ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/279198958

Big Data Analytics in Healthcare

Article *in* BioMed Research International · January 2015

CITATIONS
40

READS
5,280

6 authors, including:

Ashwin Belle
University of Michigan
29 PUBLICATIONS **220** CITATIONS
SEE PROFILE

Raghuram Thiagarajan
Pratt & Miller Engineering
10 PUBLICATIONS **231** CITATIONS
SEE PROFILE

S.M.Reza Soroushmehr
University of Michigan
98 PUBLICATIONS **205** CITATIONS
SEE PROFILE

Kayvan Najarian
University of Michigan
343 PUBLICATIONS **1,161** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Analysis of patient data in traumatic brain injuries View project

Image and Video compression View project

All content following this page was uploaded by S.M.Reza Soroushmehr on 27 June 2015.

The user has requested enhancement of the downloaded file.

Big Data Analytics in Healthcare

Ashwin Belle[*1,4], Raghuram Thiagarajan[*2], S.M.Reza Soroushmehr[*†1,4], Fatemeh Navidi[3], Daniel A. Beard[2,4] and Kayvan Najarian[1,4]

[1]Emergency Medicine Department, University of Michigan, Ann Arbor, MI, USA

[2]Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI, USA

[3]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

[4]University of Michigan Center for Integrative Research in Critical Care (MCIRCC), Ann Arbor, MI , USA

1 Abstract

The rapidly expanding field of big data analytics has started to play a pivotal role in the evolution of healthcare practices and research. It has provided tools to accumulate, manage, analyze and assimilate large volumes of disparate, structured and unstructured data produced by current healthcare systems. Big data analytics has been recently applied towards aiding the process of care delivery and disease exploration. However, the adoption rate and research development in this space is still hindered by some fundamental problems inherent within the big–data paradigm. In this article, we discuss some of these major challenges with a focus on three upcoming and promising areas of medical research; image, signal and genomics based analytics. Recent research which targets utilization of large volumes of medical data while combining multi–modal data from disparate sources are discussed. Potential areas of research within this field which have ability to provide meaningful impact on healthcare delivery are also examined.

*These authors contributed equally to this work.

†Corresponding author, Email: ssoroush@umich.edu

2 Introduction

The concept of 'big data' is not new, however the way it is defined is constantly changing. Various attempts at defining big data essentially characterize it as a collection of data elements whose size, speed, type and/or complexity require one to seek, adopt and invent new hardware and software mechanisms in order to successfully store, analyze and visualize the data [1–3]. Healthcare is a prime example of how the three V's of data, velocity (speed of generation of data), variety and volume [4], are an innate aspect of the data it produces. This data is spread among multiple healthcare systems, health insurers, researchers, government entities, etc. Furthermore, each of these data repositories is siloed and inherently incapable of providing a platform for global data transparency. To add to the three V's, the veracity of healthcare data is also critical for its meaningful use towards developing translational research.

Despite the inherent complexities of healthcare data, there is potential and benefit in developing and implementing big data solutions within this realm. A report by McKinsey Global Institute suggests that if US healthcare were to use big data creatively and effectively, the sector could create more than $300 billion in value every year. Two–thirds of the value would be in the form of reducing US healthcare expenditure [5]. Historical approaches to medical research have generally focused on the investigation of disease states based on the changes in physiology in the form of a confined view of certain singular modality of data [6]. Although this approach to understanding diseases is essential, research at this level mutes the variation and interconnectedness that define the true underlying medical mechanisms [7]. After decades of technological laggard, the field of medicine has begun to acclimatize to today's digital data age. New technologies make it possible to capture vast amounts of information about each individual patient over a large timescale. However, despite the advent of medical electronics, the data captured and gathered from these patients has remained vastly underutilized and thus wasted.

Important physiological and pathophysiological phenomena concurrently manifest as changes across multiple clinical streams. This results from strong coupling among different systems within the body (e.g., interactions between heart rate, respiration, blood pressure, etc.) thereby producing potential markers for clinical assessment. Thus, understanding and predicting diseases require an aggregated approach where structured and unstructured data stemming from a myriad of clinical and non–clinical modalities are utilized for a more comprehensive perspective of the disease–states. An aspect of healthcare research that has recently gained traction is in addressing some of the growing pains in introducing concepts of big data analytics to medicine. Researchers are studying the complex nature of healthcare data both in terms of characteristics of the data itself as well as in the taxonomy of analytics that can be meaningfully performed

on them.

In this paper, three areas of big data analytics in medicine are discussed. These three areas do not comprehensively reflect the application of big data analytics in medicine; instead they are intended to provide a perspective of broad, popular areas of research where the concepts of big–data analytics are currently being applied.

1) *Image processing;* Medical images are an important source of data frequently used for diagnosis, therapy assessment and planning [8]. Computed tomography (CT), magnetic resonance imaging (MRI), X–ray, molecular imaging, ultrasound, photoacoustic imaging, fluoroscopy, positron emission tomography–computed tomography (PET–CT), mammography are some of the examples of imaging techniques that are well established within clinical settings. Medical image data can range anywhere from a few megabytes for a single study (e.g. an histology image) to hundreds of megabytes per study (e.g. thin–slice CT studies comprising upto 2500+ scans per study). Such data requires large storage capacities if stored long term. It also demands fast and accurate algorithms if any decision assist automation were to be performed using the data. In addition, if other sources of data acquired for each patient are also utilized during the diagnoses, prognosis and treatment processes, then the problem of providing cohesive storage and developing efficient methods capable of encapsulating the broad range of data becomes a challenge.

2) *Signal processing;* Similar to medical images, medical signals also pose volume and velocity obstacles especially during continuous, high–resolution acquisition and storage from a multitude of monitors connected to each patient. However, in addition to the data size issues, physiological signals also pose complexity of a spatio–temporal nature. Analysis of physiological signals is often more meaningful when presented along with situational context awareness which needs to be embedded into the development of continuous monitoring and predictive systems to ensure its effectiveness and robustness.

Currently healthcare systems use numerous disparate and continuous monitoring devices that utilize singular physiological waveform data or discretized vital information to provide alert mechanisms in case of overt events. However, such uncompounded approaches towards development and implementation of alarm systems tends to be unreliable and their sheer numbers could cause *'alarm fatigue'* for both care givers and patients [9–11]. In this setting, the ability to discover new medical knowledge is constrained by prior knowledge that has typically fallen short of maximally utilizing high–dimensional time–series data. The reason that these alarm mechanisms tend to fail is primarily because these systems tend to rely on single sources of information while lacking context of the patients' true physiological conditions from a broader and more comprehensive viewpoint. Therefore, there is a need to develop improved and more comprehensive

approaches towards studying interactions and correlations between multi–modal clinical time–series data. This is important because studies continue to show that humans are poor in reasoning about changes affecting more than two signals [12–14].

3) *Genomics;* The cost to sequence the human genome (encompassing 30,000 to 35,000 genes) is rapidly decreasing with the development of high–throughput sequencing technology [15, 16]. With implications for current public health policies and delivery of care [17, 18], analyzing genome–scale data for developing actionable recommendations in a timely manner is a significant challenge to the field of computational biology. Cost and time to deliver recommendations are crucial in a clinical setting. Initiatives tackling this complex problem include tracking of 100,000 subjects over 20 to 30 years using the predictive, preventive, participatory and personalized health, refer to as P4, medicine paradigm [19–21] and an integrative personal omics profile [22]. The P4 initiative is using a system approach for (i) analyzing genome–scale data sets to determine disease states, (ii) moving towards blood based diagnostic tools for continuous monitoring of a subject, (iii) exploring new approaches to drug target discovery, developing tools to deal with big data challenges of capturing, validating, storing, mining, integrating, and finally (iv) modeling data for each individual. The integrative personal omics profile (iPOP) combines physiological monitoring and multiple high–throughput methods for genome sequencing to generate a detailed health and disease states of a subject [22]. Ultimately realizing actionable recommendations at the clinic level remains a grand challenge for this field [23, 24]. Utilizing such high density data for exploration, discovery and clinical translation demands novel big–data approaches and analytics.

Despite the enormous expenditure consumed by the current healthcare systems, clinical outcomes remain suboptimal, particularly in the United States, where 96 people per 100,000 die annually from conditions considered treatable [25]. A key factor attributing towards such inefficiencies is the inability to effectively gather, share and use information in a more comprehensive manner within the healthcare systems [26]. This is an opportunity for big data analytics to play a more significant role in aiding the exploration and discovery process, improving the delivery of care, helping to design and plan healthcare policy, providing a means for comprehensively measuring and evaluating the complicated and convoluted data of healthcare. More importantly, adoption of insights gained from big data analytics has the potential to save lives, improve care delivery, expand access to healthcare, align pay with performance, and help curb the vexing growth of healthcare costs.

3 Medical Image Processing from Big Data Point of View

Medical imaging provides important information on anatomy and organ function in addition to detecting diseases–states. Moreover, it is utilized for organ delineation, identifying tumors in lungs, spinal deformity diagnosis, artery stenosis detection, aneurysm detection, etc. In these applications image processing techniques such as enhancement, segmentation and denoising in addition to machine learning methods are employed. As the size and dimensionality of data increase, understanding the dependencies among the data and designing efficient, accurate and computationally effective methods demand new computer–aided techniques and platforms. The rapid growth in the number of health–care organizations as well as the number of patients has resulted in the greater use of computer–aided medical diagnostics and decision support systems in clinical settings. Many areas in health care such as diagnosis, prognosis and screening can be improved by utilizing computational intelligence [27]. The integration of computer analysis with appropriate care has potential to help clinicians improve diagnostic accuracy [28]. The integration of medical images with other types of electronic health record (EHR) data and genomic data can improve the accuracy and reduce the time taken for a diagnosis.

In the following, data produced by imaging techniques are reviewed and applications of medical imaging from a big data point of view are discussed.

3.1 Data Produced by Imaging Techniques

Medical imaging encompasses a wide spectrum of different image acquisition methodologies typically utilized for a variety of clinical applications. For example, visualizing blood vessel structure can be performed using magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and photoacoustic imaging [29]. From a data dimension point of view, medical images might have 2, 3 and four dimensions. Positron emission tomography (PET), CT, 3D ultrasound and functional MRI (fMRI) are considered as multi–dimensional medical data. Modern medical image technologies can produce high–resolution images such as respiration–correlated or "four dimensional" computed tomography (4D CT) [30]. Higher resolution and dimensions of these images generates large volumes of data requiring high performance computing (HPC) and advanced analytical methods for its utilization. For instance, microscopic scans of a human brain with high resolution can require 66TB of storage space [31]. Although the volume and variety of medical data make its analysis a big challenge, advances in medical imaging could make individualized care more practical [32] and provide quantitative information in variety of applications such as disease stratification,

predictive modeling, decision making systems and so on. In the following we refer to two medical imaging techniques and one of their associated challenges.

Molecular imaging is a non–invasive technique of cellular and sub–cellular events [33] which has the potential for clinical diagnosis of disease–states such as cancer. However, in order to make it clinically applicable for patients, the interaction of radiology, nuclear medicine and biology is crucial [34] that could complicate its automated analysis.

Microwave imaging is an emerging methodology that could create a map of electromagnetic wave scattering arising from the contrast in the dielectric properties of different tissues [35]. It has both functional and physiological information encoded in the dielectric properties which can help differentiate and characterize different tissues and/or pathologies [36]. However, microwaves have scattering behavior that makes retrieval of information a challenging task.

The integration of images from different modalities and/or other clinical and physiological information could improve the accuracy of diagnosis and outcome prediction of disease. Liebeskind and Feldmann explored advances in neurovascular imaging and the role of multimodal CT or MRI including angiography and perfusion imaging on evaluating the brain vascular disorder, and achieving precision medicine [32]. Delayed enhanced MRI is used for exact assessment of myocardial infarction scar and electroanatomic mapping (EAM) can help in identifying the subendocardial extension of infarct [37]. The role of evaluating both MRI and CT images to increase the accuracy of diagnosis in detecting the presence of erosions and osteophytes in the temporomandibular joint (TMJ) has been investigated by Hussain et al [38]. According to this study simultaneous evaluation of all the available imaging techniques is an unmet need.

Advanced Multimodal Image–Guided Operating (AMIGO) suite has been designed which has angiographic X–ray system, MRI, 3D ultrasound and PET/CT imaging in the operating room. This system has been used for cancer therapy and showed the improvement in localization and targeting an individual's diseased tissue [39].

Besides the huge space required for storing all the data and their analysis, finding the map and dependencies among different data types are challenges for which there is no optimal solution yet.

3.2 Methods

The volume of medical images is growing exponentially. For instance, ImageCLEF medical image dataset contained around 66,000 images between 2005 and 2007 while just in the year of 2013 around 300,000 images were stored everyday [40]. In addition to the growing volume of images, they differ in modality, resolution,

dimension and quality which introduce new challenges such as data–integration and mining specially if multiple datasets are involved. Compared to the volume of research that exists on single modal medical image analysis, there are considerably lesser number of research initiatives on multi–modal image analysis.

When utilizing data at a local/institutional level, an important aspect of the research is on how the developed system is evaluated and validated. Having annotated data or a structured method to annotate new data is a real challenge. This becomes even more challenging when large scale data integration from multiple institutions are taken into account. For the same applications and the same modality such as CT scans for traumatic brain injury, different institutes might use different settings for image acquisitions. In order to benefit the multi–modal images and their integration with other medical data, new analytical methods with real–time feasibility and scalability are required. In the following we look at analytical methods that deal with some aspects of big data.

3.2.1 Analytical Methods

The goal of medical image analytics is to improve the interpretability of depicted contents [8]. Many methods and frameworks have been developed for medical image processing. However, these methods are not necessarily applicable for big data analytics.

One of the frameworks developed for analyzing and transformation of very large datasets is Hadoop that uses MapReduce [41, 42]. MapReduce is a programming paradigm that provides scalability across many servers in a Hadoop cluster with a broad variety of real–world applications [43–45]. However, it doesn't perform well with input–output intensive tasks [46]. MapReduce framework has been used in [46] to increase the speed of three large–scale medical image processing use–cases, (i) employing a well–known machine learning method, support vector machines (SVM), to find optimal parameter for lung texture classification (ii) content–based medical image indexing, and (iii) wavelet analysis for solid texture classification. In this framework, a cluster of heterogeneous computing nodes with a maximum of 42 concurrent map tasks was set up and the speedup around 100 was achieved. In other words, total execution time for finding optimal SVM parameters was reduced from about 1000h to around 10h. Designing a fast method is crucial in some applications such as trauma assessment in critical care where the end goal is to utilize such imaging techniques and its analysis within what is considered as a golden–hour of care [47]. Therefore, execution time or real–time feasibility of developed methods is of importance. Accuracy is another factor that should be considered in designing an analytical method. Finding dependencies among different types of data could help improve the accuracy. A hybrid machine learning method has been developed in [48] that classifies schizophrenia

patients and healthy controls using fMRI images and single nucleotide polymorphism (SNP) data [48]. A classification accuracy of 87% has been achieved which is higher than using either data alone. Alfonso et al. have compared some organ segmentation methods when data is considered as big data. They have proposed a method that incorporates both the local contrast of the image and atlas probabilistic information [49]. An average of 33% improvement has been achieved compared to using only atlas information. Tsymbal et al. have designed a clinical decision support system that exploits discriminative distance learning with significantly lower computational complexity compared to classical alternatives and hence this system is more scalable to retrieval from big data [50]. A computer–aided decision support system was developed by Wenan et al [51] that can assist physicians to provide accurate treatment planning for patients suffering from traumatic brain injury (TBI). In this method, patient's demographic information, medical records, and features extracted from CT scans were combined to predict the level of intracranial pressure (ICP). The accuracy, sensitivity and specificity were reported to be around 70.3%, 65.2%, 73.7% respectively. In [52], molecular imaging and its impact on cancer detection and cancer drug improvement are discussed. The proposed technology is designed to aid in the early detection of cancer by integrating molecular and physiological information with anatomical information. Using this imaging technique for patients with advanced ovarian cancer, the accuracy of the predictor of response to a special treatment has been increased compared to other clinical or histopathologic criteria. A hybrid digital–optical correlator (HDOC) has been designed to speed up the correlation of images [53]. HDOC can be employed to compare images in the absence of coordinate matching or geo–registration. In this multichannel correlator method, the computation is performed in the storage medium which is a volume holographic memory. These features could help HDOC to be applicable in the area of big–data analytics [53].

3.2.2 Collecting, Sharing and Compressing Methods

In addition to developing analytical methods, efforts have been made for collecting, compressing, sharing and anonymizing medical data. One example is iDASH (integrating data for analysis, anonymization, and sharing) which is a center for biomedical computing [54]. It focuses on algorithms and tools for sharing data in a privacy–preserving manner. The goal of iDASH is to bring together a multi–institutional team of quantitative scientists to develop algorithms and tools, services, and a biomedical cyber–infrastructure to be used by biomedical and behavioral researchers [54]. Another example of a similar approach is Health–e–child consortium of 14 academic, industry, and clinical partners with the aim of developing an integrated healthcare platform for European Paediatrics [50].

Based on Hadoop platform, a system has been designed for exchanging, storing and sharing electronic medical records (EMR) among different healthcare systems [55]. This system can also help users retrieve medical images from a database. Medical data has been investigated from an acquisition point of view where patients' vital data is collected through a network of sensors [56]. This system delivers data to a cloud for storage, distribution and processing. A prototype system has been implemented in [57] to handle standard store/query/retrieve requests on a database of Digital Imaging and Communications in Medicine (DICOM) images. This system uses Microsoft Windows Azure as a cloud computing platform.

When dealing with very large volume of data, compression techniques can help overcome data storage and network bandwidth limitations. Many methods have been developed for medical image compression. However, there are a few methods developed for big data compression. A method has been designed to compress both high–throughput sequencing dataset and the data generated from calculation of log–odds of probability error for each nucleotide [54] while the maximum compression ratios of 400 and 5 haves been achieved respectively. This dataset has medical and biomedical data including genotyping, gene expression, proteomic measurements with demographics, laboratory values, images, therapeutic interventions, and clinical phenotypes for Kawasaki Disease(KD). By illustrating the data with a graph model, a framework for analyzing large–scale data has been presented [58]. For this model, the fundamental signal processing techniques such as filtering and Fourier transform are implemented. In [59], the application of simplicity and power (SP) theory of intelligence in big data has been investigated. The goal of SP theory is to simplify and integrate concepts from multiple fields such as artificial intelligence, mainstream computing, mathematics, and human perception and cognition that can be observed as a brain-like system [59]. The proposed SP system performs lossless compression through the matching and unification of patterns. However, this system is still in the design stage and cannot be supported by today's technologies.

There are some limitations in implementing the application–specific compression methods on both general–purpose processors and parallel processors such as graphics processing units (GPUs) as these algorithms need highly variable control and complex bit manipulations which are not well–suited to GPUs and pipeline architectures. To overcome this limitation, an FPGA implementation is proposed for LZ–factorization which decreases the computational burden of the compression algorithm [60]. A lossy image compression has been introduced in [61] that reshapes the image in such a way that if the image is uniformly sampled, sharp features have a higher sampling density than the coarse ones . This method is claimed to be applicable for big data compression. However, for medical applications lossy methods are not applicable in most cases as fidelity is important and information must be preserved.

