A Flexible Syndromic Surveillance and Outbreak Detection Feedback System

Johnson Chao-Sheng Huang[1], Joseph Kwong-Leung Yu[2], and Yau-Hwang Kuo[1]
[1] Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan
[2] Pingtung Christian Hospital, Taiwan
Email: {Johnson, kuoyh}@ismp.csie.ncku.edu.tw, kwongleung.yu@gmail.com

***Abstract*—In recent years, syndrome surveillance systems have become crucial in preventing emerging infectious diseases and bioterrorism attacks in many nations. Syndrome surveillance collects large amounts of prodromal-phase symptoms. After data mining, processing, and analyzing, the obtained information is used for the early detection of outbreaks. Such systems require knowledge of epidemiological factors, syndromic grouping, and outbreak detection models. Therefore, this study proposes a set of electronic medical records based on open syndrome detection systems to cope with a flexible combination of detection algorithms. These records can show visual alerts and provide notifications for investigators to immediately grasp the epidemic prevalence and adopt appropriate preventive measures.**

***Index Terms*—syndrome surveillance, early outbreak detection, emerging infectious disease, electronic medical record**

I. Introduction

Overexploitation fosters global warming and extreme weather. Threats of emerging infectious diseases are becoming increasingly difficult to address, especially in the era of globalization [1]. Moreover, convenient transportation and frequent social activities contribute to challenges in public health disease surveillance.

Infectious diseases can significantly affect humans. For instance, the 1918 flu pandemic infected 500 million people worldwide and resulted in the deaths of 100 million, which was 3%–5% of the world's population. Severe acute respiratory syndrome in southern China caused an eventual 8,096 cases and 774 deaths reported in multiple countries with the majority of cases in Hong Kong (9.6% fatality rate) between November 2002 and July 2003. The largest outbreak recently was the Ebola virus, which began in 2014 and continued into 2015 in West Africa. This outbreak had nearly 30 thousand reported cases, resulting in more than 11 thousand deaths. The ability to detect such emerging infectious diseases as early as possible and reduce their harm has become a common goal of all public health scholars. Therefore, research on monitoring various health-related data to detect infectious disease outbreaks has been very popular.

Electronic syndromic surveillance systems have been used for the early detection of outbreaks to follow the size, spread, and tempo of outbreaks; monitor disease trends; and provide reassurance that an outbreak has not occurred. Syndromic surveillance systems seek to use existing health data in real time to provide immediate analysis and feedback to those charged with investigation and follow up of potential outbreaks. Optimal syndrome definitions for continuous monitoring and specific data sources best suited to outbreak surveillance for specific diseases have not been determined. Broadly applicable signal detection methodologies and response protocols that can maximize detection while preserving scant resources are being sought [2].

Syndromic surveillance systems detect outbreaks from daily routine heterogeneous health-related information, such as data from an outpatient department, emergency department, emergency hotline, ambulance dispatch, medication distribution, and employee attendance system. This information is assessed for any abnormal phenomena and early discovery of emerging infectious disease outbreaks via different anomaly detection methods [3]. Other studies have shown that these types of systems are also used in the prevention of bioterrorist attacks and reduction of damage [4]. An example of such an attack is the Bacillus anthracis bioterrorism incident, which caused five deaths and infected 22 people. Other examples were the use of Yersinia pestis and smallpox in bioterror attacks in the history of war. These diseases spread slowly but became severe and widely spread because there was no reporting mechanism between cases and hospitals. Although biological terrorist attacks on epidemiology are different from epidemic diseases, syndromic surveillance systems must be developed to improve public awareness and strengthen public health. Once abnormalities are perceived, immediate analysis helps determine whether there is a bioterrorist attack. Necessary actions can then be carried out before future damage is made.

