Chronic Diseases and Health Monitoring Big Data: A Survey

RongHeng Lin*, ZeZhou Ye*, Hao Wang†, Budan Wu*
*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China, 100876
†Norwegian University of Science and Technology

***Abstract*—With the technology development of the data science and the network technology, the world has stepped into a big data era. Medical area includes vast amount of data which is suitable for data analysis. Lots of research work in medical big data has been done in recent years, targeting on data collection, data analysis and visualization. However, very few research work provides a full survey of the medical big data on chronic diseases and health monitoring. To get a comprehensive understanding of big data in big medical data, especially in chronic diseases and health monitoring area. This paper investigates recent research efforts, and conducts a comprehensive overview of the work on medical big data. It focuses on the full cycles of the big data processing, which includes medical big data preprocessing, big data tools and algorithm, big data visualization, security issues in big data. It also tried to fill the gap between common big data technology and medical special needs by analyzing detail implementation of medical big data. To the best of our knowledge, this is the first survey that targets the chronic diseases and health monitoring big data technologies.**

***Index Terms*—medical big data, chronic diseases, health monitoring, data analysis, data visualization**

I. Introduction

With the continuous development of medical information, the expansion of medical data accelerates and the coverage area increases. The advent of the big data era creates new opportunities for health and medical domains. As it's known, big data contain 4Vs, which are volume, velocity, variety or veracity. Medical big data has all four features. For volume, a commonly cited statistic from EMC said that 4.4 zettabytes of data existed globally in 2013. That number is predicted as 44 zettabytes by 2020 as it more than doubles each year. For velocity, the health monitoring data is generating every second. For variety, the medical domain contains the many potential big data sources, for instance, the digital medical record, MRI, CT, health monitoring data, genome data. For veracity, a medical data might be incomplete, biased, or even filled with noise. And users can't utilize insights. To analyze the data, data preprocessing, data modeling, data visualization and security are needed. Ambiguous information, repetition, noise and ultra-high dimensions influence the medical data. Therefore it is necessary to preprocess the data. Medical big data preprocessing, integration involves data ETL (Extract, Transform, Load), multi data source integration, and unified data model. Typical algorithms and tools based on existing big data platform are mainly used to make data analysis more convenient and effective. Also, medical big data visualization including treemap, circle-packing, sunburst etc. is the most effective tool when faced with complex medical data and growing medical needs. The type of big data visualization is information visualization, interaction techniques and architectures, modeling techniques, multiresolution methods, visualization algorithms and techniques and volume visualization. Moreover, there are some customized analytics based on visualization which has been listed in this paper. Medical big data is mainly used in clinical data, monitoring and early warning of chronic diseases, daily activities and physical characteristics index detection and collection nowadays. Schematic diagram of this paper organization structure is shown in figure 1.

Manuscript received March 1, 2018; revised August 26, 2018. Corresponding authors: Rongheng Lin, Hao Wang(email: rhlin@bupt.edu.cn, hawa@ntnu.no).

II. Medical big data preprocessing, integration

Twenty-five Semantic Web and Database researchers met at the 2011 STI Semantic Summit in Riga[1] to discuss the opportunities and challenges posed by Big Data for the Semantic Web, Semantic Technologies, and Database communities. The unanimous conclusion was that the greatest shared challenge was not only engineering Big Data, but also doing so meaningfully. Similar to the big data opportunities facing the e-commerce and science and technology communities, the health community is facing challenges of health-and healthcare-related content generated from numerous patient care points of contact, sophisticated medical instruments, and web-based health communities. Two main sources of health big data are genomics-driven big data and payer-provider big data.

With exponential growth in data, enterprises must act to make the most of the vast data landscape-to thoughtfully apply multiple technologies, carefully select key data for specific investigations, and innovatively tailor large integrated datasets including chronic diseases and health monitoring data to support specific queries and analyses. Obviously primary data pools are at the heart of the big-data revolution in healthcare including chronic diseases and health monitoring. Faced with unsustainable costs and enormous amounts of under-utilized data, health care needs more efficient practices, research, and tools to harness the full benefits of personal health and healthcare-related data. Chawla et al.[2] presented the foundations of work that taked a Big Data driven approach towards

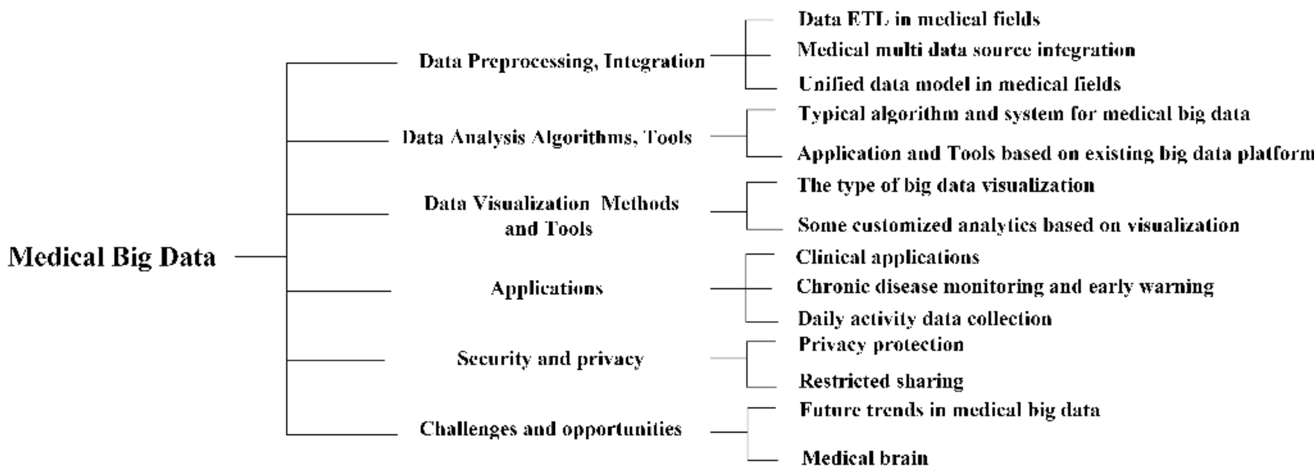

Fig. 1. Schematic diagram of this paper organization structure.

personalized healthcare, and demonstrated its applicability to patient-centered outcomes, meaningful use, and reducing re-admission rates.

To provide a more convenient service and environment of healthcare, Zhang et al.[3] proposed a cyber-physical system for patient-centric healthcare applications and services, called Health-CPS. This system consisted of a data collection layer with a united standard, and a data-oriented service layer. The results showed that the technologies of cloud and big data can be used to enhance the performance of the healthcare systems. Duschka et al.[4] described the novel class of recursive query answering plans. First step, they found a query plan that produced the maximal set of answers from the sources for arbitrary recursive queries. And then they used the presence of functional and full dependencies in the mediated schema.

A. Data ETL in medical fields

ETL (Extract, Transform, Load) is a process in database operations, especially in medical data warehousing. ETL processes are a key component of DWs (Data warehouses) because incorrect or misleading data will produce wrong medical decisions, and therefore, a correct design of these processes at early stages of a DW project is absolutely necessary to improve data quality. ETL tools in medical fields are pieces of software responsible for the extraction of data from several sources like chronic diseases and health monitoring data sources, their cleansing, customization and insertion into a data warehouse. R Kimball et al.[5] introduced the toolkits series of data ETL, such as data quality, projects managing the ETL process and real-time ETL. Vassiliadis et al.[6] have conducted a survey to present the research work in the field of ETL technology in a structured way. The software processes that facilitate the original loading and the periodic refreshment of the data warehouse contents are commonly known as ETL processes.

A typical architecture for supporting ETL in medical/clinical domains is shown in figure 2. The data over which tasks are performed often comes from different sources, such as multiple operational chronic diseases and health monitoring databases across departments within the organization, as well as external vendors. Different sources contain data of varying qualities, and use inconsistent representations, codes, and formats.

A large number of frameworks, systems and tools about ETL have been developed for medical fields. An ETL tool has been implemented to integrate and normalize the clinical data from different operational data sources. The CDW includes online analytical processing (OLAP) and complex network analysis (CNA) components to explore the various clinical relationships. Yao et al.[7] developed a system to intelligently process medical big data and uncover some features of hospital information system user behaviors. It had a data ETL module that efficiently transfers data from the CDR to the BDW in a parallel manner. Moreover, Vassiliadis et al.[8] focused on the problem of the definition of ETL activities and provided formal foundations for their conceptual representation. Juan et al.[9] presented their approach, based on the Unified Modeling Language (UML). They provided the necessary mechanisms for an easy and quick specification of the common operations defined in many ETL processes.