Table 1: Challenges facing medical image analysis

Challenges	Description and Possible Solutions
Preprocessing	Medical images suffer from different types of noise/artifacts and missing data. Noise reduction, artifact removal, missing data handling, contrast adjusting and etc could enhance the quality of images and increase the performance of processing methods. Employing multimodal data could be beneficial for this purpose [62–64].
Compression	Reducing the volume of data while maintaining important data such as anatomically relevant data [54, 60, 65].
Parallelization/ Real-time realization	Developing scalable/parallel methods and frameworks to speed up the analysis/processing [60].
Registration/Mapping	Aligning consecutive slices/frames from one scan or corresponding images from different modalities [66, 67].
Sharing /Security/ Anonymization	Integrity, privacy and confidentiality of data must be protected [54, 68–70].
Segmentation	Delineation of anatomical structure such as vessels, bones, and etc [49, 67, 71].
Data Integration/Mining	Finding dependencies/patterns among multimodal data and/or the data captured at different time points in order to increase the accuracy of diagnosis, prediction and overall performance of the system [46, 48, 51, 72].
Validation	Assessing the performance or accuracy of the system/method. Validation can be objective or subjective. For the former, annotated data is usually required [73–75].

These techniques are among a few techniques that have been either designed as prototypes or developed with limited applications. Developing methods for processing/analyzing a broad range and large volume of data with acceptable accuracy and speed is still critical. In Table 1, we summarize the challenges facing medical image processing. When dealing with big-data, these challenges seemed to be more serious and on the other hand analytical methods could benefit the big–data to handle them.

4 Medical Signal Analytics

Telemetry and physiological signal monitoring devices are ubiquitous. However, continuous data generated from these monitors have not been typically stored for more than a brief period of time, thereby neglecting extensive investigation into generated data. However, in the recent past, there has been an increase in the attempts towards utilizing telemetry and continuous physiological time series monitoring to improve patient care and management [76–79].

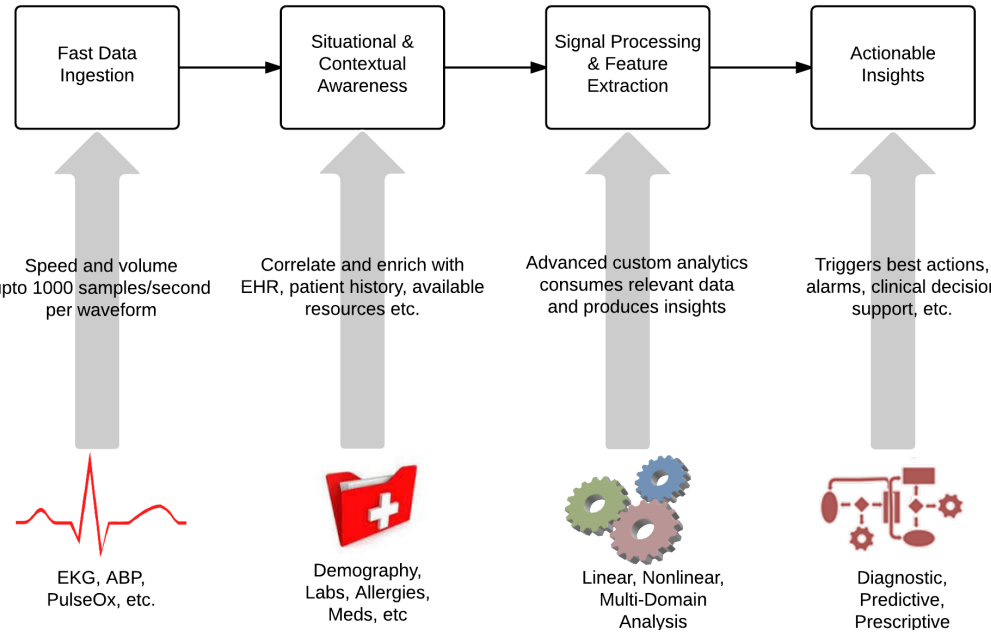

Figure 1: Generalized analytic work–flow using streaming healthcare data

Streaming data analytics in healthcare can be defined as a systematic use of continuous waveform (signal varying against time) and related medical record information developed through applied analytical disciplines (e.g. statistical, quantitative, contextual, cognitive, predictive, etc.) to drive decision making for patient care. The analytics workflow of real–time streaming waveforms in clinical settings can be broadly described using Fig 1. Firstly a platform for streaming data acquisition and ingestion is required which has the bandwidth to handle multiple waveforms at different fidelities. Integrating these dynamic waveform data with static data from the EHR is a key component to provide situational and contextual awareness for the analytics engine. Enriching the data consumed by analytics not only makes the system more robust, but also helps balance the sensitivity and specificity of the predictive analytics. The specifics of the signal processing will largely depend on the type of disease cohort under investigation. A variety of signal processing mechanisms can be utilized to extract a multitude of target features which are then consumed by a pre-trained machine learning model to produce an actionable insight. These actionable insights could either be diagnostic, predictive or prescriptive. These insights could further be designed to trigger other mechanisms such as alarms, notification to physicians, etc.

Harmonizing such continuous waveform data with discrete data from other sources for finding necessary patient information, and conducting research towards development of next generation diagnoses and treatments can be a daunting task [80]. For bed–side implementation of such systems in clinical environments,

there are several technical considerations and requirements that need to be designed and implemented at system, analytic and clinical levels. The following subsections provide an overview of different challenges and existing approaches in the development of monitoring systems that consume both high fidelity waveform data as well as discrete data from non–continuous sources.

4.1 Data acquisition

Historically streaming data from continuous physiological signal acquisition devices was rarely stored. Even if the option to store this data were available, the length of these data captures was typically short and downloaded only using proprietary software and data formats provided by the device manufacturers. Although most major medical device manufactures are now taking steps to provide interfaces to access live streaming data from their devices, such data in motion very quickly poses archetypal big data challenges. Adding to this is the fact that there are also governance challenges such as lack of data protocols, lack of data standards, data privacy issues etc. On the other side there are many challenges within the healthcare systems such as network bandwidth, scalability, cost, etc, that have stalled the wide spread adoption of such streaming data collection [81–83]. This has allowed way for system wide projects which especially cater to medical research communities [76, 78, 79, 84–92].

Research community has interest in consuming data captured from live monitors for developing continuous monitoring technologies [93,94]. There have been several indigenous and off the shelf efforts in developing and implementing systems that enable such data captures [84, 95–98]. There are also products being developed in the industry that facilitate device manufacturer agnostic data acquisition from patient monitors across healthcare systems.

4.2 Data Storage and Retrieval

With the large volumes of streaming data and other patient information that can be gathered from clinical settings, sophisticated storage mechanisms of such data is imperative. Since storing and retrieving can be computational and time expensive, it is key to have a storage infrastructure that facilitates rapid data pull and commits based on analytic demands.

With its capability to store and compute large volumes of data, usage of systems such as Hadoop, MapReduce and MongoDB [99, 100] are becoming much more common with the healthcare research communities. MongoDB is a free cross–platform document–oriented database which eschews traditional table–based relational database. Typically each health system have their own custom relational database schemas and

data models which inhibit interoperability of healthcare data for multi–institutional data sharing or research studies. Furthermore, given the nature of traditional databases integrating data of different types such as streaming waveforms and static EHR data is not feasible. This is where MongoDB and other document-based databases can provide high performance, high availability, and easy scalability for the healthcare data–needs [101, 102]. Apache Hadoop is an open source framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is a highly scalable platform which provides a variety of computing modules such as MapReduce, Spark™, etc. For performing analytics on continuous telemetry waveforms, a module like Spark™ is especially useful since it provides capabilities to ingest and compute on streaming data along with machine learning and graphing tools. Such technologies allow researchers to utilize data for both real–time as well as retrospective analysis, with the end goal to translate scientific discovery into applications for clinical settings in an effective manner.

4.3 Data Aggregation

Integration of disparate sources of data, developing consistency within the data, standardization of data from similar sources, improving the confidence in the data especially towards utilizing automated analytics are among challenges facing data aggregation in healthcare systems [103]. Medical data can be complex in nature as well as being interconnected and interdependent, hence simplification of this complexity is important. Medical data is also subject to the highest level of scrutiny for privacy and provenance from governing bodies, therefore developing secure storage, access and use of the data are very important [104]

Analysis of continuous data heavily utilizes the information in time domain, however static data does not always provide true time context. Hence, when combining the waveform data with static electronic health record data, the temporal nature of the time context during integration can also add significantly to the challenges. There are considerable efforts in compiling waveforms and other associated electronic medical information into one cohesive database that are made publicly available for researchers worldwide [105, 106]. For example, MIMIC II [107, 108] and some other datasets included in Physionet [95] provide waveform and other clinical data from a wide variety of actual patient cohorts.

4.4 Signal Analytics using Big Data

Research in signal processing for developing big data based clinical decision support systems (CDSS) is getting more prevalent [109]. In fact organizations such as the Institution of Medicine have long advocated use of health information technology including CDSS to improve care quality [110]. CDSSs provide medical prac-

titioners with knowledge and patient–specific information, intelligently filtered and presented at appropriate times, to improve the delivery of care [111].

A vast amounts of data in short periods of time is produced in intensive care units (ICU) where large volumes of physiological data is acquired from each patient. Hence, the potential for developing CDSS in an ICU environment has been recognized by many researchers. A scalable infrastructure for developing a patient care management system has been proposed which combines static data and stream data monitored from critically ill patients in the ICU for data mining and alerting medical staff of critical events in real time [112]. Similarly, Bressan et al. developed an architecture specialized for a neonatal ICU which utilized streaming data from infusion pumps, EEG monitors, cerebral oxygenation monitors, etc. to provide clinical decision support [113]. A clinical trial is currently underway which extracts biomarkers through signal processing from heart and respiratory waveforms in real time to test whether maintaining stable heart rate and respiratory rate variability throughout the spontaneous breathing trials, administered to patients before extubation, may predict subsequent successful extubation [114]. An animal study shows how acquisition of non–invasive continuous data such as tissue oxygenation, fluid content and blood flow can be used as indicators of soft tissue healing in wound care [77]. Electrocardiogrpah parameters from telemetry along with demographic information including medical history, ejection fraction, laboratory values, and medications have been used to develop an in–hospital early detection system for cardiac arrest [115].

