shows zero-error histogram with bin size of *20* bins, it is clear from the figure that absolute difference between actual and predicted values is mostly distributed near zero.

Figure 5 shows the multi-class classification confusion matrix for training, validation, test and overall data sets. Blue colored square-boxes on the main diagonal of confusion matrix represent true positive *(TP)*, red colored off-diagonal boxes represent false positive *(FP)* and false negative *(FN)* entries, bottom-most row represents corresponding sensitivity of each syndrome, and the right-most column represents positive predicted value of each syndrome. Out of 16400(45x36602/100) test cases network was able to predict 3336 cases of *respiratory* syndrome with sensitivity of *98.1%*, 4006 cases of *gastro-intestinal* syndrome with sensitivity of *98.9%*, 169 cases of *hemorrhagic* syndrome with sensitivity of *85.4%*, 427 cases of *rash* syndrome with sensitivity *97.7%*, 3464 cases of *fever* syndrome with sensitivity *99.1%*, 787 cases of *neurological* syndrome with sensitivity *99.9%* and 34 cases of *shock/coma* syndrome with sensitivity *77.3%.* In addition, 3974 test cases were classified in *other* category. There were no cases reported for *botulinic* syndrome in training and test data. Total classification accuracy for training, validation, test and overall data is *98.7%*, *98.2%*, *98.3%* and *98.4%* respectively.

Receiver Operating Characteristics (ROC) for training, test, validation, and overall data set are given in figure 6. It plots *true positive rate* of each syndrome as a function of *false positive rate.* It provides a measure of accuracy and discriminative power of trained network among different syndromes i.e. it is used to quantify sensitivity and specificity of each syndrome. It is clear from the figure that for all the syndromes the area under the ROC curve approaches unity, which explains strong discriminative power of neural network classifier. Thus for almost all the syndromes the classifier is able to identify the appropriate syndrome with sensitivity and specificity greater than *95%*, except for *botulinic* syndrome for which no case was reported in study period. Considerable results were achieved using ICD-10 diagnostics codes for classification.

