Journal of Computer Sciences and Applications, 2018, Vol. 6, No. 1, 48-55
Available online at http://pubs.sciepub.com/jcsa/6/1/7

DOI:10.12691/jcsa-6-1-7

Big Data Analytics in Biometrics and Healthcare

Lidong Wang[1,*], Cheryl Ann Alexander[2]

[1]Institute for Systems Engineering Research, Mississippi State University, Vicksburg, MS, USA
[2]Institute for Information Technology Innovations & Smart Healthcare, Vicksburg, MS, USA
*Corresponding author: lidong@iser.msstate.edu

Abstract Big Data analytics has been used in biometric systems and healthcare. Biometrics is a powerful tool used in healthcare for identification, insurance, and management, etc. There are many computational resources on the cloud, which makes the cloud a strong platform for biometric systems, healthcare systems, and Big Data analytics. Big data, Big Data analytics, and general information security and privacy in big data are presented in this paper. Cloud computing in biometric systems, Big Data analytics in biometrics, and Big Data analytics in healthcare are also presented. The challenges of biometric data on the cloud and the challenges of Big Data analytics in healthcare are discussed.

Keywords: *big data, Big Data analytics, cloud computing, biometrics, healthcare, deep learning, information security, privacy, precision medicine, personalized medicine*

Cite This Article: Lidong Wang, and Cheryl Ann Alexander, "Big Data Analytics in Biometrics and Healthcare." *Journal of Computer Sciences and Applications*, vol. 6, no. 1 (2018): 48-55. doi: 10.12691/jcsa-6-1-7.

1. Introduction

Biometrics such as fingerprints, retina, and face recognition is used primarily for identification and authentication. Smartphones, for example, now use fingerprint recognition in place of passwords for security. Biometric data, however, comes with its own set of limitation is also subject to a higher level of privacy protection than traditional forms of identification [1]. Along with biometric data, data is being produced via untraditional forms such as sensors, smartphones, genomic data, and healthcare electronic medical records (EMR). Data from such sources is rapidly rising and much of this new data being generated is unstructured, especially in the form of provider notes, clinical notes, vital sign records, etc. Also, metadata is rapidly increasing [2]. Basic information collected by one device may not be subject to legal protections, however, privacy issues may arise once it is combined with other information. Big Data will benefit medical and epidemiological research both in data processing and data protection [3].

A global open standard system for radio frequency identification (RFID) has laid the foundation for many architectures supporting Internet of Things (IoT). RFID technologies use machine-to-machine (M2M) transmissions. M2M transmissions share information without any special configuration or other setup requirements. Healthcare wearables contain wireless sensors embedded in the device and worn on the body. M2M technologies and healthcare apps, as well as healthcare wearables could improve patient outcomes, reduce health expenditures, and allow providers to deliver care in more patient-friendly ways. Healthcare providers can monitor patient blood pressures, respiration rates, and a variety of other biometric information continuously and remotely, as well as through wearables. Healthcare providers can identify trends and make better decisions by analyzing the continuous data. Microsoft's health-tracking wearable called Microsoft Band incorporates exotic sensors like galvanic skin response; it is now possible to predict a user's emotional state by adding heart rate and temperature information [3]. All these are related to data analytics because RFID-based IoT and M2M in healthcare apps generate big data.

The purpose of data protection is to ensure information privacy and security, which is a difficult challenge for the public sector because government agencies need to implement policy changes that deal with threats in real time. The collection and analysis of personal data by governments inevitably raises concerns about civil liberties. The same concerns have been raised about social websites such as Facebook which can also misuse personal data. Individuals and organizations are likely to experience both benefits and threats inherent in big data [4].

Authentication is required in stored database systems so that only authorized users can access the data and related cloud infrastructures. Single-factor authentication is insufficient for some business transactions with high-risk big data. A simple approach is using complex authentication methods with multi-factor, multi-dimensional password, or multi-level authentication. This may help large cloud user groups to perform computation on big data. Strong password generation and authentication technique is used to perform computation on big data using cloud services. It utilizes multi-source, multi-sampling, multi-level, multi-dimensional, multi-factor, multi-mode, and multi-layered user's biometric data and personal data for generating a password to identify the user [5].

The purpose of this paper is to present advances of Big Data analytics in biometric and healthcare. The organization of the paper is as follows: the second section introduces big data and Big Data analytics; Section 3 introduces information security and privacy in big data; Section 4 presents Big Data analytics, cloud computing and biometrics; Section 5 presents Big Data analytics in healthcare; and Section 6 is the conclusion.

