

Nanoinformatics for biomedicine: emerging approaches and applications

Xiong Liu¹
Thomas J Webster²

¹Intelligent Automation, Inc,
Rockville, MD, USA; ²Department
of Chemical Engineering, Northeastern
University, Boston, MA, USA

Abstract: This special issue on nanoinformatics for biomedicine is a collection of recent papers from the 2012 IEEE Workshop on Nanoinformatics for Biomedicine (NanoInfo 2012) and other work in the area. These papers illustrate different aspects of nanoinformatics to support biomedical research and to advance knowledge on nanomaterial–biological interactions. The topics covered include data curation, data standards, data mining and predictive modeling, machine learning, and translational research. The objectives of this special issue are multifold: (1) to bring together and showcase some of the latest research results in the field; (2) to introduce some useful repositories, systems, and analysis tools; and (3) to stimulate more research activities in the field.

Keywords: nanobiotechnology, nanoinformatics, data curation, data mining, machine learning, translational research

Introduction

The application of nanotechnology in biomedicine offers innovative approaches to disease diagnosis, treatment and prevention. During this process, nanoinformatics has emerged as a discipline that provides informational foundations for nanomedicine. To address the ever-growing information needs of nanomedicine, nanoinformatics is playing an important role in acquiring relevant data and information, conceptually organizing information for conversion to knowledge, and developing computing applications that are based on important parameters at the nanoscale and are applicable in a broad health perspective.

Recently, the field of nanoinformatics has witnessed a steady growth in research activities and scientific achievements. More than 35 important events have taken place over the past 5 years.¹ One of the most recent events is the 2012 IEEE Workshop on Nanoinformatics for Biomedicine (NanoInfo 2012).² NanoInfo 2012 was held on October 4, 2012 in Philadelphia, PA, in conjunction with the 2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). This workshop aimed to provide a forum for discussion on all the aspects of nanoinformatics to support biomedical research and advance knowledge on nanomaterial–biological interactions (NBIs). The workshop presented nine papers with a particular emphasis on data collection and curation, interoperability and data standards, computational or information technologies to solve specific problems, and translational research. Selected papers from the workshop and other related work in the area are included in this special issue.

Correspondence: Xiong Liu
15400 Calhoun Drive, Suite 400,
Rockville, MD 20855, USA
Tel +1 301 294 4629
Fax +1 301 294 5201
Email xliu09@gmail.com

The objectives of this special issue are multifold: (1) to bring together and showcase some of the latest research results in the field; (2) to introduce some useful repositories, systems, and analysis tools; and (3) to stimulate more research activities in the field. In this editorial, we first discuss some of the current topics in nanoinformatics. We then introduce the papers in this special issue.

Research topics

The present issue covers various nanoinformatics approaches to enhance knowledge for biomedicine. A sample list of significant topics follows.

New data standards and data models

The past decade of nanotechnology has produced an explosion of knowledge, with a large amount of heterogeneous information about various types of nanomaterials. To support cooperative research, there are needs for standard protocols, standard data formats, controlled vocabularies, and appropriate data models. Efforts to establish nanotechnology data standards include ASTM nanotechnology standards,³ ISO standards on nanotechnologies,⁴ ontologies for cancer nanotechnology research,⁵ and minimal information standards on nanomaterial characterization.⁶ In this special issue, the paper by Mills et al discusses the minimal information that currently exists about such nanomaterials criteria, which is being adopted community-wide. In addition, the paper by Tang et al introduces a new integrated data model, which describes the complex data structure for nanomaterial environmental impact-related bibliography and characterization data.

Data mining for nanomedical research

Data mining can assist in the discovery of knowledge from large amounts of bio- and nanomedical data. From the perspective of machine learning, data mining problems can be categorized into two main groups: supervised learning and unsupervised learning. Supervised learning can be further categorized into two subgroups, namely classification and numerical prediction. In a classification problem, the class label is discrete. While in a numerical prediction problem, the class label is continuous. A representative supervised learning is predicting the toxicology of nanomaterials.⁷ Classification methods are used when toxicity is defined in a categorical manner (toxic or nontoxic), and numerical prediction methods are used when toxicity is defined as a numerical attribute. Clustering is a typical problem of unsupervised learning. For example, the K-means clustering can

be applied to localize the nanoparticles within the cellular environment, based on their characteristic spectroscopic signatures.⁸ Well trained data-mining models provide predictive power. They could be used to forecast future results or risks.

