Integration of Clinical and Genomic Data for Decision Support in Cancer

Yorgos Goletsis
University of Ioannina, Greece

Themis P. Exarchos
University of Ioannina, Greece

Nikolaos Giannakeas
University of Ioannina, Greece

Markos G. Tsipouras
University of Ioannina, Greece

Dimitrios I. Fotiadis
University of Ioannina, Greece
Michaelideion Cardiology Center, Greece & Biomedical Research Institute, Greece

INTRODUCTION

Computer aided medical diagnosis is one of the most important research fields in biomedical engineering. Most of the efforts made focus on diagnosis based on clinical features. The latest breakthroughs of the technology in the biomolecular sciences are a direct cause of the explosive growth of biological data available to the scientific community. New technologies allow for high volume affordable production and collection of information on biological sequences, gene expression levels and proteins structure on almost every aspect of the molecular architecture of living organisms. For this reason, bioinformatics is asked to provide tools for biological information processing, representing today's key in understanding the molecular basis of physiological and pathological genotypes. The exploitation of bioinformatics for medical diagnosis appears as an emerging field for the integration of clinical and genomic features, maximizing the information regarding the patient's health status and the quality of the computer aided diagnosis.

Cancer is one of the prominent domains where this integration is expected to bring significant achievements. As genetic features play a significant role in the metabolism and the function of the cells, the integration of genetic information (proteomics-genomics) to cancer-related decision support is now perceived by many not as a future trend but rather as a demanding need. The usual patient management in cancer treatment involves several, usually iterative, steps consisting of diagnosis, staging, treatment selection, and prognosis. As the patient is usually asked to perform new examinations, diagnosis and staging status can change over time. On the other hand, treatment selection and prognosis depend on the available findings, response to previous treatment plan and, of course, clinical guidelines. The integration of these evolving and changing data into clinical decision is a hard task which makes fully personalised treatment plan almost impossible. The use of clinical decision support systems (CDSSs) can assist in the processing of the available information and provide accurate staging, personalised treatment selection, and prognosis. The development of electronic patient records and of technologies that produce and collect biological information have led to a plethora of data characterizing a specific patient. Although this might seem beneficial, it can lead to confusion and weakness concerning the data management. The integration of the patient data (quantitative) that are hard to be processed by a human decision maker (the clinician) further imposes the use of CDSSs in personalized medical care (Louie, Mork, Martin-Sanchez, Halevy, & Tarczy-Hornoch, 2007). The future vision—but current need—will not include generic treatment plans according to some naive reasoning, but totally personalised treatment based on the clinicogenomic profile of the patient.

In this article, we address decision support for cancer by exploiting clinical data and identifying mutations

on tumour suppressor genes. The goal is to perform data integration between medicine and molecular biology by developing a framework where clinical and genomic features are appropriately combined in order to handle cancer diseases. The constitution of such a decision support system is based on (a) cancer clinical data and (b) biological information that is derived from genomic sources. Through this integration, real time conclusions can be drawn for early diagnosis, staging and more effective cancer treatment.

BACKGROUND

Clinical Decision Support Systems are active knowledge systems which use two or more items of patient data to generate case-specific advice (Fotiadis, Goletsis, Likas, & Papadopoulos, 2006). CDSSs are used to enhance diagnostic efforts and include computer based programs that, based on information entered by the clinician, provide extensive differential diagnosis, staging (if possible), treatment, follow-up, and so forth. CDSSs consist of an inference engine that is used to associate the input variables with the target outcome. This inference engine can be developed based either on explicit medical knowledge, expressed in a set of rules (knowledge based systems) or on data driven techniques, such as machine learning (Mitchel, 2006) and data mining (intelligent systems) (Tan, Steinbach, & Kumar, 2005). CDSSs require the input of patient-specific clinical variables (medical data) and as a result provide patient specific recommendation.