2. Big Data and Big Data Analytics

Big data management and analytics research is proceeding in the following directions: 1) building infrastructure and high-performance computing techniques for big data; 2) techniques such as integrating multiple data sources and indexing and querying big data; and 3) data analytics techniques that manipulate and analyze big data to extract nuggets. With respect to building infrastructures, technologies such as Hadoop and MapReduce as well as Storm are being developed for managing substantial amounts of data in the cloud. In addition, main memory data management techniques have advanced so that a few terabytes of data can be managed in main memory. Hive, Cassandra, and NoSQL databases have been developed for managing petabytes of data [6]. Table 1 shows the domains, software, features and some approaches of Big Data.

The growth of big data and the expansion of data analytics platforms such as Hadoop and NoSQL are creating new opportunities for cloud computing to become a key enabler of Big Data analytics. The integration of Big Data analytics and cloud computing has led to another service model known as Analytics as a Service (AaaS). This model will provide companies with faster and more scalable ways to integrate, transform, analyze, and visualize various types of data including structured, semi-structured, and unstructured data in real time. Big Data analytics enables cloud services [8]. Table 2 lists some platforms and tools for Big Data analytics.

There are at least three paradoxes which lie at the heart of Big Data. Although it is accepted that Big Data has beneficial outcomes, there is an argument that the evidence of benefits has not been balanced with the limits or undesirable outcomes of Big Data. First, Big Data suffers from a transparency paradox, whereas the operations of big data itself are almost entirely shrouded in legal and commercial secrecy. Second, Big Data creates an identity paradox, whereas individuals seek control over the formation of their identity; Big Data therefore constitutes identity. In spatial terms, identity formation may occur when an individual's geolocational information is surveilled and analyzed in biopolitical spatial profiling. Third, Big Data may be called a power paradox because although it is a powerful tool that will revolutionize our lives, Big Data sensors, tools and applications are in the hands of powerful institutions rather than ordinary people. Big Data may therefore be exacerbating inequalities and exploitation, rather than ameliorating them [10].

It is often assumed that Big Data techniques are unbiased because of the scale of the data and the techniques are implemented through algorithmic systems. However, it is a mistake to assume they are objective simply because they are data-driven. The algorithmic systems of Big Data use sophisticated processes. These processes need inputs. Data sets that lack information or disproportionately represent certain populations result in skewed algorithmic systems that effectively encode discrimination because of the flawed nature of the initial inputs. Data availability, access to technology, and participation in the digital ecosystem vary greatly due to economic, linguistic, structural or socioeconomic barriers. Programmers and data scientists may inadvertently or unconsciously design, train, or use Big Data systems with biases. Therefore, an important factor for implementing the "equal opportunity by design" principle is engaging with the area of "bias mitigation". With the use of Big Data and machine learning, it will remain important not to place too much reliance on new systems without continuously testing the inputs and mechanics behind them and the results they produce [11]. A summary of Big Data challenges and solutions is shown in Table 3 [12].

Table 1. Big Data Domains, Features, Software and Analytical Approaches [7]

Domains	Software	Features/Tasks/Outcomes
Platforms and Physical Capacity	Hadoop system	Parallel distributed, multicore, data sharing and integrity, clustering
Data Storage	HDFS distributed file systems	Storage, backup, acquisition, retrieval, redundancy removal
Preprocessing for Fundamental Data Analysis	R/pbdR, SAS JMP, SPSS, Matlab	Data cleaning, aggregation/integration, extracting, visualization
Advanced Computational Approaches	R/pbdR, SAS JMP, Revolution Analytics, Tableau Software, Matlab	Modeling, computing, analysis, interpretation

Table 2. Platforms and Tools for Big Data Analytics [9]

Hadoop System	An open source platform for handling big data; the most popular implementation of the MapReduce methodology.
Cloud Storage	It uses a network of remote servers. The servers are hosted on the Internet to store, manage, and process data.
Column-oriented Databases	They store datasets as segments of columns of data rather than as rows of data; allow huge compression and very fast query times.
NoSQL Databases	In relational databases, tabular relations are used while a NoSQL (Not only SQL) database uses a different method to store and retrieve data including structured, semi-structured, and unstructured data.
Hive	It is a runtime Hadoop support architecture that leverages Structure Query Language (SQL) with the Hadoop platform.
PIG	Instead of a "SQL-like" language, it allows for query execution over data stored on a Hadoop cluster.
Cassandra	It is a distributed database system and is designated as a top-level project modelled to handle big data distributed across many utility servers.