Several papers in this special issue fall in this category. Liu et al discuss the predictive modeling of nanomaterial biological effects (eg, mortality, malformation, growth inhibition) using numerical prediction algorithms such as regression tree and instance-based learner; Sayes et al study the grouping of colloidal metal nanoparticles based on their physicochemical characteristics, using linear discriminant analysis classification; Kim and Bredel focus on predicting survival in cancer patients using gene expression profiles and algorithms such as nearest neighbor and Cox-based regression; and Paul and Maji apply rough sets and support vector machine to classify human cancers (eg, lung tumor) based on microRNA (miRNA) expression data.

To discover high-quality data-mining models, Tang et al propose a meta-optimization strategy to select most appropriate algorithms and tune algorithm parameters. The goal is to discover an optimal data-mining model with the best performance (eg, highest prediction accuracy) for a given dataset or scenario. The optimization strategy not only searches for the best algorithm from a list of applicable algorithms, but also finds the parameter settings that will result in optimal performance for a specific learning algorithm.

Novel information systems and tools

Nanoinformatics encourages innovative methods for collecting, storing, processing, and sharing nanomedical data. In this special issue, several innovative systems and tools have been proposed. Mills et al describe the Nanomaterial Registry (NR) Web-based system for data sharing and analysis. Unique features of the NR system include the minimal information about nanomaterials, the compliance levels that measure the quality and quantity of characterization of a nanomaterial entry, side-by-side comparison of nanomaterial characterizations, and rule-based nanomaterial matching tool. Tang et al present the NEIMiner data-mining system for sharing and mining nanomaterial impact data. Key features of NEIMiner include the model-driven data-mining system based on a nanomaterial impact modeling framework, integrated and collaborative data management by leveraging the Drupal content management system,⁹ optimized data-mining model discovery, model composition to form comprehensive impact/risk predictability, and the model base for managing and exposing data-mining

models. In addition, Liu et al show an example of the model query interface for obtaining predictive information from the underlying models.

NBIs

The study of NBIs is an emerging area which investigates nanomaterial exposure effects (beneficial, benign, or deleterious) in biological systems. NBIs can be defined at multiple levels of biological organization, such as molecular, cellular, and organismal level. The NBI Knowledgebase developed at Oregon State University is a repository of annotated data on NBIs as well as nanomaterial characterization, synthesis methods, and exposure scenarios.¹⁰ As more NBI data are accumulated, it is critical to adopt computational and data-mining methods to provide a logical framework for extrapolations based on experimental data across species, doses, exposure routes, and scenarios. In this special issue, Tang et al introduce the NEIMiner data-mining system for studying nanomaterial environmental and biological impact, and Liu et al discuss the modeling of various nanomaterial exposure effects.

Translational research

Translational nanoinformatics aims to translate nano-level or atomic-level measurements and data into new knowledge that can lead to high quality diagnosis and therapy for patients. Nanoinformatics faces unique challenges such as the polymorphic and polydisperse nature of nanomaterials. These challenges require new approaches for data and knowledge integration at the nanoscale. Expert annotation and curation of data are needed to ensure the quality of the data and analysis methods used in nanomedicine.¹¹

Nanoinformatics also addresses computing applications that process omics and disease information that are essential to translational research. In this special issue, three papers fall in this category. Kim and Bredel present a study on predicting cancer survival using gene expression profiles. This study demonstrates that using preexisting biological knowledge (cancer pathway genes) for survival prediction has significant advantages over other approaches that do not leverage prior knowledge. Paul and Maji describe a study on selecting differentially expressed miRNAs and predicting cancers using miRNA expression data. This study shows that rough sets are a useful approach to select significant miRNAs for cancer prediction. In addition, Ercan et al present a novel application of protein analysis related to tissue growth on nanomaterials.