Medical data are observations regarding a patient, including demographic details (i.e., age, sex), medical history (i.e., diabetes, obesity), laboratory examinations (e.g., creatinine, triglyceride), biomedical signals (ECG, EMG), medical images (i.e., MRI, CT), and so forth. Demographic details, medical history, and laboratory data are the most easily obtained and recorded and, therefore, most commonly included in electronic patient records. On the other hand, biomedical signals and medical images require more effort in order to be acquired in a digital format and must be processed for useful feature extraction. Apart from these, several types of genomic data can be generated from laboratory examinations, that is, gene DNA or protein sequences, gene expression data, microarray images, and so forth. Genomic data can also be used for medical diagnosis, disease prevention, and population genetics studies. Although medical data are sufficient for the diagnosis of several diseases, recent studies have demonstrated the high information value of genomic data, especially in specific types of diseases, such as cancer diseases.

The great amount and the complexity of the available genetic data complicates their analysis from conventional data analysis methods and requires higher order analysis methods such as data mining techniques. Lately, data mining has received much attention in bioinformatics and molecular biology (Cook, Lawrence, Su, Maglothin, & Jonyer, 2001). Data mining methods are usually applied in the analysis of data coming from DNA microarrays or mass spectrometry. Over the last few years, several scientific reports have shown the potential of data mining to infer clinically relevant models from molecular data and thus provide clinical decision support. The majority of papers published in the area of data mining for genomic medicine deals with the analysis of gene expression data coming from DNA microarrays (Jiang & Gruenwald, 2005; Shah & Kusiak, 2007; Walker et al., 2004) consisting of thousands of genes for each patient, with the aim to diagnose types of diseases and to obtain a prognosis which may lead to therapeutic decisions. Most of the research works are related to oncology (Louie et al., 2007), where there is a strong need for defining individualized therapeutic strategies. Another area where data mining has been applied is the analysis of genes or proteins, represented as sequences (Exarchos, Papaloukas, Lampros, & Fotiadis, 2006); sequential pattern mining techniques are suitable for analyzing these types of data (Zaki, 2000).

Several CDSSs for cancer have been proposed in the literature. Most of the approaches are based solely on clinical data and a few methods exist that provide cancer decision support using microarray gene expression data. The cancer CDSSs concern several different types of cancer and employ various techniques for their development. The majority of systems are still in a research level and only a few are being used in clinical practice. A CDSS which is already in clinical use is PAPNET (Boon & Kok, 2001) which deals with cervical cancer. PAPNET uses ANNs to extract abnormal cell appearances from vaginal smear slides and describe them in histological terms. Other CDSSs for cervical cancer concentrate on the evaluation of the benefits of the PAPNET system (Doornewaard, 1999; Nieminen, Hakama, Viikki, Tarkkanen, & Anttila, 2003). Colon cancer has also been studied using clinical data and fuzzy classification trees (Chiang, Shieh, Hsu, & Wong, 2005) or pattern analysis of gene expression levels

7 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/integration-clinical-genomic-data-decision/13011

Related Content

Promoting Mobile Health Adoption to Hospital Patients Through Social Influencers: A Multi-Group Analysis Among Patients With High vs. Low Hospital Usage
Daisy Leeand Gary C.H. Chong (2021). *International Journal of Healthcare Information Systems and Informatics (pp. 1-15).*
www.irma-international.org/article/promoting-mobile-health-adoption-to-hospital-patients-through-social-influencers/279328

Neuroeconomic Psychology: 3 Cognitive Training Modules for End-Users
Torben Larsen (2019). *International Journal of Patient-Centered Healthcare (pp. 1-17).*
www.irma-international.org/article/neuroeconomic-psychology/250227

Innovation in ICT-Based Health Care Provision
Synnøve Thomassen Andersenand Arild Jansen (2011). *International Journal of Healthcare Information Systems and Informatics (pp. 14-27).*
www.irma-international.org/article/innovation-ict-based-health-care/53477

From Idea to Use: Lessons Learned From a Participatory ICT Healthcare Case Study
Henrik Enquist (2013). *User-Driven Healthcare: Concepts, Methodologies, Tools, and Applications (pp. 1037-1053).*
www.irma-international.org/chapter/idea-use-lessons-learned-participatory/73878

The Effect of Clustering in Filter Method Results Applied in Medical Datasets
Nadjla Elongand Sidi Ahmed Rahal (2021). *International Journal of Healthcare Information Systems and Informatics (pp. 38-57).*
www.irma-international.org/article/the-effect-of-clustering-in-filter-method-results-applied-in-medical-datasets/267883