B. Medical multi data source integration

Medical multi data source integration involves combining data residing in different sources including chronic diseases and health monitoring data sources and providing users with a unified view of them. This process becomes significant in a variety of situations, which include both medical and scientific domains. Data integration appears with increasing frequency as the volume and the need to share existing data explodes. There are some typical cases about multi data source integration in medical areas. The current world of isolated research and proprietary data encodings is evolving into a future of standardized medical databases and integrated medical applications, such as clinical decision support systems. Catley et al.[10] explored the use of XML, and its associated Schema Language, to enhance sharing of medical data. XML enables data portability and will reach its full potential when the medical community develops a standardized basis for medical schema content and shares these schemas

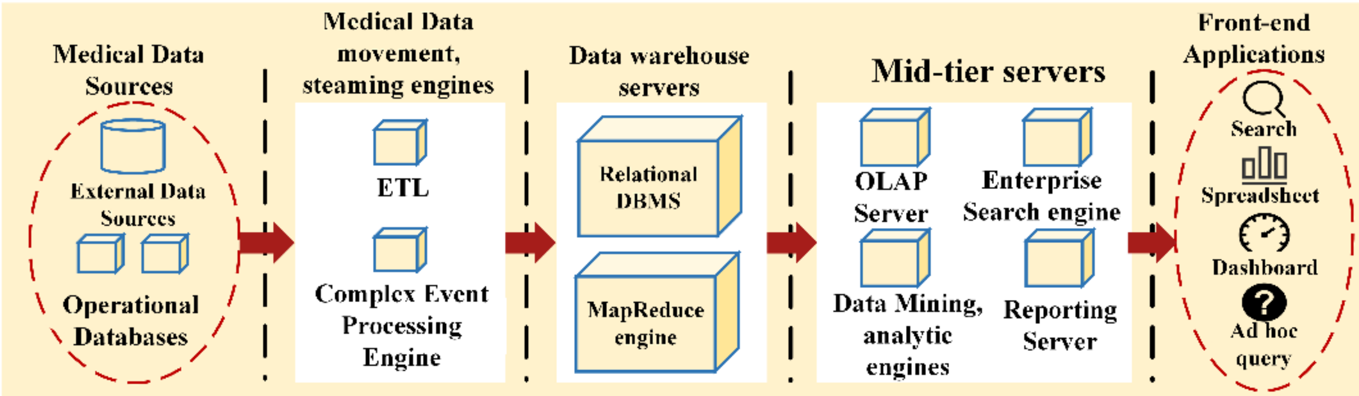

Fig. 2. A typical architecture for supporting ETL in the medical/clinical domains.

in recognized repositories. Their research group is currently harnessing XML's standardization potential by designing a standards-compliant, medical information infrastructure that will allow for seamless integration of all clinical decision support tools. Patients' data is often scattered in a variety of databases and may, in a distributed model, be held across several disparate repositories. Consequently addressing the needs of an evidence-based medicine community presented issues of biomedical data integration, clinical interpretation and knowledge management.

Most applications involve a combination of informational interactions and transactional interactions. Informational interactions involve efficiently aggregating discrete pieces of data that are potentially resident in multiple data sources including chronic diseases and health monitoring data sources, and potentially in multiple data formats. Transactional interactions involve taking a piece of data and orchestrating its propagation to the various underlying applications. This involves coordinating a business process through a formal or informal workflow, managing long-running processes, managing human interactions, handling applications.

A data-integration system provides access to a multitude of data sources through a single mediated schema. A key bottleneck in building such systems has been the laborious manual construction of semantic mappings between the source schemas and the mediated schema. In summary, Doan et al.[11] described LSD, a system that employs and extends current machine-learning techniques to semi-automatically find such mappings. LSD first asks the user to provide the semantic mappings for a small set of data sources, and then uses these mappings together with the sources to train a set of learners. Each learner exploits a different type of information either in the source schemas or in their data.

C. Unified data model in medical fields

Unified Data Model for healthcare is an industry-specific blueprint that provides data warehouse design models, business terminology and analytics to help you quickly develop business applications. It includes prebuilt reporting templates that offer a deeper view of your organization through key performance indicators and other measures. Unified data modeling supports features like document schema of NoSQL databases and reverse engineering of data from an existing database. It also supports visual refactoring of existing databases. Vaguine et al.[12] presented SFCHECK, a standalone software package that features a unified set of procedures for evaluating the structure-factor data obtained from X-ray diffraction experiments and for assessing the agreement of the atomic coordinates with these data. The evaluation is performed completely automatically, and produces a concise PostScript pictorial output similar to that of PROCHECK performed by Laskowski in 1993, and greatly facilitates visual inspection of the results. Many applications of unified information systems require not just spatial data handling but a unified approach to space and time. Worboys[13] began by motivating this requirement with some examples, continued by identifying some of the key issues in this area and then discussed a unified generic model for information.

The Unified Medical Language System is a repository of biomedical vocabularies developed by the US National Library of Medicine. UMLS concepts are not only interrelated, but may also be linked to external resources such as GenBank[14]. A unified timeline model and user interface for multimedia medical databases is described for supporting a timeline-based presentation of information[15]. Various visualization programs permit the user to view data in various ways, including full image views, graphs, and tables. The technology is applied for proof-of-concept to two areas: thoracic oncology and thermal tumor ablation therapy of the brain. This effort is part of the multidisciplinary KMeD project in collaboration with medical research and clinical treatment projects at UCLA.

Real unified data mining/analysis applications call for a framework, which adequately supports knowledge discovery as a multi-step process, where the input of one mining operation can be the output of another. Previous studies, primarily focusing on fast computation of one specific mining task at a time, ignore this vital issue. Motivated by this observation, Johnson et al.[16] developed a unified data model supporting all major mining and analysis tasks. Unified data model consists of three distinct worlds, corresponding to intentional and extensional dimensions, and to data sets like chronic diseases and health monitoring data sets. The notion of dimension is a centerpiece of the model. Equipped with hierarchies, dimensions integrate the output of seemingly dissimilar mining and analysis operations in a clean manner.

III. Medical big data analysis algorithms, tools

A. Typical algorithm and system for medical big data

The advances in information technology have witnessed great progress on healthcare technologies in various domains nowadays. Many recent paradigm shifts-such as the Internet of Things (IoT), Ambient Assisted Living (AAL), e-health and telemedicine[17]. Narrative reports in medical records contain a wealth of information that may augment structured data for managing patient information and predicting trends in diseases. Many health managers and experts believe that with the data, it is possible to easily discover useful knowledge to improve health policies, increase patient safety and eliminate redundancies and unnecessary costs[18]. The emulated data was stored and maintained in HPC Parallel File and BDA platform in HDFS (Hadoop distributed file system). The replication factor for HDFS was set to three for fault tolerance. This requires innovation in a team setting to develop stages in the methodology unique to BDA configurations related to healthcare databases in order for the database to maintain data integrity, as shown in figure 3.

Chen et al.[19] pointed out that the analysis accuracy was reduced when the quality of medical data was incomplete and different regions exhibit unique characteristics of certain regional diseases. Furthermore, patients' information is scattered on the distributed data center. Concurrently, fast progress has been made in clinical analytics. Big data also holds the promise of supporting a wide range of medical and healthcare functions, including among others disease surveillance, clinical decision support and population health management. Without a robust fundamental theory for representation, analysis and inference, a roadmap for uniform handling and analyzing of such complex data remains elusive. Dinov et al.[20] outlined various big data challenges, opportunities, modeling methods and software techniques for blending complex healthcare data, advanced analytic tools, and distributed scientific computing. Using imaging, genetic and healthcare data, they provided examples of processing heterogeneous datasets using distributed cloud services, automated and semi-automated classification techniques, and open-science protocols.

The conventional approaches for health data management have achieved limited success, as they are incapable of handling the huge amount of complex data with high volume, high velocity, and high variety. Fang[21] presented a comprehensive overview of the existing challenges, techniques, and future directions for computational health informatics in the big data era, with a structured analysis of the historical and state-of-the-art methods. On the basis of this material, they identified and discussed the essential prospects lying ahead for computational health informatics in this big data era.

Healthcare big data comprise data from different structured, semi-structured, and unstructured sources. Hossain et al.[22] proposed a healthcare big data framework using voice pathology assessment (VPA) as a case study. In the proposed VPA system, two robust features, MPEG-7 low-level audio and the interlaced derivative pattern, are used for processing the voice or speech signals. The algorithms and systems mentioned above fully illustrate the importance of large data in the medical industry. Large data play an important role in improving medical standards and the quality of human life. How to reduce the hospital cost data by including their primary care and increase business value is suggested in figure 4. It clearly mention the project hospitalization risk for individual patients by calculating the patient risk calculator.

B. Application and Tools based on existing big data platform

Since the completion of the Human Genome project at the turn of the Century, there has been an unprecedented proliferation of genomic sequence data. The exponential evolution of data in health care has brought a lot of challenges in terms of data transfer, storage, computation and analysis. Patel et al.[23] introduced the thought of data in healthcare and the results of various surveys to show the impact of big data. For example, Driscoll et al.[24] provided an overview of cloud computing and big data technologies, and discussed how such expertise can be used to deal with biology's big data sets. As shown in table 1, due to the ever growing number of elderly people coupled with limited resources in terms of medical facilities and personnel in many countries, the burden that conventional healthcare systems carry is becoming heavy.

Big Data is transforming healthcare, business, and ultimately society itself, as e-Health becomes one of key driving factors during the innovation process. Liu et al.[25] investigated Big Data e-Health Service to fulfill the Big Data applications in the e-Health service domain. They explained why the existing Big Data technologies and the like cannot be simply applied to e-Health services directly. Healthcare scientific applications require of deploying hundreds of interconnected sensors to monitor the health status of a host. Augustine et al.[26] analyzed and revealed the benefits of Big Data Analytics and Hadoop in the applications of Healthcare where the data flow to and from is in massive volume. The developing countries like India with huge population faces various problems in the field of healthcare with respect to the expenses, meeting the needs of the economically deprived people, access to the hospitals, research in the field of medicine and especially in the time of spreading epidemics.

With the increasing volumes of information gathered via patient monitoring systems, physicians have been put on increasing pressure for making sophisticated analytical decisions that exploit the various types of data that is being gathered per patient. Mary et al.[27] gave an insight on how to use apache spark for performing predictive analytics with the healthcare data. Large amount of data such as Physician notes, medical prescription, lab and scan reports generated by the healthcare industry is useless until there is a proper method to process this data interactively in real-time. Pita et al.[28] focused on Brazilian Public Health System and on large databases from Ministry of Health and Ministry of Social Development and Hunger Alleviation. They presented their Spark-based approach to data processing and probabilistic record linkage of such databases.

Rathore et al.[29] proposed a Real-time Medical Emergency Response System that involves IoT-based medical sensors

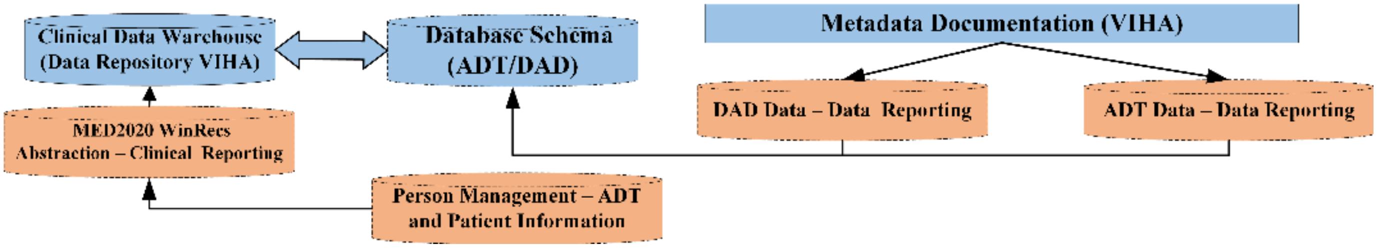

Fig. 3. The workflow of abstracting the ADT and DAD profiles including workflow steps carried out by VIHA staff only

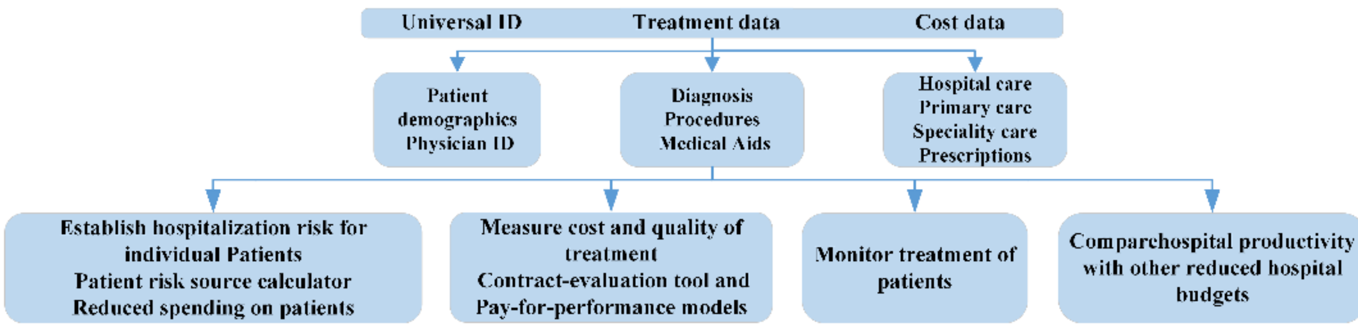

Fig. 4. Big data Plan for healthcare-industry

TABLE I
FEATURES OF WEARABLE 1.0 AND WEARABLE 2.0.

Product name	Category	Comfort index	Machine wash	Accuracy	Sustainability	Physiological index
Smart bracelet	Wearable 1.0	High	No	Low	Yes	Simple
Smart watch	Wearable 1.0	High	No	Low	Yes	Simple
ECG monitoring instrument	Wearable 1.0	Low	No	High	Yes	Simple
Heart rate monitor	Wearable 1.0	Middle	No	High	No	Simple
Fall detection device	Wearable 1.0	Middle	No	Low	No	No
Smart clothing	Wearable 2.0	High	Yes	High	Yes	Complex

deployed on the human body. The data collected from millions of body-attached sensors is forwarded to Intelligent Building for processing and for performing necessary actions using various units. The results showed that the proposed system has the capability of efficiently processing WBAN sensory data from millions of users in order to perform real-time responses in case of emergencies. Most hospitals today are dealing with the big data problem, as they generate and store petabytes of patient records most of which in form of medical imaging. Analyzing such large amounts of biomedical imaging data to enable discovery and guide physicians in personalized care is becoming an important focus of data mining and machine learning algorithms developed for biomedical Informatics (BMI). Neshatpour et al.[30] introduced a scalable and efficient hardware acceleration method using low cost commodity FPGAs that is interfaced with a server architecture through a high speed interface. The results showed promising kernel speedup of up to 27 times for large image data sets. This translates to 7.8 and 1.8 times speedup in an end-to-end Hadoop MapReduce implementation of K-mean s and Laplacian Filtering algorithm, respectively. Idris[31] proposed HDFS as convergence platform and used Hadoop No-SQL solutions to build warehouse for applications real time access to the data. They managed users clinical, personalized, and feedback data to provide clinical, physical, social, and mental health monitoring platform. Their prototype system successfully integrated various technology platforms and provided centralized health monitoring system.

IV. MEDICAL BIG DATA VISUALIZATION METHODS AND TOOLS

Big data visualization methods in areas of chronic diseases and health monitoring including[32]: (1) TreeMap, based on space-filling visualization of hierarchical data; (2) Circle Packing, which also can be included into circles from a higher hierarchy level; (3) Sunburst, which is also an alternative to Treemap, but it uses Treemap visualization, converted to polar coordinate system; (4) Circular Network Diagram, which allows us to represent aggregated data as a set of arcs between analyzed data objects, so that the analyst can get quantity information about relations between objects; (5) Parallel Coordinates, which allows visual analysis to be extended with multiple data factors for different objects; (6) Streamgraph, which is displaced around a central axis, resulting in flowing and organic shape. According to Table 2 and Table 3, now we can clearly classify visualization methods by Big Data classes as well as identify their strengths and weaknesses.

There are many different graphical visualization methods, but multidimensional data visualization is still just a little known and a topical subject of research. Graphical visualization has been already used in different aspects of human activity, but the effectiveness and even applicability of methods can become a real problem with data volumes growth and data production speed. Most of data visualization methods usually does not appear from nothing, but they become a development of earlier existing methods. Big data visualization approaches including: More Than One View per Representation Display, Dynamical Changes in Number of Factors, and Filtering. Typical cases using visualization methods including DnaSP, DNA polymorphism analyses by the coalescent and other methods[40]. DnaSP is a software package for the analysis of DNA polymorphism data. Present version introduces several new modules and features which, among other options allow: (1) handling big data sets; (2) conducting a large number of coalescent-based tests by Monte Carlo computer simulations; (3) extensive analyses of the genetic differentiation and gene flow among populations; (4) analyzing the evolutionary pattern of preferred and unpreferred codons. Data visualization tools are the exact weapons that show us various insights of the collected data.

Almost all fields of study and practice eventually will confront the big-data problem. Visualization has proven effective for not only presenting essential information in vast amounts of data but also driving complex analyses. Big-data analytics and discovery present new research opportunities to the computer graphics and visualization community. Daniel et al.[41] highlighted the latest advancements in solving the big-data problem through visual means, with four articles on new techniques, systems, or applications. Visualization tools and techniques are also increasing in value. Along with the data scientists, a new generation of computer scientists are bringing to bear techniques for working with very large data sets. Expertise in the design of experiments can help cross the gap between correlation and causation.