Other topics

Other research topics in nanoinformatics include text and literature mining for useful knowledge, modeling quantitative structure–activity relationships, quantitative modeling in toxicology, molecular modeling and simulation techniques, imaging techniques for in vivo cancer diagnosis, clinical trials, risk assessment and regulation, surveillance systems, ethical and social issues, and others. Refer to Maojo et al's review¹¹ on nanoinformatics areas.

Papers in this special issue

In this issue, Mills et al discuss their work on the Nanomaterial Registry, which has been developed to address challenges such as the needs for standard methods, data formatting, and controlled vocabularies for data sharing. The Registry is meant to be an authoritative resource, where the data is systematically curated and reviewed by subject matter experts. The Registry's data model is based on a set of minimal information about nanomaterials (MIAN). The MIAN has been used to create evaluation and similarity criteria for nanomaterials that are curated into the Registry. Currently, the Registry supports search, browse, side-by-side comparison of nanomaterials, compliance ratings based on the quality and quantity of data, and the ability to search for similar nanomaterials.

Tang et al discuss the NEIMiner data-mining system for studying the nanomaterial environmental impact (NEI). NEIMiner first aggregates data from heterogeneous data sources that are mixed with bibliographic texts and structured experimental data. It then uses a Web content management system to store and manage NEI-related bibliography and characterization data. Further, NEIMiner provides advanced data-mining modules to automate the process of deriving, sharing, and visualizing high-quality NEI prediction models from experimental data. Using NEIMiner, researchers and policymakers can obtain scientific information about the environmental fate of nanomaterials, their availability to receptor organisms (eg, uptake), and any resultant biological effects (eg, toxicity).

Liu et al give detailed examples of how to model the unintended biological effects resulting from the application of nanomaterials. The scope of modeling covers nanomaterial properties, exposure and study scenarios, biological responses, and their interactions. The modeling approach is based on data-mining techniques such as numerical prediction algorithms. Using an experimental dataset on the toxicity of nanomaterials to embryonic zebrafish, results show that high prediction accuracy can be obtained for predicting biological effects such as 24 hours post-fertilization (hpf)

mortality, 120 hpf mortality, and 120 hpf heart malformation. To make the prediction models more generally available, a model base component is developed to expose the models and associated datasets. The model base provides a what-if query interface to inform the nanomaterial design process for benign materials and processes.

Sayes et al present a study on the utility of specific statistical and pattern recognition techniques in grouping nanoparticles based on experimental data about their physicochemical properties. The goal of the study is to show that a set of selected physicochemical properties of metal-based colloidal materials could be used as nanodescriptors that facilitate the grouping of these metal-based colloids. The results of the analyses suggest that a seemingly homogeneous group of nanoparticles could be separated into subgroups depending on interdependencies observed in their nanodescriptors. This study is an example of improved nanomaterial characterization, which facilitates the study of the relationship between the physicochemical properties of nanomaterials and their induced toxicological responses.

Kim and Bredel discuss their work on predicting survival in cancer patients using cancer pathway gene expression profiles. They studied four different survival prediction methods, including 1-nearest neighbor survival prediction method, random patient selection method, Cox-based regression method with whole gene expression profiles, and Cox-based regression method with the gene expression profiles of cancer pathway genes. The results show that the Cox-based regression method using the gene expression profiles of cancer pathway genes achieved the best performance. This study demonstrates that using preexisting biological knowledge for survival prediction is a promising approach. Indeed, personalized medicine is usually based on known subcategories of a disease for better treatment.

Paul and Maji discuss a novel approach for insilico identification of differentially expressed miRNAs from microarray expression datasets. MiRNAs play an important role in various cellular processes and thus help in carrying out normal functioning of a cell. Various studies have also shown the role of miRNAs in cancer and utility of miRNAs for the diagnosis of cancer and other diseases. This study integrates the theory of rough sets and the merit of a bootstrap error estimate for selecting differentially expressed miRNAs. While rough sets select relevant and significant miRNAs from expression data, the error rate method minimizes the variability and biasedness of the derived results. The effectiveness of the proposed approach is demonstrated on several miRNA microarray expression datasets.