Table 3. Big Data Challenges and Solutions

Process	Challenges	Solutions
Data access and Collection	• Efficiently obtain detailed data for many agents • Easy access to data with standardized formats • Protocols on security, privacy, and data rights	• Web scraping • Web traffic and communications monitoring • Sensors
Data Storage	• Data reliability • Warehousing • Tools for data storage and integration of various big datasets	• Saving essential information only and updating it in real time • SQL, NoSQL, Apache Hadoop
Data Processing	• Use non-numeric data for quantitative analysis	• Text mining tools to transform text into numbers
Data Analysis	• Too large data to process • Large number of variables • Causality • Finding latent topics and attaching meaning	• Principal components analysis • Topic modelling, latent Dirichlet allocation, and deep learning • Parallelization, bags of little bootstrap, sequential analysis
Reporting and Visualization	• Difficult to understand complex patterns • Facilitating interpretation, representation with external partners and knowledge users	• Describing data sources • Describing methods and specifications • Visualization and graphic interpretations

Data mining extracts valuable information from multiple sources. Requirements of data mining include the ability to handle diverse types of data; graceful degeneration of data mining algorithms; valuable data mining results; representation of data mining requests; mining information from different sources; mining at different abstraction levels; and protection of privacy and data security [4]. Challenges of data mining lie in: dimensionality, scalability, complex and heterogeneous data, streaming data processing, data quality, data ownership and distribution, privacy preservation [13]. Real-time analytics can analyze and create insights from all available data and resources in real time. Multiple streams of heterogeneous data offer the possibility to extract insights in real time. Data stream mining can analyze and process streaming data in the present. Complex event detection can deal with the discovery and management of patterns over multiple data streams, where patterns are semantically rich and at high levels [14].

Big data mining is the capability of extracting useful information from large datasets or stream data. High-performance computing platforms are often required to support big data mining. The main source of data is cleaned, transformed, catalogued, and finally made available for data mining and online analytical processing [8]. Deep learning has the potential in big data. The advantage of deep learning algorithms is that they can be parallelized to enable the analysis of very big and very complex data such as medical images or videos, text data, or other unstructured information [14].

3. Information Security and Privacy in Big Data

Information security is an important issue in the era of big data. The lack of standards among e-commerce sites, the openness of customers, the sophistication of phishers, and the tenacity of hackers put considerable confidential information at risk [2]. Security and privacy concerns are growing because big data becomes more and more accessible. Large-scale data sharing is becoming routine among scientists, clinicians, businesses, and governmental agencies. However, tools and technologies that are being developed to manage the massive data sets are often not designed without adequate security or privacy measures. Existing approaches to security and privacy are increasingly being breached; therefore, it is needed for continuous reassessment and updating of current approaches to prevent data leakage. Data hackers have become more dangerous in the era of big data because of the availability of large volumes of publicly available data and disparate data sources. While stopping hackers from getting data is necessary, stopping hackers (and authorized users) from removing data from its authorized location is also critical [15]. The most common techniques and mechanisms for data confidentiality are access control systems and encryptions. Approaches needed for access control systems for big data [16] are as follows:

- *Enforcing access control policies in big data stores*: Some big data systems allow users to submit arbitrary jobs using programming languages, which creates big challenges in enforcing fine grained access control efficiently for different users.
- *Enforcing access control policies on heterogeneous multi-media data*: Content-based access control is important for surveillance applications which are critical for security.
- *Automatically designing, evolving, and managing access control policies*: When dealing with dynamic environments, the ability to automatically design and evolve policies is critical to ensure that data is available for use while assuring data confidentiality.
- *Merging large numbers of access control policies*: Big data entails integrating data originating from multiple sources that may be associated with their own access control policies; these policies must be enforced even when the data is integrated with other data.
- *Automatically administering authorizations for big data and for granting permissions*: If fine-grained access control is required, manual administration on large data sets is not feasible. Techniques are needed by which authorization can be automatically granted, possibly based on the user digital identity, data contents, and metadata.
- *Data lifecycle framework*: A comprehensive approach to privacy for big data needs to be based on a systematic data lifecycle approach.