A. The type of big data visualization

The type of big data visualization are information visualization, interaction techniques and architectures, modelling techniques, multiresolution methods, visualization algorithms and techniques and volume visualization. Big data visualization or data visualization is viewed by many disciplines as a modern equivalent of visual communication. It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information"[42]. The article "Data Visualization: Modern Approaches" gave an overview of seven subjects of data visualization[43]:

- Articles & resources
- Displaying connections
- Displaying data
- Displaying news
- Displaying websites
- Mind maps
- Tools and services

All these subjects are closely related to graphic design and information representation. On the other hand, from a computer science perspective, Frits in 2002 categorized the field into sub-fields[44] [45]:

- Information visualization
- Interaction techniques and architectures
- Modelling techniques
- Multiresolution methods
- Visualization algorithms and techniques
- Volume visualization

B. Some customized analytics based on visualization

Customized analytics based on visualization including visualization of differential gene expression using a novel method of RNA fingerprinting, electronic health records (EHRs), other system and so on. Visual analytics seeks to marry techniques from information visualization with techniques from computational transformation and analysis of data. Information visualization forms part of the direct interface between user and machine, amplifying human cognitive capabilities in six basic ways[46]: (1) by increasing cognitive resources; (2) by reducing search, such as by representing a large amount of data in a small space; (3) by enhancing the recognition of patterns; (4) by supporting the easy perceptual inference of relationships that are otherwise more difficult to induce.

Customized software tools for the representation, visualization, analysis and simulation of medical big data like statnet. The packages implement recent advances in network modeling based on exponential-family random graph models (ERGM). The components of the package provide a comprehensive framework for ERGM-based network modeling, including tools for model estimation, model evaluation, model-based network simulation, and network visualization[47]. There are some customized analytics based on visualization including areas of chronic diseases and health monitoring. Such as: The self-organizing map (SOM) is an efficient tool for visualization of multidimensional numerical data. Vesanto et al.[48] presented an overview and categorization of both old and new methods for the visualization of SOM. The purpose was to give an idea of what kind of information can be acquired from different presentations and how the SOM can best be utilized in exploratory data visualization. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP was realized by Richard et al.[49].

As medical organizations modernize their operations, they are increasingly adopting electronic health records (EHRs) and deploying new health information technology systems that create, gather, and manage their information. As a result, the amount of data available to clinicians, administrators, and researchers in the healthcare system continues to grow at an unprecedented rate[50]. Warner et al.[51] presented a network visualization-based system to explore, analyze, and display the phenotypic patterns that may have remained occult or hidden in basic statistical and pairwise correlation analysis. The

TABLE II
PROPERTIES AND CLASSIFICATION OF VISUALIZATION METHODS.

Method name	Large data volume	Data variety	Data dynamics	Big data class
Treemap	+	-	-	Hierarchical data
Circlepacking	+	-	-	Hierarchical data
Sunburst	+	-	+	Volume + Velocity
Circular network diagram	+	+	-	Volume + Velocity
Parallel coordinates	+	+	+	Volume + Velocity + Variety
Streamgraph	+	-	+	Volume + Velocity

TABLE III
STRENGTHS AND WEAKNESS OF VISUALIZATION METHODS.

Method name	Strength	Weakness
TreeMaps [33] [34]	• Hierarchical grouping clearly shows data relations • Extreme outliers are immediately visible using special color	• Data must be hierarchical, better for analyzing data sets where there is at least one important quantitative dimension with wide variations • Not suitable for examining historical trends and time patterns • Factor used for size calculation cannot have negative values
Circlepacking [35]	• Space-efficient visualization method compared to Treemap	• The same disadvantages as for Treemap
Sunburst [36]	• Easily perceptible by most humans	• The same disadvantages as for Treemap
Circular network diagram [37]	• Allows us to make relative data representation, which can be easily percepted • Within the circle, the resolution varies linearly, increasing with radial position. This makes the center of the circle ideal for compactly displaying summary statistics or indicating points of interest	• Method may end in imperceptible representation form and may need regrouping of data objects on the screen • Objects with smallest parameter weight can be suppressed by larger ones, ending up in total mess onto the diagram
Parallel coordinates [38]	• Factors ordering does not influence total diagram perceptions • Method allows us to analyze both whole data set of objects at once and individual data objects	• Method has limitation to the number of factors, shown at once • Visualization dynamic data end up in changing whole data representation
Streamgraph [39]	• Effective for trends visualization	• Data representation shows one factor • Method depends on data layers sorting

system allowed clinicians and researchers to quickly generate hypotheses and gain deeper understanding of subpopulations. Research conducted by Shehzad et al.[52] focused on the development of an interactive decision support environment in which users can explore epidemic models and their impact. This environment provided a spatiotemporal view where users can interactively utilize mitigative response measures and observe the impact of their decision over time.

V. Medical big data applications

In this section, we discuss the application of big data in the medical industry: clinical data monitoring and analysis; monitoring and early warning of chronic diseases; daily activity data and vital signs detection and collection, then we discuss the existing problems of medical big data and look to the future.

A. Clinical applications

Big data can play an important role in determining causal relationships in patient symptoms, predicting the risks of disease occurrence or recurrence, and improving the quality of primary care. Using systematic reviews to create the classification results for the healthcare big data utilizing[53] showed that the data was split into four broad categories: management and childbirth, clinical decision support, consumer behavior and support services. Two well-known examples of using big data in health are Google Flu Trends and HealthMap. By analyzing the medical status of Americans[54], it has been found that technological advances make it easier to collect and analyze patient record information from multiple sources, as individual patient data may come from different payers, hospitals, laboratories and doctor's offices. By digitizing, consolidating and effectively utilizing big data, significant benefits could be realized from the organization of single-physician offices and multiple health-care facilities to large-scale hospital networks and the care providers responsible for care facilities[55]. A survey by IBM found that big data analytics will help healthcare in many ways, such as evidence-based medicine, genomic analysis, pre-trial fraud analysis, device/remote monitoring, and patient file analysis[56]. Big data helps reduce clinical procedures, research and development, public health waste and inefficiencies. Data collected by sensors and smartphones identify the individual risk factors[57].

B. Chronic disease monitoring and early warning

Diabetic Mellitus is a chronic metabolic disease that requires active and sustained participation by diabetes patients, its careers and physicians for management and good control[58]. Through an 11-year follow-up of Southall diabetes among South Asians and Europeans[59], they found that more than 20% of middle-aged and older South Asians worldwide had diabetes mellitus, whereas South Asian adult diabetics compared to Europeans had a marked increase in their propensity for cardiovascular disease, especially among young people. With big data processing technology, all the technologies related to chronic diseases are extracted, converted, loaded and stored in HFDS. Then different algorithms are used to facilitate data analysis to obtain smarter and more valuable analysis results, and finally to wisdom apply to the management of chronic diseases, promote and improve the management of chronic diseases, as shown in figure 5. By using location-aware healthcare, anyone from the countryside can get the right treatment at a low cost.

Asthma is one of the most prevalent and costly chronic diseases in the United States and can not be cured, but many adverse events can be prevented through appropriate drug use and avoidance of environmental triggers[60]. After studying ED data, Tweeter data, Google data, and sensor data[61], Sudha et al. introduced a new method for using multiple data sources to predict visits to asthma-related emergency departments in a particular area based on a near real-time environment and social media data, with a precision of about 70%. The results might contribute to public health monitoring, ED preparation and targeted patient intervention[62]. Due to environmental pollution, lack of exercise, eating habits and a closed environment, the prevalence of chronic diseases, especially heart disease, has increased dramatically[63]. Atif et al. proposed an active health monitoring system for heart disease patients, through which the e-health system processed real-time data collected by the wearing electronic band to help patients take the initiative measures to prevent abnormal, to help doctors continue to monitor the patient's health status.

One example of using population database monitoring to alert potential high-risk patients, behaviors and outcomes is the UW eHealth-PHINEX project[64] at the University of Wisconsin, which developed a framework to map asthma and diabetes data from EHR to socioeconomic information found in a public health data exchange to help characterize community disease patterns. Context-aware monitoring is an emerging technology that provides real-time, personalized healthcare services and a wealth of big data applications. Forkan et al. proposed BDCaM[65], a general framework for personalized medicine, it used knowledge-based discovery and analyzed data generated in an ambient assisted living (AAL) system to allow context-aware systems to adapt its behavior in runtime, thus applied it in context-aware decision-making process. Through (1) extracting the construct from the theoretical framework, (2) using qualitative methods to extract the construct from the interview data, (3) abstraction of the HITAM-II construction and modeling, and in-depth interviews with undergraduates who self-tracking activity, sleep and diet, Kim et al.[66] illustrated that HITAM-II could successfully describe health-conscious attitudes, behavioral intentions and behaviors from another perspective.

C. Daily activity data collection

The convergence of technologies has contributed to the emergence of life logging as a mainstream activity[67]. With cheap computer storage and advanced sensing technology, it could efficiently sense personal activities, locations and environments. People can digitally record their own daily lives in varying amounts of detail for a variety of purposes.

Quantitative Self (QS) is a self-tracking individual engaged in any type of biological, physical, behavioral or environ-

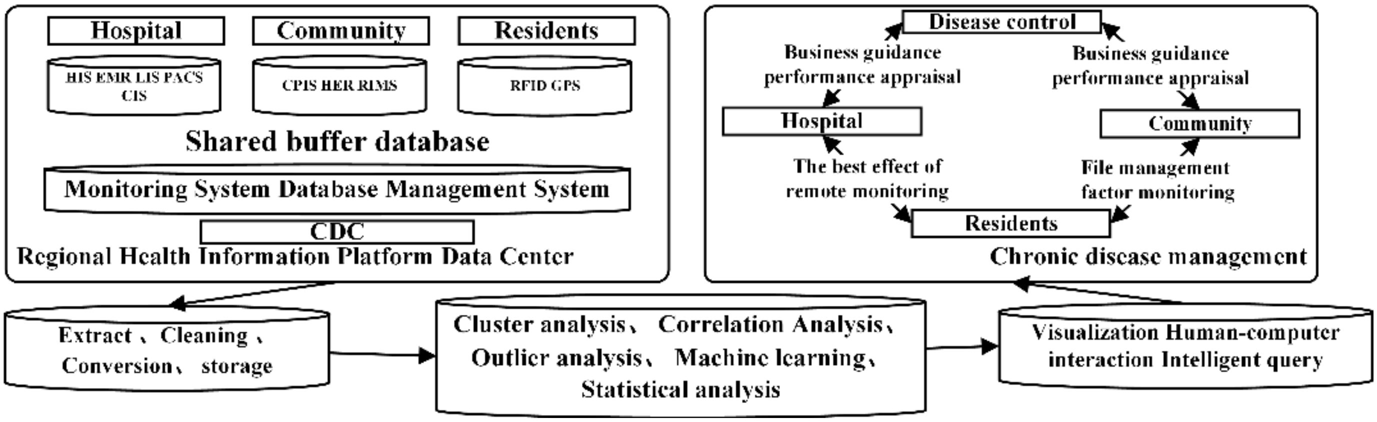

Fig. 5. Chronic disease management based on big data

mental information[68]. Various areas such as weight, energy levels, mood, time use, quality of sleep, etc. can be tracked and analyzed. Nowadays more and more digital devices with associated applications are available for self-tracking. Many of these devices can be worn on or near the body to measure the elements of the user's daily life and activities and to generate data that can be recorded and monitored by the user. Devices include not only digital cameras, smartphones, tablets, watches, wireless scales and sphygmomanometers, but also wearable tapes or patches with embedded sensors, clip-on devices and jewelry that measure body functions or move wirelessly upload data. Among these devices, many of the global positioning devices, gyroscopes, altimeter and accelerometers provide spatial location and are capable of quantized motion. These techniques allow self-followers to collect data and other variables about their emotions, diet, dreams, social contacts, posture, sexual activity, blood chemistry, heart rate, body temperature, exercise patterns, brain function, alcohol, coffee and tobacco consumption. People use sensors and smartphones to track all aspects of health and wellness behaviors, which can be important passive and manual data collection devices. Chen et al.[69] designed "smart clothing" for sustainable health monitoring and collected a variety of unobtrusive human physiological indicators. Wearable computing systems represent one of the most thrust areas used to transform traditional healthcare systems into active systems able to continuously monitor and control the patients' health in order to manage their care at an early stage. Nodes capable of sensing, processing and delivering one or more vital signs could be seamlessly integrated into a wireless personal or body network for health monitoring, such as an electrocardiogram sensor (ECG) that can be used to monitor cardiac activity. Aleksandar et al.[70] discussed the implementation issues and described a health monitoring prototype sensor network that utilized off-the-shelf 802.15.4-compliant network nodes and custom motion and heart activity sensors. Lists of wearable systems are presented in table 4.

The Human Body Sensor Network (BSN) has become another active area, with particular emphasis on enabling real-time decision-making and treatment therapies[71]. Due to the advances in data logging and sensor technology, advanced dynamic systems that measure significant human behavior in daily life have become feasible. One example is activity monitoring (AM)[72], which is based on flow acceleration measurements to assess posture and movement during long-term ($\geq$ 24 hours) measurements in normal daily life[73]. So far, it has been applied to rehabilitation, psychophysiology and cardiology. About sensors fetching data in areas of Chronic Diseases and Health Monitoring, Malasri et al.[74] proposed an architecture called "SNAP" (Sensor Network for Assessment of Patients) and its associated mechanisms for securing medical sensor networks. As shown in left part of figure 6, the mote is connected to several medical sensors, which take samples of the patient's health data when they are activated. The main tasks of the mote are the following: (a) authenticate the patient with the base station using two-tier authentication system; (b) establish a symmetric key with the base station using ECC-based secure key exchange protocol; and (c) communicate with the base station to receive queries and send sensor data.

Daojing et al.[75] suggested a Medical Sensor Network (MSN) should follow a two-tier architecture. In right part of figure 6, the whole network region is partitioned into a collection of cells, each containing a Master node (MN) in charging of a number of Sensor nodes (SNs). SNs are mainly responsible for sensing tasks, while MNs perform more resource-demanding computation and communication tasks. They assumed that MNs and SNs know their affiliated cells. Note that tiering does imply physical clustering, each SN only communicates with a sole MN. MNs and SNs differ significantly in their resources. In particular, MNs have abundant resources in storage and computation.

The problems with wearable systems at present are[76] [77]: First, user needs, perceptions and acceptability, easily using, and unobtrusive wearable systems make it easier for users to accept these devices in daily life; and second, the effectiveness, quality of life, functional ability, and thoroughly test the effectiveness of this technique in home-based monitoring of patients with heart failure; third, interoperability, if clinical software applications can seamlessly collect medical data, interest in these devices will increase; fourth, hardware and software; fifth, the benefits for medical, health, life quality; sixth, cost, psychology, socio-economic barriers; seventh, privacy, moral

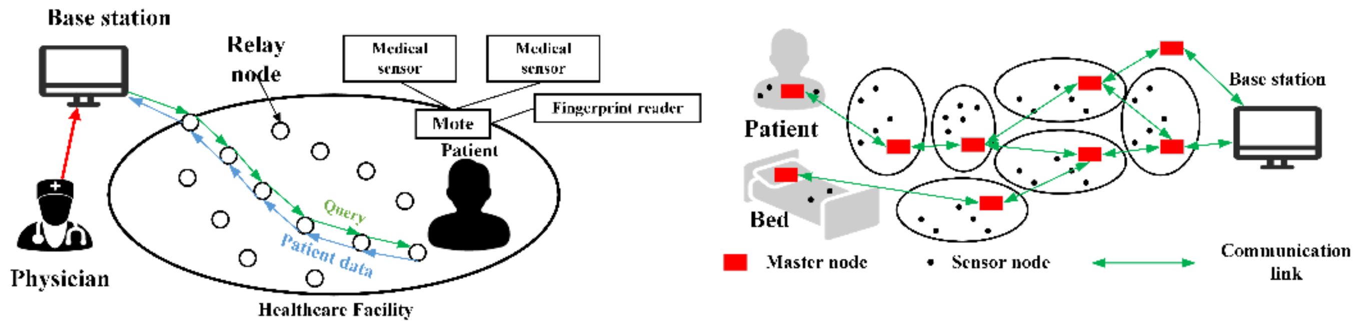

Fig. 6. Medical Sensor Network for Assessment of Patients

and legal barriers; And eighth the impact of wearable systems on society.

Big data has various applications in medical, but the applications mainly focus on three dimension. First, clinical application, which helps determining causal relationships in patient symptoms, improving the quality of primary care. Second, chronic disease monitoring and early warning, because such disease has characters like everlasting, costly, controllable, monitoring these disease could help patients prevent abnormal, help characterize community disease patterns. Third, daily activity data collecting, because activity data can be used for modeling personal habit, with help of wearable systems, it can manage people's health status.

VI. Security and Privacy of Medical Big Data

A. Privacy protection

A 2015 Gartner report noted that data processing technologies have not kept pace with the increase in the volume of healthcare data, and an integrated and trustworthy healthcare analytics solution can facilitate effective decision making in patient care and risk management. The rise of "Big Data" analytics in the private sector poses new challenges for privacy advocates. Through its reliance on existing data and predictive analysis to create detailed individual profiles, Big Data has exploded the scope of personally identifiable information. To respond to Big Data's evolving practices, Crawford et al.[78] examined several existing privacy regimes and explained why these approaches inadequately address current Big Data challenges. They proposed a new approach to mitigating predictive privacy harms-that of a right to procedural data due process by examining due process's role in the Anglo-American legal system and building on previous scholarship about due process for public administrative computer systems.

Storing and sharing of medical data in the cloud environment, where computing resources including storage is provided by a third party service provider, raise serious concern of individual privacy for the adoption of cloud computing technologies. The solution for medical dataset sharing in the cloud should support multiple data accessing paradigms with different privacy strengths. The statistics or cryptography technology alone cannot enforce the multiple privacy demands. Yang et al.[79] proposed a practical solution for privacy preserving medical record sharing for cloud computing. Based on the classification of the attributes of medical records, they used vertical partition of medical dataset to achieve the consideration of different parts of medical data with different privacy concerns. The healthcare sector deals with large volumes of electronic data related to patient services. Srinivasan and Uma[80] described two novel applications that leverage big data to detect fraud, abuse, waste, and errors in health insurance claims, thus reducing recurrent losses and facilitating enhanced patient care. The results indicated that claim anomalies detected using these applications help private health insurance funds recover hidden cost overruns that are not detectable using transaction processing systems.

Nowadays, Social networks and cloud services contain billions of users over the planet. Instagram, Facebook and other networks give the opportunity to share images. Users upload millions of pictures each day. Another domain requires sensitive medical images that retain personal details close to patients. Image perturbation have attracted a great deal of attention. Rahmani et al.[81] dealt with the problem of image perturbation for privacy preserving. The authors built three new systems that consist of hiding small details in pictures by rotating some pixels. Their models used two algorithms: the first one involves a simulation of the firework algorithm, the second system consists of a model of rotation based perturbation using iterated local search algorithm (ILS) with 2 optimization stages. Meanwhile, the third one consists of using the same principle of the previous system except by using the ILS algorithm with 3 optimization stages.

B. Restricted sharing

The vast amount of health data generated and stored around the world each day offers significant opportunities for advances such as the real-time tracking of diseases, and developing health care that is truly personalized. However, capturing, analyzing, and sharing health data is difficult. Heitmueller et al.[82] explored four central questions that policy makers should consider when developing public policy for use of data in health care. They discussed what aspects of big data are relevant for health care and presented a taxonomy of data types and levels of access. They suggested that successful policies require clear objectives and provide examples, discussed barriers to achieving policy objectives based on a recent policy experiment in the United Kingdom, and proposed levers that

TABLE IV
LISTS OF WEARABLE SYSTEMS.

Kind	Author	System description	Applications
Monitor	Santini et al.	Microchip	Autonomous controlled release implant (‘Pharmacy-on-a-chip’) or controllable tablet (‘smart tablet’) for oral drug delivery
	Lanat et al.	Wearable system	Several vital signs and physiological variables to determine the cardiopulmonary activity during emergencies
	Kario et al.	Multifunctional device	Heart rate, physical activity
	Islam et al.	Wellness monitor	Wellness for patients suffering
	Rimet et al.	Bootee	Wearable multiparameter monitor
	Sieg et al.	‘Glucowatch Biographer’	Blood glucose measure
	Vivago	‘Vivago’ (Wellness monitoring)	Vital signs
	Lifeguard	Lifeguard cigarette pack size box	Physiological signs
	Gmez et al.	Pumping, controller and power system	Insulin controller to achieve regulation of blood glucose
	Anliker et al.	‘Amon’ portable telemedical monitor	High-risk cardiac/respiratory patients
	Oliver et al.	Wireless medical monitoring system	Surgery recovering patients
	Ma et al.	Electronic second skin	Antenna, LED, strain gauge, temperature sensor, ECG, EMG, Wireless power coil, RF coil, RF diode
	Chung et al.	A u-healthcare system	ECG, blood pressure patterns transmitted to the hospital
	Xiao et al.	‘MicaZ’ mote based system	Heartbeat, ECG, blood pH, glucose, mobility, walking
	Maqbool et al.	Smart pill	Monitoring system with scintigraphy for measuring whole gut transit
Clothes	Giorgino et al.	Sensorized cloth	A Remote monitoring and control of motor rehabilitation
	Jagos et al.	Shoe	Human gait
	MyHeart	‘MyHeart’ clothing	Atrial fibrillation, ECG, activity
	Weber et al.	Bootee	ECG, GPS, biosensors and bioactuators
	Warrior garment	‘Amon’ portable telemedical monitor	Vital signs
	Borges et al.	Electronic second skin	Fetal movement in the last 4 weeks of the pregnancy
	Luprano et al.	‘Mermoth’ clothes	ECG, respiratory inductance plethysmography, skin, temperature, activity
	Curone et al.	‘ProeTEX’ smart garment	Health-state parameters, environmental variables
	Prochazka et al.	Electrical stimulator garment (glove)	Controlled grasp and hand opening in quadriplegia
	Niazmand et al.	Sens or based smart glove	Parkinson’s disease evaluation
	Jourand et al.	Wearable textile garment	Sudden infant death syndrome
	Bamberg et al.	Shoe	Gait analysis
	Luprano et al.	‘Mermoth’ clothes	ECG, respiratory inductance plethysmography, skin, temperature, activity
	Borges et al.	‘Smart-Clothing’	Fetal movement in the last 4 weeks of the pregnancy
Sensor	Okubo et al.	Home care sensor system	Respiratory diseases
	Coyle et al.	Textile-based sensor (‘Biotex’)	Measuring sweat
	Katsis et al.	‘Aubade’ sensor system	EMG, ECG, respiration, skin conductivity (EDR)
	Miwa et al.	Wearable sensor	Roll-over detection, sleep quality
	Chaudhary et al.	Biosensor	Glucose measures
	Riva et al.	‘Intrepid’ multi-sensor context-aware wearable	Anxiety
	Jovanov et al.	Wireless intelligent sensor system	Heart rate variability for stress measuring
	Wu et al.	RFID ring-type pulse sensor, optical sensor	Pulse and temperature signals, heart rate measures
	Akay et al.	Accelerometer unit	Body motion in healthy subjects, patients with Parkinson’s disease and post
	Schneider et al.	Implan table sensor	Nerve stimulation capable of alleviating acute pain in patients suffering cancer or Parkinson’s disease
Other	Verichip	‘Verichip’	Patient identity
	Adler et al.	Wireless capsule	Endoscopy
	Loew et al.	‘BASUMA’	Chronically ill patients
	Haar et al.	Electronic patch	EMG, arterial oxygen saturation
	Chaudhary et al.	‘Auranet’ personal computing devices	Cognitive impairments
	Fissel et al.	Unique technology toolkit, MEMS system	Membrane prototyped for renal replacement
	Haar et al.	Electronic patch	EMG, arterial oxygen saturation
	Fissel et al.	Unique technology toolkit, MEMS system	Membrane prototyped for renal replacement
	Sung et al.	‘LiveNet’ mobile platform	Accelerometer, ECG, EMG, galvanic skin conductance
	Chang et al.	Portable system	PCG, electrocardiography, body temperature, Bluetooth
	Vuorela et al.	Portable signal recorder	Electrocardiography, bioimpedance and user’s activity
	Beach et al.	In vivo telemetry system	Improvement of the function of an implant evaluated in situ, in blood vessel growth (angiogenesis), reduced inflammation, reduction of foreign body encapsulation

policy makers should consider using to advance data sharing. They argued that the case for data sharing can be won by providing real-life examples of the ways.

VII. Challenges and Opportunities of medical big data

A. Future trends in medical big data

Lots of research work in medical big data has been done in recent years, with the purpose of data collection, data analysis and visualization. However, very few research works provides a full survey of the medical big data on chronic diseases and health monitoring. Chronic disease and health monitoring are the popular direction of medical big data. Therefore, this article focuses on those directions, and we use those as an example to introduce the future trend of medical big data. The use of information technology[83] enables more efficient storage and retrieval of medical records and subsequent data. Before clinicians and clinical researchers can handle big data and translate it into advances in medical science, more data management and statistical analysis training courses are needed. It is crucial for future progress to promote the safety and privacy mechanisms for the sharing of EHR and other medical data[84]. In the future, EHR data may be combined with data from other sources such as social media, environmental information, and genetic sequencing data. In addition, with the globalization of biomedical research and health care, it is important to develop ways to coordinate and calculate big data from different countries, respect national and international policies and legislation, and patient preferences.

Integrating big data and next-generation analytics into clinical and population health research and practice[85] requires not only new sources of data but also new thinking, training and tools. There are still some problems: (1) data assets are growing but there is still a big gap between the quality and quantity of data; (2) medical research has little interest in adopting the new method; and (3) the isolation of the research team. Using big data tools and methods to analyze health informatics data collected at multiple levels, including molecular, tissue, patient, demographic, and human biology, clinical scale, and epidemiological issues. Khoshgoftaar et al.[86] analyzed and examined possible future work in each area and how to combine the data at each level to provide the most promising way to get the most out of health informatics. After investigating 109 case descriptions and covering 63 healthcare institutions, Wang explored the causal relationship between big data analytics capabilities and business value[87], and the path-value-chain of successful big data analytics, and provided new insights to healthcare practitioners on how to constitute big data analytics capabilities for business transformation and offered an empirical basis that can stimulate a more detailed investigation of big data analytics implementation. Nambiar et al.[88] proposed the future of big data analytics on how to improve the overall quality of the healthcare system. The authors noted that the future should focus on preventive health care, population health management and overall health. With the help of big data, based on patient-like analyzes and responses to these methods, more personalized medicines using specific data.

B. Medical brain

The medical brain is an application of AI in the field of artificial intelligence[89]-a combination of big data, supercomputers and machine learning. Through a large number of medical data, through the collection and analysis of medical data, Baidu Medical Brain is designed as an artificial intelligence product that can simulate the doctor's consultation process, communicate with the user, put forward the possible problems according to the user's symptoms, and then pass the verification give final recommendations to improve access and optimize patient access to high-quality medical resources.

VIII. Conclusion

This survey has presented the historical and methodologies developed for the medical big data on chronic diseases and health monitoring. It provides an extensive overview of the medical big data technologies, applications and future works in over 100 publications. These 100 publications are chosen from 300 publications that we collected from recent big data research. During our literature collecting, we setup a literature tree which includes major stages of a complete big data processing. First layer of the tree nodes can become roots of the sub tree. Then we find out about 60 papers for each sub tree. After analyzing the connections between papers, we choose around 100 publications for this survey, which cover the whole life-cycle of the big data processing. This paper covers the medical big data preprocessing, big data tools and algorithm, big data visualization, security issues in big data. To the best of our knowledge, this is the first survey that targets the chronic diseases and health monitoring big data technologies, which might help researcher to get a full understanding of the related research.

Acknowledgment

This work is supported by the Beijing Natural Science Foundation (L171010) and State Grid Corporation of China under the project title:The Improved Core Analysis Algorithms and Utilities for Smart Grid Big Data.

References

[1] Christian Bizer, Peter Boncz, Michael L. Brodie, and Orri Erling. The meaningful use of big data: four perspectivesfour challenges. *ACM SIGMOD Record*, 40(4):56–60, 2012.

[2] Nitesh V. Chawla and Darcy A. Davis. Bringing big data to personalized healthcare: a patient-centered framework. *Journal of general internal medicine*, 28(3):660–665, 2013.

[3] Yin Zhang, Meikang Qiu, Chun-Wei Tsai, Mohammad Mehedi Hassan, and Atif Alamri. Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Systems Journal*, 11(1):88–95, 2017.

[4] Oliver M. Duschka, Michael R. Genesereth, and Alon Y. Levy. Recursive query plans for data integration. *The Journal of Logic Programming*, 43(1):49–73, 2000.

[5] Ralph Kimball and Joe Caserta. *The Data Warehouse? ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. John Wiley & Sons, 2011.

[6] Panos Vassiliadis. A survey of ExtracttransformLoad technology. *International Journal of Data Warehousing and Mining (IJDWM)*, 5(3):1–27, 2009.

[7] Qin Yao, Yu Tian, Peng-Fei Li, Li-Li Tian, Yang-Ming Qian, and Jing-Song Li. Design and development of a medical big data processing system based on hadoop. *Journal of medical systems*, 39(3):23, 2015.

[8] Panos Vassiliadis, Alkis Simitsis, and Spiros Skiadopoulos. Conceptual modeling for ETL processes. In *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, pages 14–21. ACM, 2002.

[9] Juan Trujillo and Sergio Lujn-Mora. A UML based approach for modeling ETL processes in data warehouses. In *International Conference on Conceptual Modeling*, pages 307–320. Springer, 2003.

[10] Christina Catley and Monique Frize. Design of a health care architecture for medical data interoperability and application integration. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 3, pages 1952–1953. IEEE, 2002.

[11] AnHai Doan, Pedro Domingos, and Alon Y. Halevy. Reconciling schemas of disparate data sources: A machine-learning approach. In *ACM Sigmod Record*, volume 30, pages 509–520. ACM, 2001.

[12] Alexei A. Vaguine, Jean Richelle, and S. J. Wodak. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallographica Section D: Biological Crystallography*, 55(1):191–205, 1999.

[13] Michael F. Worboys. A unified model for spatial and temporal information. *The Computer Journal*, 37(1):26–34, 1994.

[14] Olivier Bodenreider. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270, 2004.

[15] John David N Dionisio, Alfonso F Cardenas, Ricky K Taira, Denise R Aberle, Wesley W Chu, Michael F McNitt-Gray, Jonathan Goldin, and Robert B Lufkin. A unified timeline model and user interface for multimedia medical databases. *Computerized Medical Imaging and Graphics*, 20(4):333–346, 1996.

[16] Theodore Johnson, Laks VS Lakshmanan, and Raymond T. Ng. The 3w model and algebra for unified data mining. In *VLDB*, pages 21–32, 2000.

[17] Davy Preuveneers, Yolande Berbers, and Wouter Joosen. SAMURAI: a batch and streaming context architecture for large-scale intelligent applications and environments. *Journal of Ambient Intelligence and Smart Environments*, 8(1):63–78, 2016.

[18] Mu-Hsing Kuo, Tony Sahama, Andre W. Kushniruk, Elizabeth M. Borycki, and Daniel K. Grunwell. Health big data analytics: current perspectives, challenges and potential solutions. *International Journal of Big Data Intelligence*, 1(1-2):114–126, 2014.

[19] Min Chen, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang. Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access*, 2017.

[20] Ivo D. Dinov. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *Gigascience*, 5(1):12, 2016.

[21] Ruogu Fang, Samira Pouyanfar, Yimin Yang, Shu-Ching Chen, and S. S. Iyengar. Computational health informatics in the big data age: a survey. *ACM Computing Surveys (CSUR)*, 49(1):12, 2016.

[22] M. Shamim Hossain and Ghulam Muhammad. Healthcare Big Data Voice Pathology Assessment Framework. *IEEE Access*, 4:7806–7815, 2016.

[23] Jigna Ashish Patel and Priyanka Sharma. Big data for better health planning. In *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on*, pages 1–5. IEEE, 2014.

[24] Aisling ODriscoll, Jurate Daugelaite, and Roy D. Sleator. Big data, Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, 46(5):774–781, 2013.

[25] W. Liu and E. K. Park. Big data as an e-health service. In *Computing, Networking and Communications (ICNC), 2014 International Conference on*, pages 982–988. IEEE, 2014.

[26] D. Peter Augustine. Leveraging Big Data analytics and Hadoop in developing Indias healthcare services. *International Journal of Computer Applications*, 89(16):44–50, 2014.

[27] J. Archenaa and EA Mary Anita. Interactive Big Data management in healthcare using Spark. In *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC16)*, pages 265–272. Springer, 2016.

[28] Robespierre Pita, Clicia Pinto, Pedro Melo, Malu Silva, Marcos Barreto, and Davide Rasella. A Spark-based Workflow for Probabilistic Record Linkage of Healthcare Data. In *EDBT/ICDT Workshops*, pages 17–26, 2015.

[29] M. Mazhar Rathore, Awais Ahmad, Anand Paul, Jiafu Wan, and Daqiang Zhang. Real-time medical emergency response system: exploiting IoT and big data for public health. *Journal of medical systems*, 40(12):283, 2016.

[30] Katayoun Neshatpour, Arezou Koohi, Farnoud Farahmand, Rajiv Joshi, Setareh Rafatirad, Avesta Sasan, and Houman Homayoun. Big biomedical image processing hardware acceleration: A case study for K-means and image filtering. In *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*, pages 1134–1137. IEEE, 2016.

[31] Muhammad Idris, Shujaat Hussain, Mahmood Ahmad, and Sungyoung Lee. Big data service engine (bise): Integration of big data technologies for human centric wellness data. In *Big Data and Smart Computing (BigComp), 2015 International Conference On*, pages 244–248. IEEE, 2015.

[32] Evgeniy Yur'evich Gorodov and Vasiliy Vasil'evich Gubarev. Analytical review of data visualization methods in application to big data. *Journal of Electrical and Computer Engineering*, 2013:22, 2013.

[33] Mark Bruls, Kees Huizing, and Jarke J Van Wijk. Squarified treemaps. In *Data visualization 2000*, pages 33–42. Springer, 2000.

[34] Benjamin B Bederson, Ben Shneiderman, and Martin Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. In *The craft of information visualization*, pages 257–278. Elsevier, 2003.

[35] Kenneth Stephenson. *Introduction to circle packing: The theory of discrete analytic functions*. Cambridge University Press, 2005.

[36] Nick Cawthon and Andrew Vande Moere. The effect of aesthetic on the usability of data visualization. In *Information Visualization, 2007. IV'07. 11th International Conference*, pages 637–648. IEEE, 2007.

[37] Marielle S Köberlin, Berend Snijder, Leonhard X Heinz, Christoph L Baumann, Astrid Fauster, Gregory I Vladimer, Anne-Claude Gavin, and Giulio Superti-Furga. A conserved circular network of coregulated lipids modulates innate immune responses. *Cell*, 162(1):170–183, 2015.

[38] Stephen Few. Multivariate analysis using parallel coordinates. *Perceptual edge*, pages 1–9, 2006.

[39] Lee Byron and Martin Wattenberg. Stacked graphs–geometry & aesthetics. *IEEE transactions on visualization and computer graphics*, 14(6), 2008.

[40] Julio Rozas, Juan C. Snchez-DelBarrio, Xavier Messeguer, and Ricardo Rozas. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19(18):2496–2497, 2003.

[41] Daniel Keim, Huamin Qu, and Kwan-Liu Ma. Big-data visualization. *IEEE Computer Graphics and Applications*, 33(4):20–21, 2013.

[42] Michael Friendly and Daniel J. Denis. Milestones in the history of thematic cartography, statistical graphics, and data visualization. *URL http://www. datavis. ca/milestones*, 32, 2001.

[43] Muzammil Khan and Sarwar Shah Khan. Data and information visualization methods, and interactive mechanisms: A survey. *International Journal of Computer Applications*, 34(1):1–14, 2011.

[44] Frits H. Post, Gregory Nielson, and Georges-Pierre Bonneau. *Data visualization: The state of the art*, volume 713. Springer Science & Business Media, 2012.

[45] Vitaly Friedman. Data visualization: Modern approaches. *Smashing Magazine*, 2, 2007.

[46] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[47] Mark S. Handcock, David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of statistical software*, 24(1):1548, 2008.

[48] Juha Vesanto. SOM-based data visualization methods. *Intelligent data analysis*, 3(2):111–126, 1999.

[49] Christian WB Bachem, Rutger S. Hoeven, Steef M. Bruijn, Dick Vreugdenhil, Marc Zabeau, and Richard GF Visser. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *The plant journal*, 9(5):745–753, 1996.

[50] Jesus J. Caban and David Gotz. *Visual analytics in healthcareopportunities and research challenges*. Oxford University Press, 2015.

[51] Jeremy L. Warner, Joshua C. Denny, David A. Kreda, and Gil Alterovitz. Seeing the forest through the trees: uncovering phenomic complexity through interactive network visualization. *JAMIA*, 22(2):324–329, 2015.

[52] Shehzad Afzal, Ross Maciejewski, and David S. Ebert. Visual analytics decision support environment for epidemic modeling and response evaluation. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 191–200. IEEE, 2011.

[53] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health affairs (Project Hope)*, 33(7):1123–1131, 2014.

[54] Basel Kayyali, David Knott, and Steve Van Kuiken. The big-data revolution in US health care: Accelerating value and innovation. *Mc Kinsey \& Company*, 2(8):1–13, 2013.

[55] Wullianallur Raghupathi and Viju Raghupathi. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2, February 2014.

[56] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Byers. Big data: The next frontier for innovation, competition, and productivity. May 2011.

[57] Meredith A. Barrett, Olivier Humblet, Robert A. Hiatt, and Nancy E. Adler. Big Data and Disease Prevention: From Quantified Self to Quantified Communities. *Big Data*, 1(3):168–175, 2013.

[58] Eleni I. Georga, Vasilios C. Protopappas, Christos V. Bellos, and Dimitrios I. Fotiadis. Wearable systems and mobile applications for diabetes disease management. *Health and Technology*, 4(2):101–112, August 2014.

[59] H.M. Mather, N Chaturvedi, and J.H. Fuller. *Mortality and morbidity from diabetes in South Asians and Europeans: 11- year follow-up of the Southall Diabetes Survey, London, UK*, volume 15. 1998.

[60] Iraq. guidelines for the diagnosis and management of asthma, 2012.

[61] S. Ram, W. Zhang, M. Williams, and Y. Pengetnze. Predicting Asthma-Related Emergency Department Visits Using Big Data. *IEEE Journal of Biomedical and Health Informatics*, 19(4):1216–1223, July 2015.

[62] Murdoch TB and Detsky AS. The inevitable application of big data to health care. *JAMA*, 309(13):1351–1352, 2013.

[63] A. Naseer, B. Y. Alkazemi, and E. U. Waraich. A big data approach for proactive healthcare monitoring of chronic

patients. In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 943–945, July 2016.

[64] T. W. Guilbert, B. Arndt, J. Temte, A. Adams, W. Buckingham, A. Tandias, C. Tomasallo, H. A. Anderson, and L. P. Hanrahan. The theory and application of UW ehealth-PHINEX, a clinical electronic health record-public health information exchange. *WMJ : official publication of the State Medical Society of Wisconsin*, 111(3):124–133, June 2012.

[65] A. R. M. Forkan, I. Khalil, A. Ibaida, and Z. Tari. BDCaM: Big Data for Context-Aware Monitoring #8212;A Personalized Knowledge Discovery Framework for Assisted Healthcare. *IEEE Transactions on Cloud Computing*, 5(4):628–641, 2017.

[66] Jeongeun Kim. A Qualitative Analysis of User Experiences With a Self-Tracker for Activity, Sleep, and Diet. *Interactive Journal of Medical Research*, 3(1), March 2014.

[67] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. LifeLogging: Personal Big Data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, June 2014.

[68] Swan M. The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. - Abstract - Europe PMC, June 2013.

[69] Min Chen, Yujun Ma, Jeungeun Song, Chin-Feng Lai, and Bin Hu. Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring. *Mobile Networks and Applications*, 21(5):825–845, October 2016.

[70] Aleksandar Milenkovi, Chris Otto, and Emil Jovanov. Wireless sensor networks for personal health monitoring: Issues and an implementation. *Computer Communications*, 29(13):2521–2533, August 2006.

[71] C. C. Y. Poon, B. P. L. Lo, M. R. Yuce, A. Alomainy, and Y. Hao. Body Sensor Networks: In the Era of Big Data and Beyond. *IEEE Reviews in Biomedical Engineering*, 8:4–16, 2015.

[72] J. B. Bussmann, J. H. Tulen, E. C. van Herel, and H. J. Stam. Quantification of physical activities by means of ambulatory accelerometry: a validation study. *Psychophysiology*, 35(5):488–496, September 1998.

[73] J. B. J. Bussmann, W. L. J. Martens, J. H. M. Tulen, F. C. Schasfoort, H. J. G. van den Berg-Emons, and H. J. Stam. Measuring daily behavior using ambulatory accelerometry: The Activity Monitor. *Behavior Research Methods, Instruments, & Computers*, 33(3):349–356, 2001.

[74] Kriangsiri Malasri and Lan Wang. Addressing security in medical sensor networks. In *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*, pages 7–12. ACM, 2007.

[75] Daojing He, Chun Chen, Sammy Chan, Jiajun Bu, and Athanasios V Vasilakos. Retrust: Attack-resistant and lightweight trust management for medical sensor networks. *IEEE transactions on information technology in biomedicine*, 16(4):623–632, 2012.

[76] Marie Chan, Daniel Estve, Jean-Yves Fourniols, Christophe Escriba, and Eric Campo. Smart wearable systems: Current status and future challenges. *Artificial Intelligence in Medicine*, 56(3):137–156, November 2012.

[77] Mohamed R Mahfouz, Gary To, and Michael J Kuhn. Smart instruments: wireless technology invades the operating room. In *Biomedical Wireless Technologies, Networks, and Sensing Systems (BioWireleSS), 2012 IEEE Topical Conference on*, pages 33–36. IEEE, 2012.

[78] Kate Crawford and Jason Schultz. Big data and due process: Toward a framework to redress predictive privacy harms. *BCL Rev.*, 55:93, 2014.

[79] Ji-Jiang Yang, Jian-Qiang Li, and Yu Niu. A hybrid solution for privacy preserving medical data sharing in the cloud environment. *Future Generation Computer Systems*, 43:74–86, 2015.

[80] Uma Srinivasan and Bavani Arunasalam. Leveraging big data analytics to reduce healthcare costs. *IT professional*, 15(6):21–28, 2013.

[81] Amine Rahmani, Abdelmalek Amine, Reda Mohamed Hamou, Mohamed Elhadi Rahmani, and Hadj Ahmed Bouarara. Privacy preserving through fireworks algorithm based model for image perturbation in big data. *International Journal of Swarm Intelligence Research (IJSIR)*, 6(3):41–58, 2015.

[82] Axel Heitmueller, Sarah Henderson, Will Warburton, Ahmed Elmagarmid, Alex Sandy Pentland, and Ara Darzi. Developing public policy to advance the use of big data in health care. *Health Affairs*, 33(9):1523–1530, 2014.

[83] Zhongheng Zhang. Big data and clinical research: focusing on the area of critical care medicine in mainland China. *Quantitative Imaging in Medicine and Surgery*, 4(5):426–429, 2014.

[84] M. K. Ross, Wei Wei, and L. Ohno-Machado. Big Data and the Electronic Health Record. *Yearbook of Medical Informatics*, 9(1):97–104, 2014.

[85] Harlan M. Krumholz. Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System. *Health Affairs*, 33(7):1163–1170, 2014.

[86] Taghi M. Khoshgoftaar and Randall Wald. A review of data mining using big data in health informatics. *Journal Of Big Data*, 1(1):2, December 2014.

[87] Yichuan Wang and Nick Hajli. Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, 70:287–299, January 2017.

[88] R. Nambiar, R. Bhardwaj, A. Sethi, and R. Vargheese. A look at challenges and opportunities of Big Data analytics in healthcare. In *2013 IEEE International Conference on Big Data*, pages 17–22, October 2013.

[89] Juan Sun. Baidu hopes big data can improve care for patients|China|chinadaily.com.cn, 2016.