Finally, Ercan et al explain in their paper how to predict protein adsorption that may be beneficial for reducing infection or increasing tissue growth on nanomaterials. Surface properties such as nanoscale roughness and energy significantly influence protein adsorption on to nanomaterials. This study provides an important step in developing mathematical models that can correlate surface properties to protein adsorption for improving medical devices.

Summary

In closing, the seven papers here represent some of the latest and promising research results in the emerging field of nanoinformatics, which continues to have a significant impact on biomedical research and its applications. We hope that this special issue will stimulate further research in this field.

Acknowledgments

The author gratefully acknowledges Dr Kaizhi Tang and Dr Jeffery A Steevens, co-chairs of NanoInfo 2012, for their assistance with the special issue. The author also would like to thank the committee members of NanoInfo 2012 for their overall support of this event. Committee members: Demir Akin, Stanford University, USA; Nathan Baker, Pacific Northwest National Laboratory, USA; Keith Chan, Hong Kong Polytechnic University, Hong Kong; Yoram Cohen, University of California Los Angeles, USA; Jacques Colinge, Austrian Academy of Science, Austria; Nily Dan, Drexel University, USA; Martin Fritts, National Institute of Standards and Technology, USA; Sharon Gaheen, SAIC, USA; Stacey Harper, Oregon State University, USA; Hyunsoo Kim, University of Alabama at Birmingham, USA; Martha Marrapese, Keller and Heckman, USA; Stanley Mofatt, Regent University College of Science and Technology, Ghana; Michelle Ostraat, RTI International, USA; Sue Pan, SAIC, USA; Thomas J Webster, Northeastern University, USA; Jessica Zhang, Carnegie Mellon University, USA.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Maojo V, Fritts M, de la Iglesia D, et al. Nanoinformatics: a new area of research in nanomedicine. *Int J Nanomedicine*. 2012;7:3867–3890.
2. 2012 Workshop on Nanoinformatics for Biomedicine [webpage on the Internet]. Available from: <http://workshops.i-a-i.com/nanoinfo2012>. Accessed November 25, 2012.
3. Nanotechnology Standards [webpage on the Internet]. Available from: <http://www.astm.org/Standards/nanotechnology-standards.html>. Accessed November 25, 2012.

4. ISO Technical Committee 229 on Nanotechnologies [webpage on the Internet]. Available from: http://www.iso.org/iso/iso_technical_committee?commid=381983. Accessed November 25, 2012.
5. Thomas DG, Pappu RV, Baker NA. Nanoparticle ontology for cancer nanotechnology research. *J Biomed Inform.* 2011;44(1):59–74.
6. Characterization Matters [homepage on the Internet]. Available from: <http://characterizationmatters.org>. Accessed November 25, 2012.
7. Horev-Azaria L, Kirkpatrick CJ, Korenstein R, et al. Predictive toxicology of cobalt nanoparticles and ions: comparative in vitro study of different cellular models using methods of knowledge discovery from data. *Toxicol Sci.* 2011;122(2):489–501.
8. Dorney J, Bonnier F, Garcia A, Casey A, Chambers G, Byrne HJ. Identifying and localizing intracellular nanoparticles using Raman spectroscopy. *Analyst.* 2012;137(5):1111–1119.
9. Drupal [webpage on the Internet]. Available from: <http://drupal.org/>. Accessed November 25, 2012.
10. Nanomaterial-Biological Interactions Knowledgebase [homepage on the Internet]. Available from: <http://oregonstate.edu/nbi>. Accessed November 25, 2012.
11. Maojo V, Martin-Sanchez F, Kulikowski C, Rodriguez-Paton A, Fritts M. Nanoinformatics and DNA-based computing: catalyzing nanomedicine. *Pediatr Res.* 2010;67(5):481–489.

International Journal of Nanomedicine

Publish your work in this journal

The International Journal of Nanomedicine is an international, peer-reviewed journal focusing on the application of nanotechnology in diagnostics, therapeutics, and drug delivery systems throughout the biomedical field. This journal is indexed on PubMed Central, MedLine, CAS, SciSearch®, Current Contents®/Clinical Medicine,

Submit your manuscript here: <http://www.dovepress.com/international-journal-of-nanomedicine-journal>

Journal Citation Reports/Science Edition, EMBase, Scopus and the Elsevier Bibliographic databases. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress