Identifying Sociomarkers of Pediatric Asthma Patients at Risk of Hospital Revisiting

Eun Kyong Shin*, Ruhi Mahajan, Oguz Akbilgic and Arash Shaban-Nejad

UTHSC Oak Ridge National Laboratory- (UTHSC-ORNL) Center for Biomedical Informatics, Memphis, TN, USA

Objective

Asthma is one of the most common chronic childhood diseases in the United States [2, 3]. In addition to its pervasiveness, pediatric asthma shows high sensitivity to the environment. Combining medical-social dataset with machine learning methods we demonstrate how socio-markers play an important role in identifying patients at risk of hospital revisits due to pediatric asthma within a year.

Introduction

A socio-marker is a measurable indicator of social conditions where a patient is embedded in and exposed to, being analogous with a biomarker indicating the severity or presence of some disease state. Social factors are one of the most clinical health determinants [1], which play a critical role in explaining health outcomes. Socio-markers can help medical practitioners and researchers to reliably identify high-risk individuals in a timely manner.

Methods

We collected data from three different sources: pediatric asthma encounter records from Jan 1st, 2016 to Dec 31st, 2016 at a children's hospital, the 2010 U.S census data and neighborhood quality survey data by Memphis Property Hub. After merging these datasets we examine the effect of social features in identifying the patients who visited the hospital more than once during the observation period. We only use the first time visit (3,678 cases) to avoid over-counting of the same patients. In addition to demographic features (age, gender, insurance type, and race (African American and White)), we incorporate the social features such as the proportion of individuals living below the federal poverty level, blight prevalence, neighborhood quality, neighborhood quality inequality, trash dumping presence, the broken window pervasiveness within the zip code area of patients' residence are included.

We then implemented a Support Vector Machine (SVM) based classification model using abovementioned 11 social features. The classification outcome is either *patient visits the hospital only one-time* (class 0) or *revisits the hospital within a year* (class 1). Among 3,678 unique patients in the dataset, only 823 pediatric patients revisited hospital with asthma. So, to overcome the class imbalance issue, we have used 823 patients' data (randomly selected in 1,000 iterations) from each class. Further, to avoid overfitting and ensure generalizability, we divided the dataset as training, test, and validation with a proportion of 60%, 20%, and 20%, respectively. The reported test (5-folds cross-validation using training and testing data) and validation accuracy of the SVM method are averaged over 1,000 iterations to avoid sampling error and bias.

Results

The proposed socio-marker model resulted in an average classification accuracy of 63.70% for the test set and 63.67 % for the validation set. Further, the average specificity (the total true negative cases divided by the sum of true negative and false positive) and sensitivity (the total number of true positive cases divided by the sum of positive predicted cases) is found to be 62.79% and 64.77%, respectively for the test set and 62.79% and 64.83%, respectively for the validation set. Results of this study suggest that socio-marker features that are not directly related to a patient's medical conditions can still predict whether the patient will come back to the hospital within a year or not with approximately 64% accuracy.

Conclusions

Bringing the socio-marker features in the surveillance system may ease the burden of detecting the patients at risk of revisiting the hospital. The results should be interpreted with caution because we only used 12-month period of observation and the visit beyond the observation window is not considered. Also the patients may have visited different hospitals which are not captured in the data.

Keywords

Sociomarkers; Asthma Patients at Risk; Machine learning; Health Surveillance; Social Components of Health

References

1. Booske BC, Athens JK, Kindig DA, Park H, Remington PL: Different perspectives for assigning weights to determinants of health. University of Wisconsin: Population Health Institute 2010.
2. Subbarao P, Mandhane PJ, Sears MR: Asthma: epidemiology, etiology and risk factors. Canadian Medical Association Journal 2009, 181(9):E181-E190.
3. Gold DR, Wright R: Population disparities in asthma. Annu Rev Public Health 2005, 26:89-113.

*Eun Kyong Shin
E-mail: eshin3@uthsc.edu