A study presented by Lee et al. uses the MIMIC II database to prompt therapeutic intervention to hypotensive episodes using cardiac and blood pressure time series data [116]. Another study shows the use of physiological waveform data along with clinical data from the MIMIC II database for finding similarities among patients within the selected cohorts [117]. This similarity can potentially help care givers in the decision making process while utilizing outcomes and treatments knowledge gathered from similar disease cases from the past. A combination of multiple waveform information available in the MIMIC II database is utilized to develop early detection of cardiovascular instability in patients [118]. Many types of physiological data captured in the operative and pre–operative care settings and how analytics can consume these data to help continuously monitor the status of the patients during, before and after surgery are described in [119]. The potential of developing data fusion based machine learning models which utilizes biomarkers from breathomics (metabolomics study of exhaled air) as a diagnostic tool is demonstrated in [120].

Research in neurology has shown interest in electro–physiologic monitoring of patients to not only examine complex diseases under a new light but to also develop next generation diagnostics and therapeutic devices. An article focusing on neurocritical care explores the different physiological monitroing systems specifically

developed for the care of patients with disorders who require neurocritical care [121]. The authors of this article do not make specific recommendations about treatment, imaging, and intraoperative monitoring, instead they examine the potentials and implications of neuromonitoring with differeing quality of data and also provide guidance on developing research and application in this area. The development of multimodality monitoring for traumatic brain injury patients and individually tailored, patient specific care are examined in [122]. Zanatta et al. have investigated whether multimodal brain monitoring performed with TCD, EEG, and SEPs reduces the incidence of major neurologic complications in patients who underwent cardiac surgery. The authors evaluated whether the use of multimodality brain monitoring shortened the duration of mechanical ventilation required by patients as well as ICU and healthcare stays. The concepts of multimodal monitoring for secondary brain injury in neurocritical care as well as outline initial and future approaches using informatics tools for understanding and applying such data towards clinical care are described in [123].

As complex physiological monitoring devices are getting smaller, cheaper and more portable personal monitoring devices are being used outside of clinical environments both by patients and enthusiasts alike. However, similar to clinical applications, combining information simultaneously collected from multiple portable devices can become challenging. Pantelopoulos et al. discussed the research and development of wearable biosensor systems and identify the advantages and shortcomings in this area of study [124]. Similarly, portable and connected electrocardiogram, blood pressure and body weight devices are used to setup a network based study of telemedicine [125]. The variety of fixed as well as mobile sensors available for data mining in the healthcare sector and how such data can be leveraged for developing patient care technologies are surveyed in [126].

5 Big Data Applications in Genomics

The advent of high–throughput sequencing methods has enabled researchers to study genetic markers over a wide range of population [21,127], improve efficiency by more than five orders of magnitude since sequencing of the human genome was completed [128] and ability to associate genetic causes of the phenotype in disease states [129]. Genome–wide analysis utilizing microarrays have been successful in analyzing traits across a population and contributed successfully in treatments of complex diseases such as Crohn's disease and age–related muscular degeneration [129].

Analytics of high–throughput sequencing techniques in genomics is an inherently big data problem as the human genome consists of 30,000 to 35,000 genes [15,16]. Initiatives are currently being pursued over

the timescale of years to integrate clinical data from the genomic level to the physiological level of a human being [21, 22]. These initiatives will help in delivering personalized care to each patient. Delivering recommendations in a clinical setting requires fast analysis of genome–scale big data in a reliable manner. This field is still in a nascent stage with applications in specific focus areas, such as cancer [130–133], because of cost, time and labor intensive nature of analyzing this big data problem.

Big Data applications in genomics cover a wide variety of topics. Here we focus on pathway analysis, in which functional effects of genes differentially expressed in an experiment or gene set of particular interest are analyzed, and the reconstruction of networks, where the signals measured using high–throughput techniques are analyzed to reconstruct underlying regulatory networks. These networks influence numerous cellular processes which affect the physiological state of a human being [134].

5.1 Pathway Analysis

Resources for inferring functional effects for '-omics' big data are largely based on statistical associations between observed gene expression changes and predicted functional effects. Experiment and analytical practices lead to error [135] as well as batch effects [136]. Interpretation of functional effects has to incorporate continuous increases in available genomic data and corresponding annotation of genes [24]. There are variety of tools, but no "gold standard" for functional pathway analysis of high–throughput genome–scale data [137]. Three generations of methods used for pathway analysis [24] are described as follows.

The first generation encompasses over–representation analysis approaches that determine the fraction of genes in a particular pathway found among the genes which are differentially expressed [24]. Examples of the first generation tools are Onto–Express [138, 139], GoMiner [140] and ClueGo [141]. The second generation includes functional class scoring approaches which incorporate expression level changes in individual genes as well as functionally similar genes [24]. GSEA [142] is a popular tool that belongs to the second generation of pathway analysis. The third generation includes pathway topology based tools which are publicly available pathway knowledge databases with detailed information of gene products interactions: how specific gene products interact with each other and the location where they interact [24]. Pathway–Express [143] is an example of a third generation tool that combines the knowledge of differentially expressed genes with biologically meaningful changes on a given pathway to perform pathway analysis.

5.2 Reconstruction of Regulatory Networks

Pathway analysis approaches do not attempt to make sense of high–throughput big data in biology as arising from the integrated operation of a dynamical system [24]. There are multiple approaches to analyzing genome–scale data using a dynamical system framework [134, 144, 145]. Due to the breadth of the field, in this section we mainly focus on techniques to infer network models from biological big data. Applications developed for network inference in systems biology for big data applications can be split into two broad categories consisting of reconstruction of metabolic networks and gene regulatory networks [134]. Various approaches of network inference vary in performance, and combining different approaches has shown to produce superior predictions [145, 146].

Reconstruction of metabolic networks has advanced in last two decades. One objective is to develop an understanding of organism–specific metabolism through reconstruction of metabolic networks by integrating genomics, transcriptomics and proteomics high–throughput sequencing techniques [147–154]. Constraint–based methods are widely applied to probe the genotype–phenotype relationship and attempt to overcome the limited availability of kinetic constants [155,156]. There are multitude of challenges in terms of analyzing genome–scale data including the experiment and inherent biological noise, differences among experimental platforms, and connecting gene expression to reaction flux used in constraint–based methods [157, 158].

Available reconstructed metabolic networks include Recon 1 [147], Recon 2 [153], SEED [149], IOMA [151] and MADE [159]. Recon 2 (an improvement over Recon 1) is a model to represent human metabolism and incorporates $7,440$ reactions involving $5,063$ metabolites. Recon 2 has been expanded to account for known drugs for drug target prediction studies [160] and to study off–target effects of drugs [161].

Reconstruction of gene regulatory networks from gene expression data is another well developed field. Network inference methods can be split into five categories based on the underlying model in each case: Regression, mutual information, correlation, Boolean regulatory networks, and other techniques [145]. Over 30 inference techniques were assessed post–DREAM5 challenge in 2010 [145]. Performance varied within each category and there was no category that was found to be consistently better than the others. Different methods utilize different information available in experiments which can be in the form of time series, drug perturbation experiments, gene knockouts and combinations of experimental conditions. A tree–based method (using ensembles of regression trees) [162] and two way ANOVA (analysis of variance) method [163] gave the highest performance in a recent DREAM challenge [146].

Boolean regulatory networks [134] are a special case of discrete dynamical models where the state of a node or a set of nodes exists in a binary state. The actual state of each node or set of nodes is determined by using

Table 2: Summary of popular methods and toolkits with their applications.

Toolkit Name	**Category**	**Selected Applications**
Onto–Express [138, 139]	Pathway Analysis	Breast Cancer [168]
GoMiner [140]	Pathway Analysis	Pancreatic Cancer [169]
ClueGo [141]	Pathway Analysis	Colorectal Tumors [170]
GSEA [142]	Pathway Analysis	Diabetes [171]
Pathway–Express [143]	Pathway Analysis	Leukemia [172]
Recon 2 [153]	Reconstruction of Metabolic Networks	Drug Target Prediction Studies [160]
Boolean Methods [134, 145, 164]	Reconstruction of Gene Regulatory Networks	Cardiac Differentiation [173]
ODE models [174–177]	Reconstruction of Gene Regulatory Networks	Cardiac Development [177]

Boolean operations on the state of other nodes in the network [164]. Boolean networks are extremely useful when amount of quantitative data is small [134, 164], but yield high number of false positives (when a given condition is satisfied when actually that is not the case) that may be reduced by using prior knowledge [165, 166]. Another bottleneck is that Boolean networks are prohibitively expensive when the number of nodes in network is large. This is due to the number of global states rising exponentially in the number of entities [134]. A method to overcome this bottleneck is to use clustering to break down the problem size. For example, Martin *et al.* [167] broke down a 34,000–probe microarray gene expression data set into 23 sets of meta–genes using clustering techniques. This Boolean model successfully captured the network dynamics for two different immunology microarray datasets. The dynamics of gene regulatory network can be captured using ordinary differential equations (ODEs) [174–177]. This approach has been applied to determine regulatory network for yeast [174]. The study successfully captured the regulatory network, which has been characterized using experiments by molecular biologists. Reconstruction of a gene regulatory network on a genome–scale system as a dynamical model is computationally intensive [134]. A parallelizeable dynamical ODE model has been developed to address this bottleneck [178]. It reduces the computational time to $\mathcal{O}(N^2)$ from time taken in other approaches which is $\mathcal{O}(N^3)$ or $\mathcal{O}(N^2 \log N)$ [178]. Determining connections in the regulatory network for a problem of the size of the human genome, consisting of 30,000 to 35,000 genes [15, 16], will require exploring close to a billion possible connections. The dynamical ODE model has been applied to reconstruct

the cardiogenic gene regulatory network of the mammalian heart [177]. A summary of methods and toolkits with their applications is presented in Table 2.

6 Conclusion

Big data analytics which leverages legions of disparate, structured and unstructured data sources is going to play a vital role in how healthcare is practiced in the future. One can already see a spectrum of analytics being utilized, aiding in the decision making and performance of healthcare personnel and patients. Here we focused on three areas of interest: medical image analysis, physiological signal processing and integration of physiological data with genomic data. The exponential growth of the volume of medical images forces computational scientists to come up with innovative solutions to process this large volume of data in tractable timescales. The trend of adoption of computational systems for physiological signal processing from both research and practicing medical professionals is growing steadily with the development of some very imaginative and incredible systems that help save lives. Developing a detailed model of a human being by combining physiological data and high–throughput '–omics' techniques has the potential to enhance our knowledge of disease states and help in the development of blood based diagnostic tools [19–21]. Medical image analysis, signal processing of physiological data, and integration of physiological and '–omics' data face similar challenges and opportunities in dealing with disparate structured and unstructured big data sources.

Medical image analysis covers many areas such as image acquisition, formation/reconstruction, enhancement, transmission, and compression. New technological advances have resulted in higher resolution, dimension and availability of multi–modal images which lead to the increase in accuracy of diagnosis and improvement of treatment. However, integrating medical images with different modalities or with other medical data is a potential opportunity. New analytical frameworks and methods are required to analyze these data in a clinical setting. These methods address some concerns, opportunities and challenges such as: features from images which can improve the accuracy of diagnosis, ability to utilize disparate sources of data to increase the accuracy of diagnosis and reducing cost, and improving the accuracy of processing methods such as medical image enhancement, registration and segmentation to deliver better recommendations at the clinical level.

Although there are some very real challenges for signal processing of physiological data to deal with, given the current state of data competency and non–standardized structure, there are opportunities in each

step of the process towards providing systemic improvements within the healthcare research and practice communities. Apart from the obvious need for further research in the area of data wrangling, aggregating and harmonizing continuous and discrete medical data formats, there is also an equal need for developing novel signal processing techniques specialized towards physiological signals. Research pertaining to mining for bio–markers and clandestine patterns within bio–signals to understand and predict disease cases has shown potential in providing actionable information. However, there are opportunities for developing algorithms to address data filtering, interpolation, transformation, feature extraction, feature selection, etc. Furthermore, with the notoriety and improvement of machine learning algorithms, there are opportunities in improving and developing robust CDSS for clinical prediction, prescription and diagnostics [179, 180].

Integration of physiological data and high–throughput '–omics' techniques to deliver clinical recommendations is the grand challenge for systems biologists. Although associating functional effects with changes in gene expression has progressed, the continuous increase in available genomic data and its corresponding effects of annotation of genes, and errors from experiment and analytical practices make analyzing functional effect from high–throughput sequencing techniques a challenging task.

Reconstruction of networks on the genome–scale is an ill–posed problem. Robust applications have been developed for reconstruction of metabolic networks and gene regulatory networks. Limited availability of kinetic constants is a bottleneck and hence various models attempt to overcome this limitation. There is an incomplete understanding for this large–scale problem as gene regulation, effect of different network architectures, and evolutionary effects on these networks are still being analyzed [134]. To address these concerns, the combination of careful design of experiments and model development for reconstruction of networks will help in saving time and resources spent in building understanding of regulation in genome–scale networks. The opportunity to address the grand challenge requires close co–operation among experimentalists, computational scientists and clinicians.

Author's contributions

AB is the primary author for the section on signal processing and contributed to the whole article, RT is the primary author for the section on genomics and contributed to the whole article, and SS is the primary author for the image processing section and contributed to the whole article. FN contributed to the section on image processing. DAB contributed and supervised the whole article. KN contributed and supervised the whole article. All authors have read and approved the final version of this manuscript.

Acknowledgement

Authors would like to thank Dr. Jason N. Bazil for his valuable comments on the article.

References

[1] Andrew McAfee, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. Big data:the management revolution. *Harvard Bus Rev*, 90(10):60–68, 2012.

[2] Clifford Lynch. Big data: How do your data grow? *Nature*, 455(7209):28–29, 2008.

[3] Adam Jacobs. The pathologies of big data. *Communications of the ACM*, 52(8):36–44, 2009.

[4] Paul Zikopoulos, Chris Eaton, et al. *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media, 2011.

[5] James Manyika, Michael Chui, Brad Brown, et al. Big data: The next frontier for innovation, competition, and productivity. 2011.

[6] Jeffrey J Borckardt, Michael R Nash, Martin D Murphy, Mark Moore, Darlene Shaw, and Patrick O'Neil. Clinical practice as natural laboratory for psychotherapy research: a guide to case-based time-series analysis. *American psychologist*, 63(2):77, 2008.

[7] Leo Anthony Celi, Roger G Mark, David J Stone, and Robert A Montgomery. Big data in the intensive care unit. closing the data loop. *American journal of respiratory and critical care medicine*, 187(11):1157, 2013.

[8] Felix Ritter, Tobias Boskamp, André Homeyer, Hendrik Laue, Michael Schwier, Florian Link, and H-O Peitgen. Medical image analysis. *IEEE Pulse*, 2(6):60–70, 2011.

[9] Barbara J Drew, Patricia Harris, Jessica K Zègre-Hemsey, et al. Insights into the problem of alarm fatigue with physiologic monitor devices: A comprehensive observational study of consecutive intensive care unit patients. *PloS ONE*, 9(10):e110274, 2014.

[10] Kelly Creighton Graham and Maria Cvach. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care*, 19(1):28–34, 2010.

[11] Maria Cvach. Monitor alarm fatigue: an integrative review. *Biomedical Instrumentation & Technology*, 46(4):268–277, 2012.

[12] Jeffrey M Rothschild, Christopher P Landrigan, John W Cronin, Rainu Kaushal, et al. The critical care safety study: The incidence and nature of adverse events and serious medical errors in intensive care. *Critical care medicine*, 33(8):1694–1700, 2005.

[13] Pascale Carayon and Ayşe P Gürses. A human factors engineering conceptual framework of nursing workload and patient safety in intensive care units. *Intensive and Critical Care Nursing*, 21(5):284–301, 2005.

[14] Pascale Carayon. Human factors of complex sociotechnical systems. *Applied ergonomics*, 37(4):525–535, 2006.

[15] Eric S. Lander, Lauren M. Linton, Bruce Birren, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, Feb 15 2001. Date revised - 2013-05-01; Last updated - 2013-09-25.

[16] Radoje Drmanac, Andrew B. Sparks, Matthew J. Callow, et al. Human genome sequencing using unchained base reads on self-assembling dna nanoarrays. *Science*, 327(5961):78–81, 2010.

[17] Timothy Caulfield, Jim Evans, Amy McGuire, et al. Reflections on the cost of "low-cost" whole genome sequencing: Framing the health policy debate. *PLoS Biol*, 11(11):e1001699, 11 2013.

[18] Dewey FE, Grove ME, Pan C, et al. Clinical interpretation and implications of whole-genome sequencing. *JAMA*, 311(10):1035–1045, 2014.

[19] Leroy Hood and Stephen H Friend. Predictive, personalized, preventive, participatory (p4) cancer medicine. *Nature Reviews Clinical Oncology*, 8:184–187, 2011.

[20] Leroy Hood and Mauricio Flores. A personal view on systems medicine and the emergence of proactive {P4} medicine: predictive, preventive, personalized and participatory. *New Biotechnology*, 29(6):613 – 624, 2012.

[21] Leroy Hood and Nathan D. Price. Demystifying disease, democratizing health care. *Science Translational Medicine*, 6(225):225ed5, 2014.

[22] Rui Chen, George I. Mias, Jennifer Li-Pook-Than, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, 148(6):1293 – 1307, 2012.

[23] Guy Haskin Fernald, Emidio Capriotti, Roxana Daneshjou, Konrad J. Karczewski, and Russ B. Altman. Bioinformatics challenges for personalized medicine. *Bioinformatics*, 27(13):1741–1748, 2011.

[24] Purvesh Khatri, Marina Sirota, and Atul J. Butte. Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Comput Biol*, 8(2):e1002375, Feb 2012.

[25] Olawole O Obembe. Bioinformatics, healthcare informatics and analytics: An imperative for improved healthcare system. *International Journal of Applied Information Systems*, 8(5), 2015.

[26] Thomas G Kannampallil, Amy Franklin, Trevor Cohen, and Timothy G Buchman. Sub-optimal patterns of information use: A rational analysis of information seeking behavior in critical care. In *Cognitive Informatics in Health and Biomedicine*, pages 389–408. Springer London, 2014.

[27] H. Elshazly, A.T. Azar, A. El-korany, and A.E. Hassanien. Hybrid system for lymphatic diseases diagnosis. In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 343–347, Aug 2013.

[28] Geoff Dougherty. *Digital image processing for medical applications*. Cambridge University Press, 2009.

[29] Ryan C Gessner, C Brandon Frederick, F Stuart Foster, and Paul A Dayton. Acoustic angiography: a new imaging modality for assessing microvasculature architecture. *Journal of Biomedical Imaging*, 2013:14, 2013.

[30] K Bernatowicz, P Keall, P Mishra, A Knopf, A Lomax, and J Kipritidis. Quantifying the impact of respiratory-gated 4d ct acquisition on thoracic image quality: A digital phantom study. *Medical Physics*, 42(1):324–334, 2015.

[31] Ingrid Scholl, Til Aach, Thomas M Deserno, and Torsten Kuhlen. Challenges of medical image processing. *Computer science-Research and development*, 26(1-2):5–13, 2011.

[32] David S Liebeskind and Edward Feldmann. Imaging of cerebrovascular disorders: precision medicine and the collaterome. *Annals of the New York Academy of Sciences*, 2015.

[33] Timon Hussain and Quyen T Nguyen. Molecular imaging for cancer diagnosis and surgery. *Advanced drug delivery reviews*, 66:90–100, 2014.

[34] G Baio. Molecular imaging is the key driver for clinical cancer diagnosis in the next century. *J Mol Imaging Dynam*, 2:e102, 2013.

[35] S Mustafa, B Mohammed, and A Abbosh. Novel preprocessing techniques for accurate microwave imaging of human brain. *IEEE Antennas and Wireless Propagation Letters*, 12:460–463, 2013.

