'Big data' helps delivery of smart healthcare in intensive care units

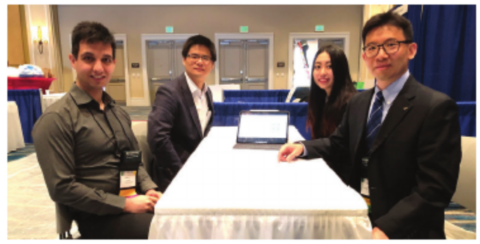
Intensive care units (ICUs) house the most seriously ill patients in the hospital. The annual admission to ICUs in the U.S. is more than 5.7 million with the projected cost of $81.7 billion. To cope with disease complexity and enhance information visibility, advanced sensing is increasingly invested and leads to data-rich environments in ICU.

For example, modern ICUs require the monitoring of heterogeneous types of clinical variables including laboratory tests (urea, creatinine, sodium and potassium), heart rate and rhythm, blood pressure, respiratory rate, blood-oxygen saturation and so on.

As such, healthcare practitioners are facing "big data" each day which is underutilized to improve the quality of healthcare delivery. Physicians often make medical decisions based on the most recent data while overlooking the historical data and the hidden relationships among the heterogeneous set of clinical variables.

In addition, as human discretion dominates the collection of clinical data, missing data are a general problem in the ICU environment. There is an urgent need to go beyond current clinical practices and further develop advance data-driven methods to enable the delivery of smart healthcare in ICUs.

In "Nested Gaussian Process Modeling for High-dimensional Data Imputation in Healthcare Systems," Ph.D. students Farhad Imani and Ruimin Chen and professor Hui Yang of The Pennsylvania State University and professor Changqing Cheng of Binghamton University aim to handle the patient heterogeneity, time asynchronization and variable heterogeneity in the big ICU data. They developed a new nested Gaussian process approach to account for correlation across time stamps, similarity among patients and closeness of variables for the prediction of missing ICU data and the dynamics of patient recovery process.



From left, Farhad Imani, Changqing Cheng, Ruimin Chen and Hui Yang take part in the group discussion.

Such analytical methods are timely in helping healthcare practitioners leverage the increasing availability of big data to achieve a substantial boost in smart ICU health management. This research is supported by a National Science Foundation sponsored project "Sensing, Modeling and Optimization of Postoperative Heart Health Management," for which Yang is the principal investigator.

CONTACT: Hui Yang; *huy25@psu.edu*; (814) 865-7397; Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, 310 Leonhard Building, University Park, PA 16802

Combining AI with human expertise to find rare causes of injuries

Injuries are a major public health problem that annually cause about 5 million fatalities and many more disabilities worldwide. To design interventions that prevent these injuries, we first need a more comprehensive understanding of circumstances surrounding the injuries. Many organizations conducting injury surveillance routinely collect short narratives describing the incident transcribed by triage nurses in hospital emergency rooms or other settings. Some also assign external cause of injury codes (E-codes) based on the international classification of diseases.

While few injury mechanisms and E-codes are frequent, such as transport-related incidents and falls, there are several rarer causes of injuries such as suffocation and heat exhaustion.

E-coded injury data are helpful in analyzing causation patterns and prioritizing prevention efforts. However, the manual coding of injury data are time-consuming and suffers from consistency and accuracy issues. Machine learning (ML) models trained on coded injury data can be used to automatically assign E-codes based on the injury narrative but their prediction accuracy is severely low.

Gaurav Nanda and Mark Lehto

Researchers from Purdue University, including postdoctoral researcher Gaurav Nanda, former Ph.D. student Hsin-Ying Huang and professor Mark Lehto, and Queensland University of Technology professor Kirsten Vallmuur, have been developing ML-based intelligent decision support systems that can identify injury cases for which E-codes can be accurately predicted using ML models and filter the remaining cases to be reviewed by human coders.

In "Semi-automated Text Mining Strategies for Identifying Rare Causes of Injuries from Emergency Room Triage Data," the team investigated various approaches of efficiently filtering out cases likely to belong to rare E-codes based on predictions of logistic regression and naive Bayes ML models. They examined the performance of these filtering approaches on a manually coded emergency department triage dataset of about a half million cases provided by the Queensland Injury Surveillance Unit. They found expert-designed linguistic rules combined with logistic regression prediction strength to be the most efficient; other automated filtering approaches that performed well included debiased sampling of training data and stagewise hierarchical classification.

With efficient filtering, such intelligent decision support systems can effectively utilize the strengths of artificial