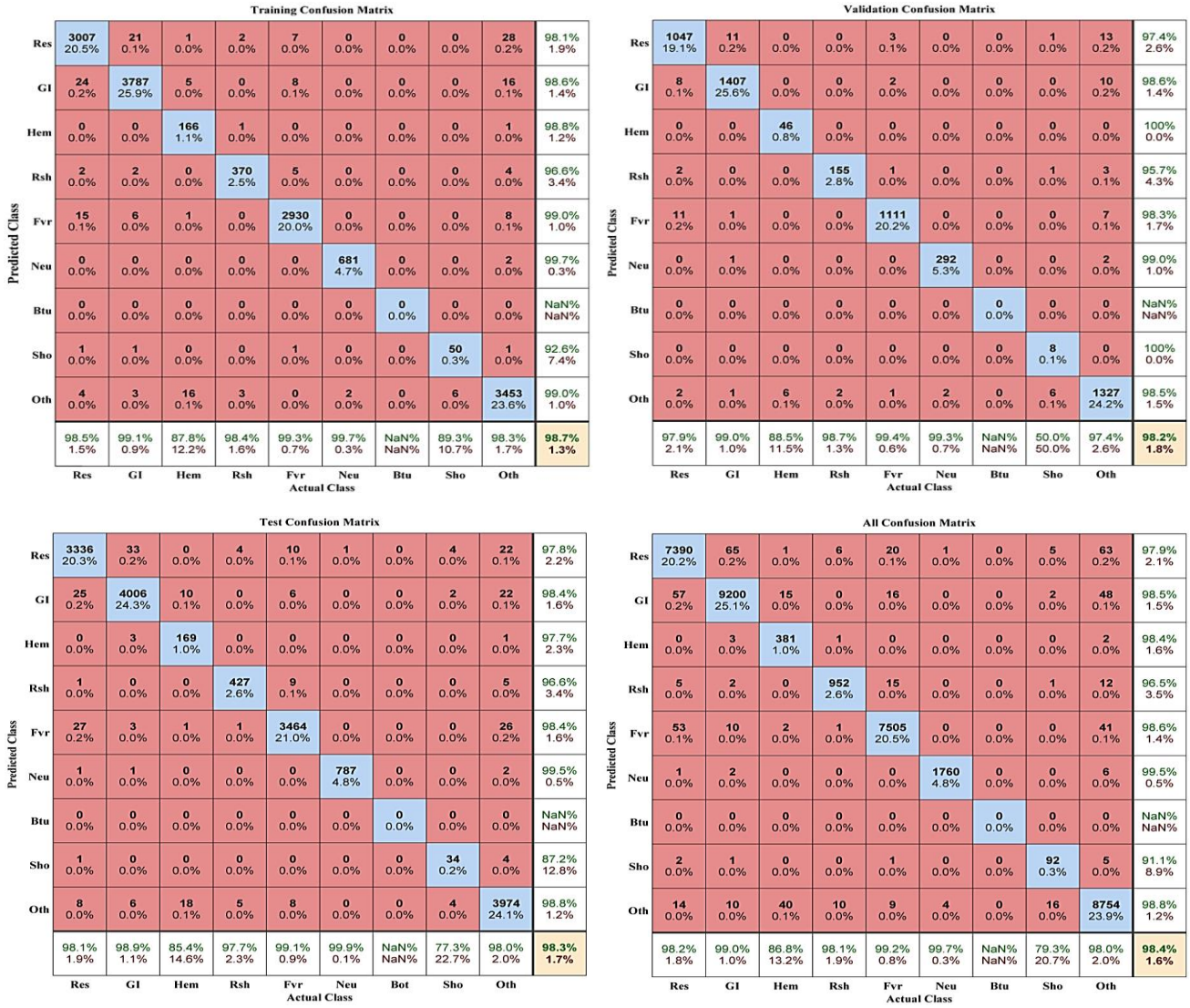

Figure 5: Confusion matrices for syndromic classification data (training, test, validation and overall data)