In 1977, researchers in hospitals constructed information systems by randomly selecting 10% of patients to collect health-related information for the basis and follow up of disease surveillance [5]. After the 911 attack in 2001, the University of Pittsburgh in the United States developed the Real-Time Outbreak and Disease Surveillance system (RODS), which is used to collect

Manuscript received March 1, 2016; revised January 14, 2017.

doi: 10.18178/jomb.6.1.6-10

information from outpatient, emergency, and OTC medication sales. RODS analyzes chief complaints through natural language processing and classifies these complaints into seven categories. Collected information is analyzed primarily through Recursive Least Squares and WSARE1.0 (What's Strange About Recent Events) [6]. Centers for Disease Control in the United States developed Early Aberration Reporting System, which collects information from emergencies, 911 ambulance hotline, clinics, attendance system, and OTC medication sales; the information is then classified into 11 class syndromes. Cumulative Sum Control Chart and Spatial Distribution/Clustering of Syndromes use marking time anomaly detection alerts and the space detection method [7]. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II) is the only system that collects both military hospital and civil hospital data among all systems. ESSENCE II was developed by the United States Department of Defense and collects clinical data such as emergency-related information, billing codes in private practice, veterinary relevant information, and non-clinical data, including attendance records, ambulance hotline records, prescription records, and OTC medication sale information. Finally, the autoregressive modeling algorithm and exponentially weighted moving average are applied to detect abnormal time warning and smallest spatial resolution for space anomaly warning detection [8].

Compared with diagnosis, syndrome surveillance does not only predict the trend of a disease early but provides researchers with more time to investigate and reduce the damage caused by a disease. Meanwhile, the active sensitivity and positive predictive values are both higher than passive notification [9]. Therefore, with the automatic collection, detection, and notification of syndrome surveillance systems, reactions can be made faster to monitor infectious diseases. Anomaly syndrome surveillance detection usually considers a number of epidemiological characteristics, such as area, climate, season, population, clinical diagnosis, and medical records. A variety of suitable anomaly detection modules for different materials must be designed. Once a new symptom or a statistical method is found, numerous adjustments will have to be made.

The Mobile Health Information System (mHIS) is an open-source health information system used by health providers worldwide [10]. It empowers individuals to create their own HIS locally, nationally, and globally. This system is particularly helpful for middle-income countries that have to deal with much paperwork. The system works on low-cost mobile devices and computers, and it is available on cross platforms with a synchronizing online/offline function. The present study adopts the mHIS's health information system to establish a cloud syndrome surveillance system, which automatically operates self-selecting epidemiological flexible statistical model parameters that allow efficient and fast data mining. Furthermore, the proposed system is designed to send different alert notification modules for notifications of various epidemic levels.

II. System Overview

An overview of the system is shown in Fig. 1. The system can be divided into two parts, namely, the hospital aspect and the syndrome surveillance system (mHIS SSS module). The collected clinical data, such as a patient's basic profile, chief complaint, and disease code (ICD-9-CM), are translated through a gateway into an electronic XML format file based on Clinical Document Architecture Release 2 (CDAR2) before the data are uploaded and exchanged.

Given that data privacy is a primary concern, the data are transmitted through a virtual private network (VPN). Once the system receives data, it will activate an extraction using a hash method to de-identify certain information, such as patient ID, names, addresses, IDs, and phone numbers.

After the data are pre-processed, the system will classify the data depending on the epidemiological parameters and syndrome grouping for the detection algorithm. Observations are based on the web management interface that includes a GIS map and dashboard. Once an abnormal outbreak occurs, alerts will be sent via SMS messages and emails. If the system marks an individual case as highly risky, the system will automatically apply for the patient medical records for future monitors.