[36] Amir H Golnabi, Paul M Meaney, and Keith D Paulsen. Tomographic microwave imaging with incorporated prior spatial information. *IEEE Transactions on Microwave Theory and Techniques*, 61(5):2129–2136, 2013.

[37] Benoit Desjardins, Thomas Crawford, Eric Good, et al. Infarct architecture and characteristics on delayed enhanced magnetic resonance imaging and electroanatomic mapping in patients with postinfarction ventricular arrhythmia. *Heart Rhythm*, 6(5):644–651, 2009.

[38] AM Hussain, G Packota, PW Major, and C Flores-Mir. Role of different imaging modalities in assessment of temporomandibular joint erosions and osteophytes: a systematic review. *Dentomaxillofac Radiol*, 37(2):63–71, 2014.

[39] Clare Tempany, Jagadeesan Jayender, Tina Kapur, Raphael Bueno, Alexandra Golby, Nathalie Agar, and Ferenc A Jolesz. Multimodal imaging for improved diagnosis and treatment of cancers. *Cancer*, 2014.

[40] Antoine Widmer, Roger Schaer, Dimitrios Markonis, and Henning Müller. Gesture interaction for content–based medical image retrieval. In *Proceedings of International Conference on Multimedia Retrieval*, pages 503:503–503:506. ACM, 2014.

[41] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, and Robert Chansler. The hadoop distributed file system. In *IEEE Symposium on Mass Storage Systems and Technologies (MSST)*, pages 1–10. IEEE, 2010.

[42] Dalia Sobhy, Yasser El-Sonbaty, and M Abou Elnasr. Medcloud: healthcare cloud computing system. In *IEEE International Conference on Internet Technology And Secured Transactions*, pages 161–166. IEEE, 2012.

[43] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.

[44] Fei Wang, Vuk Ercegovac, Tanveer Syeda-Mahmood, Akintayo Holder, Eugene Shekita, David Beymer, and Lin Hao Xu. Large-scale multimodal mining for healthcare with mapreduce. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 479–483. ACM, 2010.

[45] Wen-Syan Li, Jianfeng Yan, Ying Yan, and Jin Zhang. Xbase: cloud-enabled information appliance for healthcare. In *EDBT*, pages 675–680, 2010.

[46] D. Markonis, R. Schaer, I. Eggel, H. Muller, and A. Depeursinge. Using mapreduce for large-scale medical image analysis. In *IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB)*, pages 1–1, Sept 2012.

[47] K. Shackelford. System & method for delineation and quantification of fluid accumulation in efast trauma ultrasound images, July 2014. US Patent App. 14/167,448.

[48] Honghui Yang, Jingyu Liu, Jing Sui, Godfrey Pearlson, and Vince D Calhoun. A hybrid machine learning method for fusing fmri and genetic data: combining both improves classification of schizophrenia. *Frontiers in human neuroscience*, 4, 2010.

[49] Oscar Alfonso Jimenez del Toro and Henning Muller. Multi atlas-based segmentation with data driven refinement. In *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 605–608. IEEE, 2014.

[50] Alexey Tsymbal, Eugen Meissner, Michael Kelm, and Martin Kramer. Towards cloud-based image-integrated similarity search in big data. In *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 593–596. IEEE, 2014.

[51] Wenan Chen, Charles Cockrell, KR Ward, and Kayvan Najarian. Intracranial pressure level prediction in traumatic brain injury by extracting features from multiple sources and using machine learning methods. In *Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on*, pages 510–515. IEEE, 2010.

[52] Ralph Weissleder. Molecular imaging in cancer. *Science*, 312(5777):1168–1171, 2006.

[53] Tianxiang Zheng, Liangcai Cao, Qingsheng He, and Guofan Jin. Full-range in-plane rotation measurement for image recognition with hybrid digital-optical correlator. *Optical Engineering*, 53(1):011003–011003, 2013.

[54] Lucila Ohno-Machado, Vineet Bafna, Aziz A Boxwala, et al. iDASH: integrating data for analysis, anonymization, and sharing. *Journal of the American Medical Informatics Association*, pages amiajnl–2011, 2011.

[55] Chao-Tung Yang, Lung-Teng Chen, Wei-Li Chou, and Kuan-Chieh Wang. Implementation of a medical image file accessing system on cloud computing. In *IEEE International Conference on Computational Science and Engineering (CSE)*, pages 321–326. IEEE, 2010.

[56] Carlos Oberdan Rolim, Fernando Luiz Koch, Carlos Becker Westphall, Jorge Werner, Armando Fracalossi, and Giovanni Schmitt Salvador. A cloud computing solution for patient's data collection in health care institutions. In *International Conference on eHealth, Telemedicine, and Social Medicine (ETELEMED)*, pages 95–99. IEEE, 2010.

[57] Chia-Chi Teng, Jonathan Mitchell, Christopher Walker, Alex Swan, Cesar Davila, David Howard, and Travis Needham. A medical image archive solution in the cloud. In *IEEE International Conference on Software Engineering and Service Sciences (ICSESS)*, pages 431–434. IEEE, 2010.

[58] A Sandryhaila and J Moura. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine*, 31(5):80–90, 2014.

[59] J.G. Wolff. Big data and the SP theory of intelligence. *IEEE Access*, 2:301–315, 2014.

[60] Sang Woo Jun, K.E. Fleming, M. Adler, and J. Emer. Zip-io: Architecture for application-specific compression of big data. In *IEEE International Conference on Field-Programmable Technology (FPT)*, pages 343–351, Dec 2012.

[61] Bahram Jalali and Mohammad H Asghari. The anamorphic stretch transform: Putting the squeeze on âĂIJbig dataâĂİ. *Optics and Photonics News*, 25(2):24–31, 2014.

[62] Dan Feldman, Cynthia Sung, and Daniela Rus. The single pixel gps: learning big data signals from tiny coresets. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pages 23–32. ACM, 2012.

[63] Lionel Chiron, Maria A van Agthoven, Bruno Kieffer, Christian Rolando, and Marc-André Delsuc. Efficient denoising algorithms for large experimental datasets and their applications in fourier transform ion cyclotron resonance mass spectrometry. *Proceedings of the National Academy of Sciences*, 111(4):1385–1390, 2014.

[64] A Gilbert, Piotr Indyk, Mark Iwen, and Ludwig Schmidt. Recent developments in the sparse fourier transform: A compressed fourier transform for big data. *Signal Processing Magazine, IEEE*, 31(5):91–100, 2014.

[65] Wei-Yen Hsu. Segmentation-based compression: New frontiers of telemedicine in telecommunication. *Telematics and Informatics*, 32(3):475–485, 2015.

[66] Francisco PM Oliveira and João Manuel RS Tavares. Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, 17(2):73–93, 2014.

[67] Lei Qu, Fuhui Long, and Hanchuan Peng. 3-d registration of biological images and models: Registration of microscopic images and its uses in segmentation and annotation. *Signal Processing Magazine, IEEE*, 32(1):70–77, 2015.

[68] Mustafa Ulutas, Güzin Ulutas, and Vasif V Nabiyev. Medical image security and EPR hiding using shamir's secret sharing scheme. *Journal of Systems and Software*, 84(3):341–353, 2011.

[69] Hitoshi Satoh, Noboru Niki, Kenji Eguchi, Hironobu Ohmatsu, Masahiko Kusumoto, Masahiro Kaneko, and Noriyuki Moriyama. Teleradiology network system on cloud using the web medical image conference system with a new information security solution. In *SPIE Medical Imaging*, pages 86740X–86740X. International Society for Optics and Photonics, 2013.

[70] Chun Kiat Tan, Jason Changwei Ng, Xiaotian Xu, Chueh Loo Poh, Yong Liang Guan, and Kenneth Sheah. Security protection of dicom medical images using dual-layer reversible watermarking with tamper detection capability. *Journal of Digital Imaging*, 24(3):528–540, 2011.

[71] Fusheng Wang, Rubao Lee, Qiaoling Liu, Abulimiti Aji, Xiaodong Zhang, and Joel Saltz. Hadoop-gis: A high performance query system for analytical medical imaging with mapreduce. *Atlanta–USA: Technical report, Emory University*, pages 1–13, 2011.

[72] Nikolaos Koutsouleris, Stefan Borgwardt, Eva M Meisenzahl, Ronald Bottlender, Hans-Jürgen Möller, and Anita Riecher-Rössler. Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the fepsy study. *Schizophrenia bulletin*, page sbr145, 2011.

[73] Kevin W Bowyer. Validation of medical image analysis techniques. *Handbook of medical imaging*, 2:567–607, 2000.

[74] Pierre Jannin, Elizabeth Krupinski, and Simon Warfield. Validation in medical image processing. *IEEE Transactions on Medical Imaging*, 25(11):1405–9, 2006.

[75] Aleksandra Popovic, Matías de la Fuente, Martin Engelhardt, and Klaus Radermacher. Statistical validation metric for accuracy assessment in medical image segmentation. *International Journal of Computer Assisted Radiology and Surgery*, 2(3-4):169–181, 2007.

[76] Colin F Mackenzie, Peter Hu, Ayan Sen, Rick Dutton, Steve Seebode, Doug Floccare, and Tom Scalea. Automatic pre-hospital vital signs waveform and trend data capture fills quality management, triage and outcome prediction gaps. In *AMIA Annual Symposium Proceedings*, volume 2008, page 318. American Medical Informatics Association, 2008.

[77] Michael Bodo, Timothy Settle, Joseph Royal, Eric Lombardini, Evelyn Sawyer, and Stephen W Rothwell. Multimodal noninvasive monitoring of soft tissue wound healing. *Journal of clinical monitoring and computing*, 27(6):677–688, 2013.

[78] Peter Hu, Samuel M Galvagno Jr, Ayan Sen, et al. Identification of dynamic prehospital changes with continuous vital signs acquisition. *Air medical journal*, 33(1):27–33, 2014.

[79] Daniele Apiletti, Elena Baralis, Giulia Bruno, and Tania Cerquitelli. Real-time analysis of physiological data to support medical applications. *IEEE Transactions on Information Technology in Biomedicine*, 13(3):313–321, 2009.

