

Preface

In the past decade, we have witnessed tremendous growth in biomedical data generation and substantial improvement of computational capacities (both hardware and computational methods) that can be used to handle these data. As a result, these “Big Data” provide great opportunities to health informatics and healthcare in general. In particular, the available data and the data-driven approach have started to empower precision medicine, which provides personalized disease treatment and prevention by taking into account individual variability in genes, environment, and lifestyle. On the other hand, the huge amount of the data and how to use these data raise unparalleled challenges to data scientists and informatics researchers. It is highly nontrivial to provide useful computer-aided analyses of heterogeneous biomedical datasets accumulated in various databases and electronic health records (EHRs). The biomedical data are notorious for its diversified scales, dimensions and volumes, and require interdisciplinary technologies for visual illustration and digital characterization. Various computer programs and servers have been developed for these purposes. But how to choose and use them are often difficult, especially for beginners. In addition, integrating different data and tools together to assist medical diagnosis and treatment is even more challenging.

A number of edited books have been published to discuss different aspects of health informatics data analysis. However, these books typically focus more on individual research. The authors of each chapter often emphasize their own methods. There is lack of comprehensive overview for the field, and hence the existing books are often difficult for beginners. This book is an attempt to systematically review the computational methods and tools for different aspects of health informatics data analyses. We have designed this handbook to comprehensively cover major topics in the field, as well as to provide concrete examples. Each chapter provides the detailed review of the state-of-the-art computer programs and an example procedure of data analysis and data fusion for each of 13 important biomedical questions. By following the step-by-step procedure, you will be exploring the biomedical questions with various programs and servers like a pro. Each chapter in the book is a self-contained review of a specific topic. Hence, a

reader does not need to read through the chapters sequentially. A brief description of each chapter is given below.

Chapter “[ECG Annotation and Diagnosis Classification Techniques](#)” reviews the general techniques of ECG beat annotation and classification. It shows a preliminary study on deep learning application in ECG classification, which leads to better results and has a high potential both for performance improvement and unsupervised learning applications.

Chapter “[EEG Visualization and Analysis Techniques](#)” presents the current status of EEG research with projected applications in the areas of health care. As an example, it describes a method of quick prototyping an EEG headset in a cost-effective way and with state-of-the-art technologies.

Chapter “[Biomedical Imaging Informatics for Diagnostic Imaging Marker Selection](#)” discusses challenges and techniques of biomedical imaging informatics in the context of imaging marker extraction. In particular, it focuses on how to regulate image quality, extract image features, select useful features, and validate them.

Chapter “[Big Health Data Mining](#)” demonstrates different data levels involved in health informatics and introduces some general data mining approaches. An example case study is illustrated for mining long-term EHR data in epidemiological studies.

Chapter “[Computational Infrastructure for Telehealth](#)” introduces telehealth systems and their computational architecture, as well as challenges associated with creation of the ‘complete-loop’ solution. It also includes a practical use case describing an application for monitoring patients with hypertension.

Chapter “[Healthcare Data Mining, Association Rule Mining, and Applications](#)” introduces popular data mining algorithms and their applications in health care. It focuses on association rule mining that can provide a more flexible solution for personalized and evidence-based clinical decision support.

Chapter “[Computational Methods for Mass Spectrometry Imaging: Challenges, Progress, and Opportunities](#)” examines current and emerging methods for analysis of mass spectrometry imaging (MSI) data. It highlights associated challenges and opportunities in computational research for MSI, especially in proteomics, lipidomics, and metabolomics with spatially resolved molecular information.

Chapter “[Identification and Functional Annotation of lncRNAs in Human Disease](#)” describes the current bioinformatics methods to identify long noncoding RNAs (lncRNAs) and annotate their functions in mammal. It also provides several ways to further analyze the interactions between lncRNAs and targets, such as miRNAs and protein coding genes.

Chapter “[Metabolomics Characterization of Human Diseases](#)” summarizes popular bioinformatics analysis tools for characterizing human diseases based on their metabolomics profiles. Pathway analysis using metabolite profiles and disease classification using metabolite biomarkers are presented as two examples.

Chapter “[Metagenomics for Monitoring Environmental Biodiversity: Challenges, Progress, and Opportunities](#)” gives an overview of metagenomics, with particular emphasis on the steps involved in a typical sequence-based

metagenome project. It describes and discusses sample processing, sequencing technology, assembly, binning, annotation, experimental design, statistical analysis, and data storage and sharing.

Chapter “[Global Nonlinear Fitness Function for Protein Structures](#)” examines the problem of constructing fitness landscape of proteins for generating amino acid sequences that would fold into a structural fold for protein sequence design. It introduces two geometric views and proposes a formulation using mixture of nonlinear Gaussian kernel functions.

Chapter “[Clinical Assessment of Disease Risk Factors Using SNP Data and Bayesian Methods](#)” reviews new statistical methods based on Bayesian modeling, Bayesian variable partitioning, and Bayesian graphs and networks. As an example, it outlines how to use Bayesian approaches in clinical applications to perform epistasis analysis while accounting for the block-type genome structure.

Chapter “[Imaging Genetics: Information Fusion and Association Techniques between Biomedical Images and Genetic Factors](#)” covers recent studies of correlative and association analysis of medical imaging data and high-throughput genomic data. It also provides an example of parallel independent component analysis in an imaging genetic study of schizophrenia.

We have selected these topics carefully so that the book would be useful to a broad readership, including students, postdoctoral fellows, faculty and professional practitioners in bioinformatics, medical informatics, and other biomedical studies. We expect that the book can be used as a reference for upper undergraduate-level or beginning graduate-level bioinformatics/medical informatics courses.

We would like to thank the chapter authors for their excellent contributions to the book. We also would like to thank all the reviewers for their helpful comments and suggestions. This book would not have been possible without the professional support from Springer International Publishing AG, Cham.

Columbia, USA
Atlanta, USA
Changchun, China
Shenzhen, China

Dong Xu
May D. Wang
Fengfeng Zhou
Yunpeng Cai

Health Informatics Data Analysis

Methods and Examples

Xu, D.; Wang, M.D.; Zhou, F.; Cai, Y. (Eds.)

2017, X, 210 p. 54 illus., Hardcover

ISBN: 978-3-319-44979-1