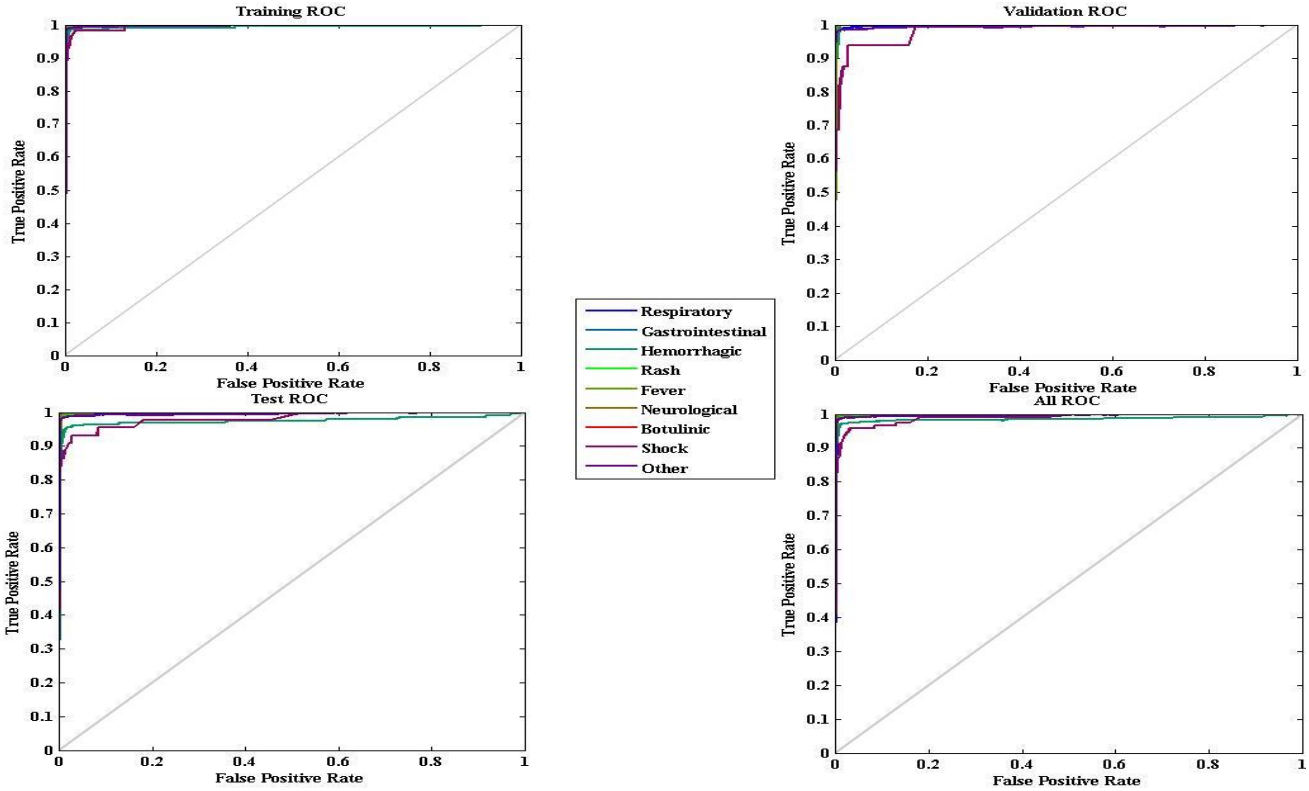

Figure 6: Receiver Operating Characteristics for training, validation, test and overall data

DISCUSSION:

This study was intended to setup an automated technique aimed at classification of chief-complaints information for individuals seeking care at hospital emergency departments and ANN was able to classify most of the complaints data to appropriate syndromes with high sensitivity, high specificity and overall test accuracy of *98.3%.*

In terms of performance metrics, our system out-performed other already implemented syndromic classifiers showing high classification accuracy, sensitivity and specificity levels. Considering statistical chief-complaint classifiers: high sensitivity was achieved in comparison with classifier developed by Chapman et al. [3], which claimed to have sensitivity ranging from *30%* to *75%* using Naïve-Bayes classifier for classification; similarly our system shows high classification performance as compared to n-gram classifier by Philip Brown et al. [5] which shows sensitivity of *70%* at specificity *96%*, however further rise in sensitivity causes huge decline in corresponding specificity for each syndrome. Classification results were also compared with dominant keyword-based classifiers namely: NC-DETECT, CC-EDS and Boston Public Health Chief-Complaint Classification System. NC-DETECT [11] shows low classification sensitivity varying form *43.78%* to *53.3%* with corresponding deterioration in specificities ranging from *95.14%* to *92.78%*; likewise, CC-EDS [14] and Boston Public Health Commission's Communicable Disease Division Chief-Complaint Classification System [13] showed low sensitivity levels as compared to our system trained using neural networks.

From ROCs given in figure 6, it can be seen that for trained neural-network, improving test case sensitivity level of syndromes does not depreciate the corresponding specificity level to a large extent. Least test case sensitivity level is achieved for Shock syndrome i.e. *77.3%* for specificity of *99%*. Table 2 shows the sensitivity level of shock syndrome under varying levels of specificity. It is evident from the table that if we increase the sensitivity beyond *95%,* there is a sudden decline in corresponding specificity level, which drops below 89%. If sensitivity level is further increased to *98.4%* corresponding specificity drops to a low value of *74%*.