[80] Jie Chen, Edward Dougherty, Semahat S Demir, Charels Friedman, Chung Sheng Li, and Stephen Wong. Grand challenges for multimodal bio-medical systems. *Circuits and Systems Magazine*, 5(2):46–52, 2005.

[81] Nir Menachemi, Askar Chukmaitov, Charles Saunders, and Robert G Brooks. Hospital quality of care: does information technology matter? the relationship between information technology adoption and quality of care. *Health care management review*, 33(1):51–59, 2008.

[82] Catherine M DesRoches, Eric G Campbell, Sowmya R Rao, et al. Electronic health records in ambulatory careâĂŤa national survey of physicians. *New England Journal of Medicine*, 359(1):50–60, 2008.

[83] Jeffrey S McCullough, Michelle Casey, Ira Moscovice, and Shailendra Prasad. The effect of health information technology on quality in us hospitals. *Health Affairs*, 29(4):647–654, 2010.

[84] James M Blum, Heyon Joo, Henry Lee, and Mohammed Saeed. Design and implementation of a hospital wide waveform capture system. *Journal of clinical monitoring and computing*, pages 1–4, 2014.

[85] D Freeman. The future of patient monitoring. *Health management technology*, 30(12):26, 2009.

[86] Bilal Muhsin and Anand Sampath. Systems and methods for storing, analyzing, retrieving and displaying streaming medical data, November 13 2012. US Patent 8,310,336.

[87] David Malan, Thaddeus Fulford-Jones, Matt Welsh, and Steve Moulton. Codeblue: An ad hoc sensor network infrastructure for emergency medical care. In *International workshop on wearable and implantable body sensor networks*, volume 5, 2004.

[88] Alex Page, Ovunc Kocabas, Scott Ames, Muthuramakrishnan Venkitasubramaniam, and Tolga Soyata. Cloud-based secure health monitoring: Optimizing fully-homomorphic encryption for streaming algorithms. In *IEEE Globecom Workshop on Cloud Computing Systems, Networks, and Applications (CCSNA)*, 2014.

[89] Joe Bange, Mark Gryzwa, Kenneth Hoyme, David C Johnson, John LaLonde, and William Mass. Medical data transport over wireless life critical network, July 12 2011. US Patent 7,978,062.

[90] Nadjia Kara and O Andrei Dragoi. Reasoning with contextual data in telehealth applications. In *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMOB)*, pages 69–69. IEEE, 2007.

[91] Gang Li, Jinzhen Liu, Xiaoxia Li, Ling Lin, and Rong Wei. A multiple biomedical signals synchronous acquisition circuit based on over-sampling and shaped signal for the application of the ubiquitous health care. *Circuits, Systems, and Signal Processing*, pages 1–15, 2014.

[92] Amir Bar-Or, J Healey, L Kontothanassis, and JM Van Thong. Biostream: A system architecture for real-time processing of physiological signals. In *IEEE International Conference of the Engineering in Medicine and Biology Society (IEMBS)*, volume 2, pages 3101–3104. IEEE, 2004.

[93] Wullianallur Raghupathi and Viju Raghupathi. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1):3, 2014.

[94] Saif Ahmad, Tim Ramsay, Lothar Huebsch, Sarah Flanagan, Sheryl McDiarmid, Izmail Batkin, Lauralyn McIntyre, Sudhir R Sundaresan, Donna E Maziak, Farid M Shamji, et al. Continuous multiparameter heart rate variability analysis heralds onset of sepsis in adults. *PLoS One*, 4(8):e6642, 2009.

[95] Ary L Goldberger, Luis AN Amaral, Leon Glass, et al. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[96] Eleftheria J Siachalou, Ilias K Kitsas, Konstantinos J Panoulas, et al. ICASP: an intensive-care acquisition and signal processing integrated framework. *Journal of medical systems*, 29(6):633–646, 2005.

[97] Mohammed Saeed, C Lieu, G Raber, and RG Mark. Mimic ii: a massive temporal icu patient database to support research in intelligent patient monitoring. In *Computers in Cardiology*, pages 641–644. IEEE, 2002.

[98] Anton Burykin, Tyler Peck, and Timothy G Buchman. Using âĂIJoff-the-shelfâĂİ tools for terabyte-scale waveform recording in intensive care: Computer system design, database description and lessons learned. *Computer methods and programs in biomedicine*, 103(3):151–160, 2011.

[99] Gómez Adrián, García Eijó Francisco, Martínez Marcela, Analia Baum, Luna Daniel, and González Bernaldo de Quirós Fernán. Mongodb: An open source alternative for hl7-cda clinical documents management. In *Open Source International Conference-CISL*, 2013.

[100] Karamjit Kaur and Rinkle Rani. Managing data in healthcare information systems: Many models, one solution. *Computer*, (3):52–59, 2015.

[101] Srikrishna Prasad and MS Nunifar Sha. Nextgen data persistence pattern in healthcare: Polyglot persistence. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pages 1–8. IEEE, 2013.

[102] Weider D Yu, Manjula Kollipara, Roopa Penmetsa, and Sumalatha Elliadka. Computer engineering department, san jose state university,(silicon valley), california, usa, 95192-0180. In *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pages 476–480. IEEE, 2013.

[103] Manuel Santos and Filipe Portela. Enabling ubiquitous data mining in intensive care: features selection and data pre-processing. In *International Conference on Enterprise Information Systems (ICEIS)*. SciTePress, 2011.

[104] Donald J Berndt, John W Fisher, Alan R Hevner, and James Studnicki. Healthcare data warehousing and quality assurance. *Computer*, 34(12):56–65, 2001.

[105] Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 2011.

[106] Brian D Athey, Michael Braxenthaler, Magali Haas, and Yike Guo. transmart: An open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Summits on Translational Science Proceedings*, 2013:6, 2013.

[107] Mohammed Saeed, Mauricio Villarroel, Andrew T Reisner, et al. Multiparameter intelligent monitoring in intensive care ii (MIMIC-II): a public-access intensive care unit database. *Critical care medicine*, 39(5):952, 2011.

[108] Daniel J Scott, Joon Lee, Ikaro Silva, Shinhyuk Park, George B Moody, Leo A Celi, and Roger G Mark. Accessing the public mimic-ii intensive care relational database for clinical research. *BMC medical informatics and decision making*, 13(1):9, 2013.

[109] Ashwin Belle, Mark A Kon, and Kayvan Najarian. Biomedical informatics for computer-aided decision support systems: a survey. *The Scientific World Journal*, 2013, 2013.

[110] BS Bloom. Crossing the quality chasm: A new health system for the 21st century (committee on quality of health care in america, institute of medicine). *JAMA-Journal of the American Medical Association-International Edition*, 287(5):645, 2002.

[111] Eta S Berner. Clinical decision support systems: state of the art. *AHRQ Publication*, (09-0069), 2009.

[112] Hyoil Han, Han C Ryoo, and Herbert Patrick. An infrastructure of stream data mining, fusion and management for monitored patients. In *IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pages 461–468. IEEE, 2006.

[113] Nadja Bressan, Andrew James, and Carolyn McGregor. Trends and opportunities for integrated real time neonatal clinical decision support. In *IEEE International Conference on Biomedical and Health Informatics (BHI)*, pages 687–690. IEEE, 2012.

[114] Andrew JE Seely, Andrea Bravi, Christophe Herry, et al. Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients? *Critical Care*, 18(2):R65, 2014.

[115] Mina Attin, Gregory Feld, Hector Lemus, Kayvan Najarian, Sharad Shandilya, Lu Wang, Pouya Sabouriazad, and Chii-Dean Lin. Electrocardiogram characteristics prior to in-hospital cardiac arrest. *Journal of clinical monitoring and computing*, pages 1–8, 2014.

[116] Joon Lee and RG Mark. A hypotensive episode predictor for intensive care based on heart rate and blood pressure time series. In *Computing in Cardiology*, pages 81–84. IEEE, 2010.

[117] Jimeng Sun, Daby Sow, Jianying Hu, and Shahram Ebadollahi. A system for mining temporal physiological data streams for advanced prognostic decision support. In *IEEE International Conference on Data Mining (ICDM)*, pages 1061–1066. IEEE, 2010.

[118] Hanqing Cao, Larry Eshelman, Nicolas Chbat, Larry Nielsen, Brian Gross, and Mohammed Saeed. Predicting icu hemodynamic instability using continuous multiparameter trends. In *IEEE International Conference on Engineering in Medicine and Biology Society (EMBS)*, pages 3803–3806. IEEE, 2008.

[119] David L Reich. *Monitoring in anesthesia and perioperative care*. Cambridge University Press, 2011.

[120] A Smolinska, A-Ch Hauschild, RRR Fijten, JW Dallinga, J Baumbach, and FJ van Schooten. Current breathomicsâĂŤa review on data pre-processing techniques and machine learning in metabolomics breath analysis. *Journal of breath research*, 8(2):027105, 2014.

[121] Peter Le Roux, David K Menon, Giuseppe Citerio, et al. Consensus summary statement of the international multidisciplinary consensus conference on multimodality monitoring in neurocritical care. *Intensive care medicine*, 40(9):1189–1209, 2014.

[122] MM Tisdall and M Smith. Multimodal monitoring in traumatic brain injury: current status and future directions. *British journal of anaesthesia*, 99(1):61–67, 2007.

[123] J Claude Hemphill, Peter Andrews, and Michael De Georgia. Multimodal monitoring and neurocritical care bioinformatics. *Nature Reviews Neurology*, 7(8):451–460, 2011.

[124] Alexandros Pantelopoulos and Nikolaos G Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(1):1–12, 2010.

[125] Sebastian Winkler, Michael Schieber, Stephanie Lücke, et al. A new telemonitoring system intended for chronic heart failure patients using mobile telephone technologyâĂŤfeasibility study. *International Journal of Cardiology*, 153(1):55–58, 2011.

[126] Daby Sow, Deepak S Turaga, and Michael Schmidt. Mining of sensor data in healthcare: A survey. In *Managing and Mining Sensor Data*, pages 459–504. Springer, 2013.

