Using Data Analytics to Predict Hospital Mortality in Sepsis Patients

Yazan Alnsour, University of Illinois at Springfield, Springfield, USA
Rassule Hadidi, University of Illinois at Springfield, Springfield, USA
Neetu Singh, University of Illinois at Springfield, Springfield, USA

ABSTRACT

Predictive analytics can be used to anticipate the risks associated with some patients, and prediction models can be employed to alert physicians and allow timely proactive interventions. Recently, health care providers have been using different types of tools with prediction capabilities. Sepsis is one of the leading causes of in-hospital death in the United States and worldwide. In this study, the authors used a large medical dataset to develop and present a model that predicts in-hospital mortality among Sepsis patients. The predictive model was developed using a dataset of more than one million records of hospitalized patients. The independent predictors of in-hospital mortality were identified using the chi-square automatic interaction detector. The authors found that adding hospital attributes to the predictive model increased the accuracy from 82.08% to 85.3% and the area under the curve from 0.69 to 0.84, which is favorable compared to using only patients’ attributes. The authors discuss the practical and research contributions of using a predictive model that incorporates both patient and hospital attributes in identifying high-risk patients.

KEYWORDS

Health Analytics, Health IT, In-Hospital Mortality, Predictive Models, Sepsis

INTRODUCTION

The Centers for Medicare and Medicaid Services (CMS) estimated that the healthcare expenditure in the United States in 2016 alone was $3.4 trillion, an increase of about 4.8% from 2015. According to the CMS, this trend will continue, with an average growth rate of 5.6% per year until 2025 (“National Health Expenditure Data,” 2018), and this significant growth is partially the result of an aging population, increased lifespans, and growing costs associated with repeated hospital visits for patients with serious infectious diseases.

Sepsis is an inflammatory condition caused by infection and results in a relatively high mortality rate (Amland & Hahn-Cover, 2016; Singer et al., 2016). Sepsis and septic shock are common causes of morbidity and mortality (Raghavan & Marik, 2006). The condition of patients with sepsis can change from stable to near death in a very short period, from days to even just several hours (Taneja et al., 2017). Early diagnosis and prompt treatment have been associated with improved outcomes and
alleviated risks (Nachimuthu & Haug, 2012). During the past decade, cases of sepsis have increased significantly (Raghavan & Marik, 2006), and sepsis is considered to be one of the main causes for admission to intensive care (Vincent et al., 1996). In the United States, for example, sepsis accounts for more cases of death than prostate cancer, breast cancer, and AIDS combined (“Sepsis Fact Sheet,” 2016). The Agency for Healthcare Research and Quality (AHRQ) indicated that sepsis is the most expensive condition treated in U.S. hospitals, costing more than $20 billion in 2011, with an average annual increase of 11.9%. Fleischmann et al. (2016) argued that reducing the burden of sepsis is a global challenge, and many countries suffer from a high level of mortality and morbidity from it. Given the current state of sepsis complications and the significant social and economic benefits of better treatment of this disease, it is paramount for health care providers to find more effective ways to deal with sepsis and improve the health outcomes of sepsis patients.

According to the Centers for Disease Control and Prevention, sepsis can be a challenge for medical providers and health care professionals (HCPs) because there is no standard diagnostic test for it (Epstein, 2016). Sepsis diagnosis relies on the judgment of HCPs. Furthermore, it has been estimated that if a country like the United States can achieve an earlier sepsis identification and evidence-based treatment, there will be 92 thousand fewer deaths and savings of more than 1.5 billion in medical cost annually (Shorr, Micek, Jackson, & Kollef, 2007). Predictive analytics in healthcare facilities has mostly been limited to the simple heuristics and scoring systems. HCPs can leverage predictive analytics and machine-learning (ML) techniques to harness the variables available through electronic health records (EHRs) to better predict patient outcomes (Bhattacharjee, Edelson, & Churpek, 2017).

In this study, we explored how predictive models can be used to identify patients with a high mortality risk while they are hospitalized. We empirically developed and examined our model using a large archival dataset from the National Inpatient Sample (NIS), which was created by the AHRQ through a federal-state-industry partnership. A total of 1,048,575 sepsis patients’ records were collected between the years 2008 and 2012. This sample was subdivided randomly into training, testing, and validation data partitions. Death during hospitalization (in-hospital mortality) was defined as a target variable. The independent predictors of mortality were identified using the Chi-square Automatic Interaction Detector (CHAID) model in testing. Additionally, the model was validated using receiver-operating characteristic (ROC) curves and accuracy metrics.

BACKGROUND

The healthcare literature indicates that patients’ mortality risk is higher when patients do not receive appropriate medical interventions and follow-ups (Schmitt et al., 2013). HCPs need the appropriate tools to make the right decisions and take the needed actions at the right time to enhance healthcare outcomes (Koh & Tan, 2011; Lee et al., 2003; Martin et al., 2009; Mehta et al., 2002; Varlamis et al., 2017). A major benefit anticipated from advancements in healthcare information systems and big data is the ability to create new insights by applying advanced data analytics to enhance health care delivery and quality (Hagland, 2011; Kociol et al., 2012; Varlamis et al., 2017).

Clinical support tools can help overwhelmed medical professionals better manage a patient’s health (Chen, Chiang, & Storey, 2012; Srinivas, Rani, & Govrdhan, 2010). One of the distinct new features of clinical support tools is the ability to provide HCPs with some type of outcome predictions. Clinical support tools that have predictive capabilities can provide HCPs with early alerts to enhance clinical decisions and actions for a better level of care and a reduction of unnecessary risk (Mitchell et al., 2016; Varlamis et al., 2017). Insights regarding the prediction of mortality risks, for example, can help HCPs and hospital management identify patients at high risk of in-hospital mortality. This will allow HCPs to be proactive in their actions and apply early interventions to improve patients’ health outcomes, alleviate unnecessary risks, and unnecessary cost (Mitchell et al., 2016; Roshanov et al., 2013; Varlamis et al., 2017). Further, predictive models
Related Content

A Semantic-Based Dynamic Search Engine Design and Implementation for Electronic Medical Records
[www.igi-global.com/article/semantic-based-dynamic-search-engine/43917?camid=4v1a](www.igi-global.com/article/semantic-based-dynamic-search-engine/43917?camid=4v1a)

Non-Traditional Data Mining Applications in Taiwan National Health Insurance (NHI) Databases: A Hybrid Mining (HM) Case for the Framing of NHI Decisions
SMART: Mobile Patient Monitoring in an Emergency Department
*Healthcare Delivery Reform and New Technologies: Organizational Initiatives* (pp. 261-275).
[www.igi-global.com/chapter/smart-mobile-patient-monitoring-emergency/50164?camid=4v1a](www.igi-global.com/chapter/smart-mobile-patient-monitoring-emergency/50164?camid=4v1a)

Automated Generation of SNOMED CT Subsets from Clinical Guidelines
Carlos Rodríguez-Solano, Leonardo Lezcano and Miguel-Ángel Sicilia (2013).
*Information Systems and Technologies for Enhancing Health and Social Care* (pp. 190-204).
[www.igi-global.com/chapter/automated-generation-snomed-subsets-clinical/75629?camid=4v1a](www.igi-global.com/chapter/automated-generation-snomed-subsets-clinical/75629?camid=4v1a)