

## Guest Editorial Preface

# Special Issue on Knowledge Discovery in Biomedicine

Samah Jamal Fodeh, Yale University, New Haven, USA

Knowledge discovery is the process of discovering new knowledge in large volumes of data. Knowledge discovery methodologies have been widely and successfully adopted in the field of biomedicine to mine and harness big biomedical data. In this special issue, we invited papers concerning the application of knowledge discovery approaches such as classification and prediction, text mining, and information retrieval to biomedical and health related data. The International Journal of Knowledge Discovery in Bioinformatics (IJKDB) is particularly interested in publishing methodological reviews on topics that introduce methodological innovations. At IJKDB, the published articles present topics such as systems biology, protein structure, gene expression, and biological data integration. IJKDB journal presents a cross-disciplinary platform that bridges many sciences for better outcomes.

## INSIDE THIS SPECIAL ISSUE

I invited papers submitted to both the fifth Data Mining in Biomedical Informatics and Healthcare (DMBIH'05) Workshop held in conjunction with the IEEE International Conference on Data Mining (ICDM'17) and the first Medical Informatics and Healthcare (MIH) Workshop held with the 23rd ACM SIGKDD Conference of Knowledge Discovery and Data Mining to be considered for publication. Three submissions were selected to be included in the final special issue. In this preface, I provide a snapshot of the research described in the special issue.

In the first article, authors investigated hospitalizations and prolonged length-of-stay in hospitals and the associated cost for providers, insurers, and patients. They proposed a prediction tool of patient annual length-of-stay in hospitals. This tool is important for resource allocation and improving patient health outcomes. Unlike previous studies that mainly incorporated clinical information about the patients, their tool predicts annual length-of-stay using information that is symmetrical between insurers, providers, and the government. They used machine learning approaches including boosted trees, random forests, and artificial neural networks for prediction.

Results suggest that the Colombia's contributory health care system would achieve significant cost savings if insurers implemented prevention programs based on predictive modeling with efficacies of more than 40% and for any intervention cost between 100,000 and 700,000 pesos per patient.

In the second article, the topic is concerned with augmenting decision making by mining text-in radiology teaching files. The authors built on their Integrated Radiology Image Search (IRIS 1.1) search engine and added two components: a new medical ontology called SNOMED\_CT and the ICD10 dictionary. IRIS 1.1 integrates public data sources and applies query expansion with exact and partial matches to find relevant teaching files. Using a set of 28 representative queries from multiple sources, their search engine finds more relevant teaching cases versus other publicly available search engines.

In the third article, the authors leveraged social media data, namely Twitter, to better detect high risk suicide behavior(s). They first evaluated the presence and density of language related to twelve suicide risk factors on Twitter then they analyzed the relationships between suicide risk factors. They have shown that language patterns of the majority of suicide risk factors have strong presence on Twitter such as “depressive symptoms”, “drug abuse” and “prior suicide attempts”. Certain linguistics patterns pertaining to suicide risk factors are more frequently used by those who are at higher risk of suicide. They also concluded that tweets may be highly reflective of emotional and behavioral attributes of users and could be a valuable resource for predicting suicide.

The variety of the topics addressed in the articles published in this special issue shed light on the utility of using knowledge discovery approaches in biomedicine. Looking at this from a different angle, I conclude that incorporating data scientist with skills in knowledge discovery and data mining in biomedicine and healthcare is becoming a must to cope with the explosive growth of digital biomedical and health related data as well as social media data.

*Samah Jamal Fodeh*

*Guest Editor*

*IJKDB*