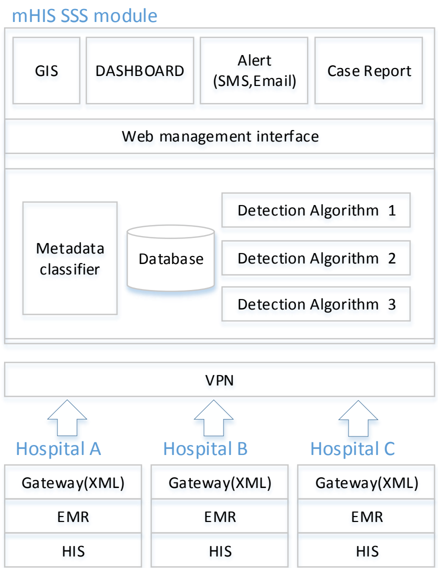


Figure 1. mHIS syndromic surveillance system module.

III. Data and Methodology

A. Data Resourse (Please see Fig. 2 and Table I)

Data used in syndrome detection systems come from a variety of different sources, so we adopt the EMR XML exchange program in Taiwan as the main object [11].

```
<cdp:ContentPackage ...>
    <ClinicalDocument ...>
        <Component>
            <StructuredBody...>
                Content of EMR...
            </StructuredBody...>
        </Component>
    </ClinicalDocument ...>
    <ds:Signature>
        Digital Signature with timestamp...
    </ds:Signature>
</cdp:ContentPackage ...>
```

Figure 2. Exchange EMR file based on CDAR2 standard.

TABLE I. EXAMPLE OF OUTPATIENT DEPARTMENT EXCHANGE PATIENT RECORD

Block	Element
Hospital info.	Hospital ID, Hospital Name
Patient info.	Personal ID Number, Chart No., Name, Gender, Birth Date, Blood Type, Rh Type, Major Illness, History of allergies, Age, Occupation, Identity Type
Date	OPD Date
Department	Department
Diagnosis	Code(International Classification of Diseases), ICD Name(International Classification of Diseases), Note
Clinical summary	Subjective, Objective, Assessment
Prescription item	Item, Procedure Code, Procedure Name, Frequency, Amount, Units, Part, Note
Prescription content	Item, Types of Prescription, Drug Code, Brand Name, Generic Name, Dosage Form, Dose, Dose Units, Frequency, Route of Administration, Medication Days, Total Amount, Total Units, Actual Amount, Actual Units, powdered, Note
Physician info.	Physician Name

B. Disease Classifier

In this paper, classification is discussed and categories are defined by public health researchers and clinic doctors. Eight syndrome groups are set according to chief complaints and ICD-9-CM. These groups are respiratory syndrome, upper GI syndrome, lower GI syndrome, hemorrhage syndrome, neurological syndrome, skin rash syndrome, influenza-like syndrome, and FEVER. For system flexibility, red eye, enterovirus, pneumonia, diarrhea, and dyspnea are also included.

C. Detection Algorithm

To detect abnormal signals effectively, the definitions of epidemiological parameters, syndrome classification, and detection algorithms are crucial. In this paper, we use two historical limits methods as examples. Please see Fig. 3. One adopts a short time period (i.e., one month) as a baseline for detecting limited information, and another adopts a long time period (i.e., 3–5 years). The threshold depends on these two methods to detect disease outbreak.

1) Historical limits method (past 2–3 years)

The HLM method relies on a straightforward comparison of the number of reported cases in the current four-week period with comparable historical data from the preceding 2–4 years. Its major strengths include simplicity, interpretability, and implicit accounting for seasonal disease patterns. These strengths make the HLM method a potentially useful aberration detection method for health departments. The Bureau of Communicable Disease of the New York City Department of Health and Mental Hygiene implemented the HLM in the early 2000s as a weekly analysis for all reported diseases for which at least 5 years of historical data were available. In applying this method in New York City, only increases in case counts >2 SD above the historical mean are considered because artifactual decreases in case counts would be detected by separate quality-control measures[12].

