Chapter 22
Predictive Modeling of Surgical Site Infections Using Sparse Laboratory Data

Prabhu RV Shankar

University of California Davis, USA

Anupama Kesari

Sri Jayachamarajendra College of Engineering, India

Priya Shalini

Sri Jayachamarajendra College of Engineering, India

N. Kamalashree

Sri Jayachamarajendra College of Engineering, India

Charan Bharadwaj

Sri Jayachamarajendra College of Engineering, India

Nitika Raj

Sri Jayachamarajendra College of Engineering, India

Sowrabha Srinivas

Sri Jayachamarajendra College of Engineering, India

Manu Shivakumar

State University of New York at Buffalo, USA

Anand Raj Ulle

Sri Jayachamarajendra College of Engineering, India

Nagabhushana N. Tagadur

Sri Jayachamarajendra College of Engineering, India

ABSTRACT

As part of a data mining competition, a training and test set of laboratory test data about patients with and without surgical site infection (SSI) were provided. The task was to develop predictive models with training set and identify patients with SSI in the no label test set. Lab test results are vital resources that guide healthcare providers make decisions about all aspects of surgical patient management. Many machine learning models were developed after pre-processing and imputing the lab tests data and only the top performing methods are discussed. Overall, RANDOM FOREST algorithms performed better than Support Vector Machine and Logistic Regression. Using a set of 74 lab tests, with RF, there were only 4 false positives in the training set and predicted 35 out of 50 SSI patients in the test set (Accuracy 0.86, Sensitivity 0.68, and Specificity 0.91). Optimal ways to address healthcare data quality concerns and imputation methods as well as newer generalizable algorithms need to be explored further to decipher new associations and knowledge among laboratory biomarkers and SSI.

INTRODUCTION

Surgical Site Infections are one of the most common Hospital Acquired Infections (HSI). The United States based Centers for Disease Control and Prevention (CDC) has developed criteria, as part of its SSI surveillance program, that define surgical site infection (SSI) as any infection that occurs at or near the surgical incision after an operative procedure within 30 days of the procedure, expanded up to 90 days if prosthetic material is implanted at surgery (Horan, Gaynes, Martone, Jarvis, & Emori, 1992). SSIs are commonly confined to the incision site even though, they can extend deep into the underlying tissues. SSI is a significant burden to the post-operative patients, overall responsible for up to 30% of all hospital acquired infections and complicates 23% of abdominal surgeries (Magill et al., 2012; Aga et al., 2015; de Lissovoyet al., 2009). SSI contributes to increased morbidity, poor quality of life, mortality (up to 3%), prolonged hospital stay (up to 10 days), increased readmission rate and healthcare expenditure (Owens, Barrett, Raetzman, Maggard-Gibbons, & Steiner, 2014). There are many known and yet-to-be-known factors that play a role in SSI and a clear understanding of those factors are needed to reduce SSI and improve surgical outcomes (Lawson, Hall, & Ko, 2013; Ke et al., 2017).

LABORATORY TESTS IN THE ERA OF BIG DATA

One of the time-tested resource that has played a significant role in helping providers make decisions about choosing patients, readiness for surgery as well as intra and post-operative management are laboratory test results, such as blood counts, blood sugars including metabolic panels and inflammatory markers. These lab test results are indispensable for all aspects of patient management. The information content in lab test results could be a valuable resource to decipher unknown associations between various post-operative complications such as SSI.

Widespread adoption of Health Information Systems (HIS) including Electronic Health Records (EHR) and laboratory Information Systems has generated massive amounts of both quantitative data (e.g., laboratory values) and qualitative data (e.g., text based clinical notes), partly fueled by funding, by the Health Information Technology for Economic and Clinical Health Act of 2009. It is but inevitable that the big data management and analytics that are successfully applied in retail sales and other domains will be researched in healthcare (Murdoch & Detsky, 2013). Recent advances in computational machine learning methods and predictive modeling could be applied to analyze the information content of those lab tests and draw valuable knowledge from any associations with SSI (Sohn et al., 2017). But, clinical laboratory data (and most healthcare data) is inherent with various constraints that need to be addressed and pre-processed before applying machine learning algorithms. The sampling of lab tests is purely based on the clinical condition of the patient, as assessed by the providers, and thus the temporality of sampling is extremely variable in time/ space as well as the type of tests done. Some tests are usually done just once (e.g., ABO blood grouping) and some others are repeated multiple times (e.g., glucose, electrolytes, hemoglobin). Other than in research settings, only a core set of lab tests are done routinely on most patients during day to day clinical care and a subset of patients may be tested for disease specific lab biomarkers. This variability of sampling, both the type/ number of tests as well as temporality (time and space) contrasts with the statistical and predictive analytic tools which very much rely on systematic and regularly sampled data.

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/predictive-modeling-of-surgical-site-infections-using-sparse-laboratory-data/243123

Related Content

Loan Fraud Detection Using Machine Learning as a Data Mining Approach
Nabila Hamdoun (2022). *International Journal of Data Analytics (pp. 1-10).*
www.irma-international.org/article/loan-fraud-detection-using-machine-learning-as-a-data-mining-approach/309096

Fitting a Three-Phase Discrete SIR Model to New Coronavirus Cases in New York State
Kris H. Green (2021). *International Journal of Data Analytics (pp. 59-74).*
www.irma-international.org/article/fitting-a-three-phase-discrete-sir-model-to-new-coronavirus-cases-in-new-york-state/285468

Use of Artificial Intelligence in the Employee Selection Process
Filiz Mzrak (2023). *HR Analytics in an Era of Rapid Automation (pp. 230-248).*
www.irma-international.org/chapter/use-of-artificial-intelligence-in-the-employee-selection-process/327757

The Impact of Creating a Business Intelligence Platform on Higher Education: The Case of the American University in Cairo
Sherif H. Kamel, Iman Megahedand Heba Atteya (2019). *Managerial Perspectives on Intelligent Big Data Analytics (pp. 232-259).*
www.irma-international.org/chapter/the-impact-of-creating-a-business-intelligence-platform-on-higher-education/224342

A Fuzzy Measure of Vulnerability for the Optimization of Vehicle Routing Problems With Time Windows
Kris Braekersand Gerrit K. Janssens (2018). *Optimization Techniques for Problem Solving in Uncertainty (pp. 167-197).*
www.irma-international.org/chapter/a-fuzzy-measure-of-vulnerability-for-the-optimization-of-vehicle-routing-problems-with-time-windows/206634