Table 2: Sensitivity and Specificity for Shock Syndrome

Sensitivity	Specificity
77.3%	*99.1%*
90.9%	*98.3%*
93.18%	*97.7%*
95.45%	*89.4%*
97.73%	*84.2%*
98.4%	*74.1%*

Misclassifications in syndromic classification are caused due to the fact that chief-complaints classification is a tedious task, because patient(s) may experience same set of symptoms under different clinical conditions as disease symptoms are quite overlapping. Classification results may further be improved by considering other clinical information present at early stages of the diagnosis, or by following a hierarchical approach that keeps on updating the classification results on the basis of new information available *e.g.* lab test reports and manual diagnosis by medical health officers.

Generally, for public health surveillance systems, this classification sensitivity raises confidence level of epidemic alerts/alarms, however in case of Shock syndrome with low sensitivity it will require large number of cases to trigger epidemic alarms or adjusting the threshold values to a lower limit. High sensitivity and specificity levels of classifier allow accurate detection of syndromic outbreaks and corresponding IDs epidemic outbursts at any geographical location. Here, high sensitivity level provides the provision of detecting the ID outbreaks with minimal possibility of missing an event, whereas high specificity ensures fewer numbers of false positive alerts generated for any syndrome.

Using ICD-10 diagnostic codes the classification performance metrics did not show any substantial change, thus the trained neural network is robust to different standards being used for recording patient's chief-complaints information.

Surveillance systems aim to detect the potential ID outbreaks well in time to mitigate the IDs epidemic spreads. Our study shows

that chief-complaints contain meaningful information and may lead to detection of relevant generic category of IDs even prior to diagnosis by medical practitioner/physician. Presently in Pakistan, all the public-sector hospitals are using paper-based methods to store medical records. In current situation, it is not possible for a doctor/researcher to put together all the historical data, form hypothesis, model ID spread patterns and detect outbreaks based on multi-source data. Instead, it is necessary to develop tools that can put together related epidemiological data sources in a unified framework to support analysis and exploration.

The important purpose of syndromic classification, besides timely identifying the newer cases of the IDs, is to explore and detect the presence or absence of disease hotspots at a given place at a given time.

Consequently, we have the ability to detect disease outbreaks in any region (without relying on or waiting for the time-consuming diagnosis results) and apply corresponding counter measures strategies. As an example rise in gastro-intestinal cases at a particular location will imply spread of food-borne or water-borne epidemics and immediate control schemes will be applicable to mitigate the epidemic spread, likewise rise in vector-borne diseases i.e. constitutional and hemorrhagic fever cases will imply vector control activities. As a result, syndromic classification can be pursued for the sake of early detection of any of the above-mentioned general categories of IDs found in Pakistan.

CONCLUSION:

We have augmented the capability of nation-wide health monitoring system by introducing automated syndromic surveillance, which includes classification of chief-complaints information for individuals reporting at emergency departments. Classification results are helpful in early ID outbreak detection and reporting.

Using automated syndromic surveillance, ID outbreaks may be detected two to three weeks in advance as compared to traditional diagnosis-based disease reporting systems.

In Pakistan, syndromic classification will be pursued for the sake of early detection of any of the priorly discussed general categories of IDs. This provides us a framework for timely awareness of prevailing situations of ID with added reference to the geographic and temporal (space-and-time-stamped) origin of the disease event associated with the patient(s) in question.

In other words, proposed surveillance system will be able to timely detect and report emerging IDs, be it respiratory, gastro-intestinal or any other above-mentioned category. Introducing automated syndromic classification and surveillance in Pakistan will significantly improve health standards. It will provide us with the ability to plan and impose disease control activities as a well-timed response to ID outbreaks. Moreover, owing to well-timed ID spread control activities and counter measures planning, the associated deaths caused by IDs will reduce by many folds.

ACKNOWLEDGEMENT:

We are grateful to Centre for System Simulation and Visual Analytics Research (C-SVAR) for providing us with the research opportunity. We are also grateful to Information and Communication Technology Research and Development (ICT R&D) for providing us with the necessary tools and funds for development of this classification model.