	Last week	*This week*	*Next Week*
Current Year		X_0	
Year-1	X_1	X_2	X_3
Year-2	X_4	X_5	X_6
Year-3	X_7	X_8	X_9
Year-4	X_{10}	X_{11}	X_{12}

Figure 3. An example of 5-year period analysis based on historical limits method.

$$\bar{X} = \frac{1}{n}\sum_{i=0}^{n} X_i$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

X_i: *Average weekly cases number*

σ: *Standard Deviation*

2) Historical Limits Method (past four weeks)

This method runs each day for the four-week interval. For a signal to be generated, the current period must contain at least three cases, and the ratio of cases to the historical mean should be greater than historical limits. Disease reviewers are promptly notified of any signals and provided with a corresponding case line list.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_{t+id}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (X_{t+id} - \bar{X})^2}$$

d: *Interval of day*

X_t: *Case number*

σ: *Standard Deviation*

D. Alerm Notification

The system does not only alert the notification module but actively informs staff, including infection control units, frontline units, and hospital management units, that an abnormal signal has reached a certain level.

The system provides flexibility on levels of sending alerts under different detection methods. To prevent unnecessary false alarms upgrading the efficiency of observation, the following alarm notifications are used:

1) By severe case report

Alerts will be sent once single/multiple sever syndromes is/are detected.

2) By detection algorithm

Alerts will be sent once single/multiple syndromes match certain algorithms.

3) By alarm numbers

Alerts will be sent once single/ multiple syndromes is/are detected within certain times.

4) By report location

Alerts will be sent once single/ multiple syndromes is/are detected within a day or a certain time.

Alert messages include information such as name of the syndrome, date, and listed suspected names for further follow-ups.

IV. Result

mHIS SSS module has been designed to have different functions for different user roles. The purpose is to allow users to quickly obtain the information they need. Please see Fig. 4-8.

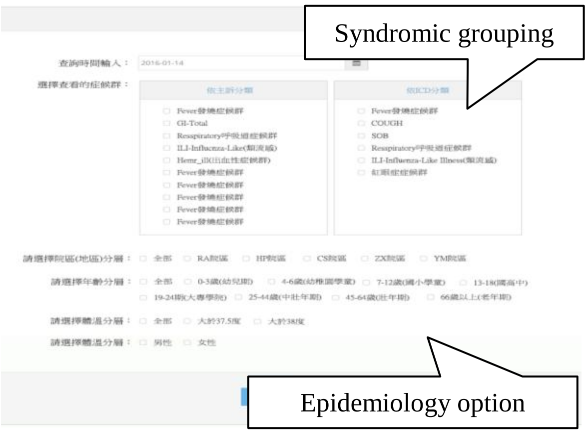


Figure 4. Create a new syndromic grouping and detention model definition.

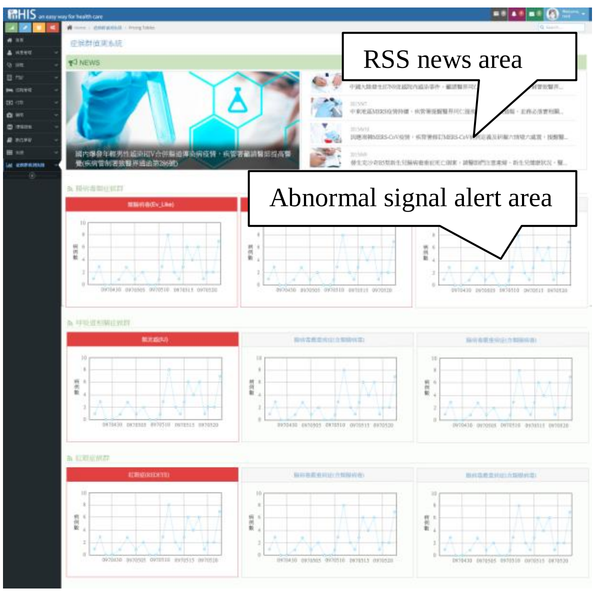


Figure 5. Syndromic surveillance system dashboard.(RSS news area shows related infectious information and data collected by RSS. Abnormal signal alert area shows types of defined surveillance groups. The red ones display outbreak detection).

