Foreword

A pressure of competitiveness is growing in the healthcare sector while quality demands are rising. These trends confront medical institutions more than ever with the necessities of critically reviewing their own efficiency and quality under both medical and economical aspects. At the same time, growing capture of medical data and integration of distributed and heterogeneous databases create a completely new environment for medical quality and cost management. These advances have come to the foreground with the vast amounts of biomedical and genomic data in an electronic form, the Internet ability to transfer the data efficiently, and the wide application of computer use in all aspects of medical, biological, and health care research and practice. In medical and health care areas, due to regulations, a large amount of data is becoming available, and practitioners are expected to use these data in their everyday work. But such a large amount of data cannot be processed by humans in a short time to support timely and accurate medical decisions. Medical datasets have reached enormous capacities. Just as an illustration, note that almost a billion persons living in North America and Europe have at least some of their medical information collected in electronic form, at least transiently. These data may contain valuable information that awaits extraction, and a new knowledge may be encapsulated in various patterns and regularities that were hidden in the data. Such knowledge may prove to be priceless in future medical decision making.

This book presents several case studies of medical data mining developed by faculty and graduates of the University of Louisville’s PhD program in Applied and Industrial Mathematics. Each contributor provides valued insights in the utilization of data mining technology supporting various aspects of a data mining process (e.g. data collection, reduction, cleansing and data integration). The book is organized in three main sections covering physician decision making, healthcare delivery, and medical data modeling. The studies include data analyses of the treatment of osteomyelitis, cardiovascular bypass versus angioplasty, the treatment of asthma, and both lung and breast cancer. Generally, the cases in this book use datasets that are publicly available for the purpose of research including the National Inpatient Sample and the Medical Expenditure Panel Survey, while data mining techniques used include market basket analysis, predictive modeling, time series analysis, survival data mining, kernel density estimation, and text mining.

This book describes applications of data mining in a very accessible form: both as a text to give ideas to the next generation of data mining practitioners, and to inform lifelong learners about potentials of data mining technology in the medical field. No chapter in this book will give you the full and final explanation about how to solve problems with data mining technology; every chapter will give you insights into how unpredictable and messy is a data mining process. Medical data mining must not be
regarded as an independent effort, but it should rather be integrated into the broader environment that is aligned with decision-making processes in the medical field. The collection of papers illustrates the importance of maintaining close contact between data mining practitioners and the medical community in order to keep a permanent dialogue in order to identify new opportunities for applications of existing data mining technologies.

Mehmed Kantardzic  
Louisville, February 10, 2009

Mehmed Kantardzic received B.S., M.S., and Ph.D. degrees in computer science from University of Sarajevo, Bosnia. Currently, he is a Professor at the Computer Engineering and Computer Science Department, University of Louisville. He is also Director of CECS Graduate Programs, Director of Data Mining Lab, and Co-Director of Data Mining Graduate Certificate Program. His research interests are: data mining & knowledge discovery, machine learning, soft computing, click fraud detection, text mining & link analysis, sequence mining, and distributed intelligent systems. His recent research projects are supported by NSF, KSTC, US Treasury Department, and NASA. Dr. Kantardzic published more than 200 articles in refereed journals and conference proceedings including two books on data mining and applications. His book on data mining, published by John Wiley, is accepted as a textbook at universities all over the world. Dr. Kantardzic is a member of IEEE, IEE, ISCA, WSEAS, and SPIA. He is on editorial board or guest editor for several international journals, NSF and NSERC panel member and reviewer, program or conference chair for several international conferences and ICMLA Steering Committee member.