[127] John W Davey, Paul A Hohenlohe, Paul D Etter, Jason Q Boone, Julian M Catchen, and Mark L Blaxter. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7):499–510, 2011.

[128] Todd J Treangen and Steven L Salzberg. Repetitive dna and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 13(1):36–46, 2012.

[129] Daniel C Koboldt, Karyn Meltz Steinberg, David E Larson, Richard K Wilson, and Elaine R Mardis. The next-generation sequencing revolution and its impact on genomics. *Cell*, 155(1):27–38, 2013.

[130] Institute of Medicine. *Informatics needs and challenges in cancer research: Workshop summary*. The National Academies Press, 2012.

[131] Eliezer M. Van Allen, Nikhil Wagle, and Mia A. Levy. Clinical analysis and interpretation of cancer genome data. *Journal of Clinical Oncology*, 31(15):1825–1833, 2013.

[132] Adel Tabchy, Cynthia X. Ma, Ron Bose, and Matthew J. Ellis. Incorporating genomics into breast cancer clinical trials and care. *Clinical Cancer Research*, 19(23):6371–6379, 2013.

[133] F. Andre, E. Mardis, M. Salm, J.-C. Soria, L. L. Siu, and C. Swanton. Prioritizing targets for precision cancer medicine. *Annals of Oncology*, 25(12):2295–2303, 2014.

[134] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, 2008.

[135] Jakob LovÃĺn, David A. Orlando, Alla A. Sigova, Charles Y. Lin, Peter B. Rahl, Christopher B. Burge, David L. Levens, Tong Ihn Lee, and Richard A. Young. Revisiting global gene expression analysis. *Cell*, 151(3):476 – 482, 2012.

[136] JT Leek, SB Robert, AI Irizarry, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11:733–739, 2010.

[137] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, 37(1):1–13, 2009.

[138] Purvesh Khatri, Sorin Draghici, G.Charles Ostermeier, and Stephen A. Krawetz. Profiling gene expression using onto-express. *Genomics*, 79(2):266 – 270, 2002.

[139] Sorin Draghici, Purvesh Khatri, Rui P. Martins, G.Charles Ostermeier, and Stephen A. Krawetz. Global functional profiling of gene expression. *Genomics*, 81(2):98 – 104, 2003.

[140] Barry R Zeeberg, Weimin Feng, Geoffrey Wang, et al. Gominer: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*, 4(4):R28, 2003.

[141] Gabriela Bindea, Bernhard Mlecnik, Hubert Hackl, et al. Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8):1091–1093, 2009.

[142] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005.

[143] Sorin Draghici, Purvesh Khatri, Adi Laurentiu Tarca, et al. A systems biology approach for pathway level analysis. *Genome Research*, 17(10):1537–1545, 2007.

[144] Bernhard O Palsson. *Systems biology*. Cambridge university press, 2006.

[145] Daniel Marbach, James C Costello, Robert Kuffner, et al. Wisdom of crowds for robust gene network inference. *Nature Methods*, 9(8):796–804, 2012.

[146] Daniel Marbach, Robert J. Prill, Thomas Schaffter, et al. Revealing strengths and weaknesses of methods for gene network inference. *Proceedings of the National Academy of Sciences*, 107(14):6286–6291, 2010.

[147] Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard Ø. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6):1777–1782, 2007.

[148] Karthik Raman and Nagasuma Chandra. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, 10(4):435–449, 2009.

[149] Christopher S Henry, Matthew DeJongh, Aaron A Best, Paul M Frybarger, Ben Linsay, and Rick L Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, 28:977–982, 2010.

[150] Karin Radrich, Yoshimasa Tsuruoka, Paul Dobson, Albert Gevorgyan, Neil Swainston, Gino Baart, and Jean-Marc Schwartz. Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. *BMC Systems Biology*, 4:114, 2010.

[151] Keren Yizhak, Tomer Benyamini, Wolfram Liebermeister, Eytan Ruppin, and Tomer Shlomi. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 26(12):i255–i260, 2010.

[152] Charles R. Haggart, Jennifer A. Bartell, Jeffrey J. Saucerman, and Jason A. Papin. Chapter twenty-one - whole-genome metabolic network reconstruction and constraint-based modeling. In Malkhey Verma Daniel Jameson and Hans V. Westerhoff, editors, *Methods in Systems Biology*, volume 500 of *Methods in Enzymology*, pages 411 – 433. Academic Press, 2011.

[153] I Thiele, N Swainston, R M T Fleming, et al. A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, 31:419–425, 2012.

[154] Douglas McCloskey, Bernhard Ø Palsson, and Adam M Feist. Basic and applied uses of genome-scale metabolic network reconstructions of escherichia coli. *Molecular Systems Biology*, 9(1), 2013.

[155] Erwin P. Gianchandani, Arvind K. Chavali, and Jason A. Papin. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(3):372–382, 2010.

[156] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotypeâĂŞphenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, 2012.

[157] Weiwen Zhang, Feng Li, and Lei Nie. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 156(2):287–301, 2010.

[158] Anna S. Blazier and Jason A. Papin. Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiology*, 3(299), 2012.

[159] Paul A Jensen and Jason A Papin. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 27(4):541–547, 2011.

[160] Ori Folger, Livnat Jerby, Christian Frezza, Eyal Gottlieb, Eytan Ruppin, and Tomer Shlomi. Predicting selective drug targets in cancer through metabolic networks. *Molecular systems biology*, 7(1), 2011.

[161] Roger L Chang, Li Xie, Lei Xie, Philip E Bourne, and Bernhard Ø Palsson. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS computational biology*, 6(9):e1000938, 2010.

[162] Alexandre Irrthum, Louis Wehenkel, Pierre Geurts, et al. Inferring regulatory networks from expression data using tree-based methods. *PloS one*, 5(9):e12776, 2010.

[163] Robert Küffner, Tobias Petri, Pegah Tavakkolkhah, Lukas Windhager, and Ralf Zimmer. Inferring gene regulatory networks by anova. *Bioinformatics*, 28(10):1376–1382, 2012.

[164] Rui-Sheng Wang, Assieh Saadatpour, and Réka Albert. Boolean modeling in systems biology: an overview of methodology and applications. *Physical biology*, 9(5):055001, 2012.

[165] Robert J Prill, Julio Saez-Rodriguez, Leonidas G Alexopoulos, Peter K Sorger, and Gustavo Stolovitzky. Crowdsourcing network inference: the dream predictive signaling network challenge. *Science signaling*, 4(189):mr7, 2011.

[166] Treenut Saithong, Somkid Bumee, Chalothorn Liamwirat, and Asawin Meechai. Analysis and practical guideline of constraint-based boolean method in genetic network inference. *PloS one*, 7(1):e30232, 2012.

[167] Shawn Martin, Zhaoduo Zhang, Anthony Martino, and Jean-Loup Faulon. Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics*, 23(7):866–874, 2007.

[168] Sorin Draghici, Purvesh Khatri, Rui P Martins, G Charles Ostermeier, and Stephen A Krawetz. Global functional profiling of gene expression. *Genomics*, 81(2):98–104, 2003.

[169] Kay L Pogue-Geile, Julie A Mackey, Ryan D George, Paul G Wood, Kenneth KW Lee, A James Moser, Derrick L Tilman, James Lyons-Weiler, and David C Whitcomb. A new microarray, enriched in pancreas and pancreatic cancer cDNAs to identify genes relevant to pancreatic cancer. *Cancer Genomics-Proteomics*, 1(5-6):371–386, 2004.

[170] Gabriela Bindea, Jérôme Galon, and Bernhard Mlecnik. Cluepedia cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*, 29(5):661–663, 2013.

[171] Vamsi K Mootha, Cecilia M Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, Emma Carlsson, Martin Ridderstråle, Esa Laurila, et al. Pgc-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*, 34(3):267–273, 2003.

[172] Marie-Helene Teiten, Serge Eifes, Simone Reuter, Annelyse Duvoix, Mario Dicato, and Marc Diederich. Gene expression profiling related to anti-inflammatory properties of curcumin in k562 leukemia cells. *Annals of the New York Academy of Sciences*, 1171(1):391–398, 2009.

[173] Wuming Gong, Naoko Koyano-Nakagawa, Tongbin Li, and Daniel J Garry. Inferring dynamic gene regulatory networks in cardiac differentiation through the integration of multi-dimensional data. *BMC bioinformatics*, 16(1):74, 2015.

[174] Katherine C Chen, Laurence Calzone, Attila Csikasz-Nagy, et al. Integrative analysis of cell cycle control in budding yeast. *Molecular biology of the cell*, 15(8):3841–3862, 2004.

[175] Shuhei Kimura, Kaori Ide, Aiko Kashihara, et al. Inference of s-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, 21(7):1154–1163, 2005.

[176] Jutta Gebert, Nicole Radde, and G-W Weber. Modeling gene regulatory networks with piecewise linear differential equations. *European Journal of Operational Research*, 181(3):1148–1165, 2007.

[177] Jason N. Bazil, Karl D. Stamm, Xing Li, Raghuram Thiagarajan, Timothy J. Nelson, Aoy Tomita-Mitchell, and Daniel A. Beard. The inferred cardiogenic gene regulatory network in the mammalian heart. *PLOS ONE*, 2014.

[178] Jason N. Bazil, Feng Qi, and Daniel A. Beard. A parallel algorithm for reverse engineering of biological networks. *Integrative Biology*, 3:1215–1223, 2011.

[179] Ashwin Belle, Soo-Yeon Ji, Wenan Chen, Toan Huynh, and Kayvan Najarian. Rule-based computer aided decision making for traumatic brain injuries. In *Machine Learning in Healthcare Informatics*, pages 229–259. Springer, 2014.

[180] Illhoi Yoo, Patricia Alafaireet, Miroslav Marinov, Keila Pena-Hernandez, Rajitha Gopidi, Jia-Fu Chang, and Lei Hua. Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, 36(4):2431–2448, 2012.

View publication stats