A. Syndromic Surveillance Management Group

The users include public health researcher, epidemic investigators and infection control administrator that are allowed to use the system tool for outbreak investigations.

B. Outbreak Monitor Group

The users include Infection control staff and policy maker that are allow to release news and notification.

C. Hospital staff

The users include all the hospital staffs witch can improve alertness during service.

The Fig. 7 demonstrates how syndromic surveillance system works in emergency department. When a patient receives emergency triage, doctors and nurses will be reminded at once if he/she is found high-risk possibility through ER HIS system. The acquisition of specimen will be immediately taken and sent to the laboratory for testing, then the patient will be waiting in ER ward fore further confirmation.

Figure 6. Detailed results including case number, threshold, patient profile, and GIS map.

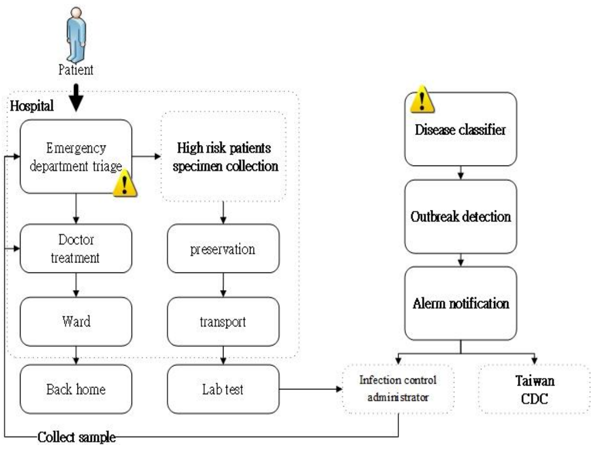


Figure 7. Syndromic surveillance system works in emergency department.

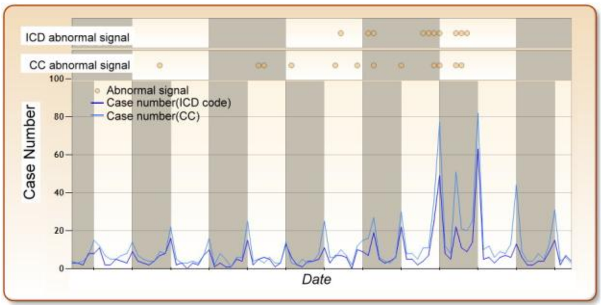


Figure 8. As different detection sources based on chief complain (CC) and ICD9.

V. Conclusion

The current disease surveillance systems in Taiwan include Communicable Disease Reporting System, Sentinel Physician Surveillance System, School Absenteeism Log, Infectious Disease Surveillance System for Institutions with Dense Population, and Preventive Inoculation Information System [13]. All these systems depend on human judgment to decide whether to send alerts. In this study, we have constructed a surveillance system with high flexibility. Automated collection and detection will be conducted when hospitals upload medical EMR files through EEC. Meanwhile, the tools provided with the mHIS SSS module allow operators to efficiently select and monitor target audiences. The proposed system is more active and rapid compared with traditional ones.

Syndrome detection systems are effective for early prevention worldwide. This open source project will be under the GPL license, so programmers can further develop and use resources. We expect the involvement of other parties and development of more detection algorithms that can benefit the public.

Acknowledgment

We would like to show our gratitude to the promotion and past experience of Pingtung Christian Hospital. The cloud system framework design is based in part upon work supported by the Ministry of Science and Technology under Grant Nos. NSC101-2221-E006-259-MY3.

References

[1] A. J. McMichael, “Globalization, climate change, and human health,” *New England Journal of Medicine*, vol. 368, no. 14, pp. 1335-1343, 2013.

[2] K. J. Henning, “What is syndromic surveillance?” *Morbidity and Mortality Weekly Report*, vol. 53, pp. 7-11, 2004.