REFERENCES:

1. Conway , Dowling N, Chapman W. Using chief complaints for syndromic surveillance: A review of chiefcomplaint based classifiers in North America. Journal of Biomedical Informatics. 2013; 46: p. 734-743.
2. Halasz S, Brown P, Goodall C, Allegra J. The ngram cc classifier: a novel method of automatically creating CC classifers based on ICD-9 groupings. In Syndromic surveillance conference; 2006.
3. Chapman WW, Dowling JN, Wagner MM. Classification of Emergency Department Chief Complaints Into 7 Syndromes: A Retrospective Analysis of 527,228 Patients. Annals of Emergency Medicine. 2005 November; 46(5): p. 445-455.

4. Espino J, etal. SyCo: a probabalistic machine learning method for classifying chief complaints into syndromes and syndrome categories. Adv Dis Surveillance. 2007; 2(5).
5. Brown P, Halász , Goodall C, Cochrane , Milano , Allegra. The ngram chief complaint classifier: a novel method of automatically creating chief complaint classifiers based on international classification of disease groupings. Journal of biomedical informatics. 2010; 43(2): p. 268-272.
6. O'Connor K, Moore K, Edgar B, McGuinness M. Maximum entropy models in chief complaint classification. In International society for disease surveillance; 2006; Baltimore.
7. Christensen M, Haug , Fiszman. Mplus: a probabilistic medical language understanding system. Proc Workshop Nat Language Process Biomed Domain. 2002;: p. 29-36.
8. Tsui F, etal. Probabilistic Case Detection for Disease Surveillance Using Data in Electronic Medical Records. Online Journal of Public Health Informatics. 2011; 3(3).
9. Lu H, Zeng D, Komatsu K, Chen H. Ontology-enhanced automatic chief complain classification for syndromic surveillance. J Biomed Inform. 2008; 41(2): p. 340-346.
10. Sniegoski A. Automated syndromic classification of chief complaint records. Johns Hopkins APL Technical Digest. 2004; 25(1): p. 68-75.
11. Scholer , Ghneim , Wu , Westlake , Travers , Waller , et al. Defining and applying a method for improving the sensitivity and specificity of an emergency department early event detection system. AMIA Annu Symp Proc. 2007;: p. 651-655.
12. Arboleda , Tokars , McMurray. BioSense real-time data initiative:improving emergency preparedness-monitoring injury sub-syndromes. Advances in disease surveillance. 2007; 4: p. 142.
13. Pendarvis J, Gunn J, Kirkwood-Smith A, Barry A. Syndromic surveillance in Boston: validating the detection of small events and integrating response from a local health department. Advances in disease surveillance. 2006; 1: p. 59.
14. Thompson DA, Eitel , Fernandes , Pines , Amsterdam , Davidson. Coded Chief Complaints—Automated Analysis of Free-text Complaints. Academic Emergency Medicine. 2006; 13(7): p. 774782.
15. Elkin P, Brown S, Husser C, Bauer B, Wahner-Roedler D, Rosenbloom S. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc. 2006; 81(6): p. 741-748.
16. Brown , Speroff , Fielstein , Bauer , Wahner-Roedler DL, Greevy , et al. eQuality: electronic quality assessment from narrative clinical reports. Mayo Clin Proc. 2006; 81(11): p. 1472-1481.
17. Travers A, Haas. Evaluation of emergency medical text processor, a system for cleaning chief complaint text data. Acad Emerg Med. 2004; 11(11): p. 1170-1176.
18. Chapman WW, etal. Developing syndrome definitions based on consensus and current use. J Am Med Inform Assoc. 2010;: p. 595-601.
19. Møller MF. A scaled conjugate gradient algorithm for fast supervised learning. Neural networks. 1993; 6(4): p. 525-533.

Submitted for publication: 06-11-2014

Accepted for publication: 10-02-2015

HE WHO IS DESERTED BY FRIENDS AND RELATIVES WILL OFTEN FIND HELP AND SYMPATHY FROM STRANGERS.

Hazrat Ali (Karmulha Wajhay)