[3] K. D. Mandl, *et al.*, “Implementing syndromic surveillance: A practical guide informed by the early experience,” *Journal of the American Medical Informatics Association*, vol. 11, no. 2, pp. 141-150, 2004.

[4] N. Meyer, J. McMenamin, C. Robertson, M. Donaghy, G. Allardice, and D. Cooper, “A multi-data source surveillance system to detect a bioterrorism attack during the G8 Summit in Scotland,” *Epidemiology and Infection*, vol. 136, no. 7, pp. 876-885, 2008.

[5] A. L. Kaszuba and G. Gibson, “Hospital emergency department surveillance system: A data base for patient care, management, research and teaching,” *Journal of the American College of Emergency Physicians*, vol.6, no.7, pp. 304-307, 1977.

[6] J. U. Espino, *et al.*, “The rods open source project: Removing a barrier to syndromic surveillance,” in *Proceeding Medinfo.: Studies in Health Technology and Informatics*, 2004, pp. 1192-1196.

[7] L. Hutwagner and J. V. Barson, “Use of the early aberration reporting system (EARS) for detection of bioterrorism agent attacks,” *Aviation, Space, and Environmental Medicine*, vol. 76, no. 10, pp. 1001-1002, 2005.

[8] J. Lombardo, *et al*., “A systems overview of the electronic surveillance system for the early notification of community-based epidemics (ESSENCE II),” *Journal of urban health*, vol. 80, no. 1, pp. i32-i42, 2003.

[9] P. Piriyawat, *et al*., “Comparison of active and passive surveillance for cerebrovascular disease the brain attack surveillance in corpus christi (BASIC) project,” *American Journal of Epidemiology*, vol. 156, no. 11, pp. 1062-1069, 2002.

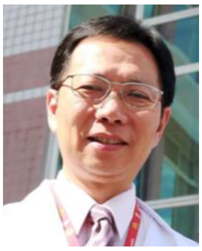
[10] mHIS Open Project, [Online]. Available: http://www.mhis.pro/

[11] W. S. Jian, *et al*., “Building a portable data and information interoperability infrastructure—framework for a standard Taiwan electronic medical record template,” *Computer methods and programs in biomedicine*, vol. 88, no. 2, pp. 102-111, 2007.

[12] A. Levin-Rector, *et al*., “Refining historical limits method to improve disease cluster detection, New York city, New York, USA,” *Emerging Infectious Diseases*, vol. 21, no. 2, pp. 265-272, 2015.

[13] F. Y. Shih, *et al*., “Challenges faced by hospital Healthcare workers in using a syndrome - based surveillance system during the 2003 outbreak of severe acute respiratory syndrome in Taiwan,” *Infection Control and Hospital Epidemiology*, vol. 28, no. 3, pp. 354-357, 2007.

Johnson Chao-Sheng Huang is a Ph.D. student at Department of Computer Science in National Cheng Kung University and IT consultant in Ping-tung Christian Hospital. His master degree was in department of medical informatics from Taipei medical university. He has been dedicated in the development of domestic and foreign health IT and led EMR Program which serving over 250 hospitals as a technical team leader. Now interested in building a cloud based open source health information system that allows developing countries to implement health IT easily.



Joseph Kwong-Leung Yu is an associate professor department of medical in Kaohsiung Medical University. He received his Ph.D. in institute of medical research from the same university. Currently Dr.Yu is anesthesiologists and also the superintendent of Pingtung Christian Hospital. His research interests focus on the international cooperation and oversea medical mission.

Yau-Hwang Kuo is a distinguished Professor in the Department of Computer Science in National Cheng Kung University. He published several journals and international conference paper from his research work. He received his Ph.D. in computer science and engineering from Department of Computer Science in National Cheng Kung University. He is interested in embedded System and Application, Sensor Network and Ubiquitous Computing, Broadband Networking, Intelligent Information System, Web Service Technology, Object-oriented Design, Fuzzy System, Knowledge-based System, SoC Design, Artificial Intelligence, Context Aware, Digital Home.