The National Science Foundation (NSF) in the US has made substantial investments both in cyber security and big data. It is therefore critical that the two areas work together to determine the direction for big data security. The collection, storage, manipulation, and retention of massive amounts of data have resulted in serious security and privacy problems. Various regulations are being proposed to handle big data so that the privacy of individuals is not violated. A security challenge for big data management and analytics is securing infrastructures. Many technologies such as Hadoop, MapReduce, Hive, Cassandra, PigLatin, Mahout and Storm do not have adequate security protections [6]. The challenges that Big Data infrastructures, data analytics and SaaS (Software as a service) present to cryptology and privacy can be summarized as follows [17]:

- *Erosion of data confidentiality*: Services are organized to conduct computation on the server side, meaning that all data generated during using the services flows to the service providers to be stored in Big Data infrastructures.
- *Integrity of user devices, servers, and algorithms*: Increasingly, service providers limit the ability of users to view or control what services are doing on devices or servers. This is intensified with accelerated feature updates.
- *Reconfiguration of service provision*: User-facing services are typically made up of multiple services, implying that the number of entities with access to user data increases while transparency decreases.
- *Intensification of identity management and tracking*: SaaS providers require user authentication as a prerequisite to service use and user tracking as central to their business models and licensing schemes, which has culminated in what might be called 'tracking and data analytics as a service'.
- *Ethics of experimenting with (user) populations for service optimization:* The application of machine learning to populations and their environments amplifies concerns about discrimination, unfair treatment, and human experimentation. In case multiple data owners have sensitive data sets such as health databases and they want to analyze and mine these big datasets, the first step is probably to share the datasets. However, the owners may not trust each other. A good option is to use biometrics to authenticate the users.

4. Big Data Analytics, Cloud Computing and Biometrics

Cloud computation supports biometric systems, healthcare, and Big Data analytics due to unbounded computational resources on the cloud. Biometrics is a powerful tool for security which aims to identify individuals based on their behavioral or physiological features. Face, fingerprint, iris, gait, and voice recognition, etc. are biometric technologies. Securer multi-party computation techniques are often used to protect privacy introduced by biometric based security systems job [18]. Every smartphone, tablet or other mobile and Internet-enabled device contains a host of different sensors. Sensors and biometrics have taken over the mass market. Biometric identification technology is a growing area of experimentation in the mobile devices sector. Using a motion sensor or microphone, it's possible to measure how people walk, run, drive, or speak. Algorithms are then used to create voice, gait, or driving profiles which enable smartphones to identify their owners from the way in which they walk, talk or drive [19].

Three trends of biometric applications have been identified as follows: the growth of unsupervised biometric systems for identity verification, accessed via mobile devices; the development of "second-generation" biometric technologies which can authenticate individuals remotely without their knowledge; and the linking of biometric data with other types of 'big data' as a part of efforts to 'profile' individuals [20].

Biometrics can be used to secure cloud storage. Cloud computing can take advantage of the strong authentication capability of biometrics to improve the security of the cloud and develop new service models such as Biometric authentication as a Service (BioAaaS). Biometric technologies can leverage cloud's unbounded computational resources and properties of scalability, flexibility, and cost reduction to reduce the cost of the biometrics system requirements of various computational resources (i.e. data storage or processing power) and to improve the performance of biometrics systems' processes (i.e. biometric matching). Leveraging the cloud for big data biometrics was proposed and a prototype of biometrics system that relies on the cloud computation resources was developed based on Hadoop and Hadoop Distributed File System (HDFS) which support the distribution processing of big data sets. The developers used Hadoop MapReduce for biometric database organization and bulk operations, ZooKeeper as a coordination service, and HBase for real time operations (i.e. verification and identification). Similarly, an iris recognition system was implemented based on Hadoop. The system only uses Hadoop MapReduce and HDFS. The experiment of the iris recognition system showed that the cloud's unbounded computation resources can be leveraged effectively to speed up the biometrics systems processes [21].

Storing biometric data in the cloud makes the system highly scalable and allows quick and reliable adaptation of the technology to an increasing user base, which may also raise privacy concerns and may not be in accordance with national legislation. Cloud based biometric services have a great potential market value. Migrating more biometric modalities to the cloud and, if possible, devising a multi-modal cloud-based biometric solution are interesting research topics [22]. The privacy of biometric data must be preserved while the data is stored in or in transfer to the cloud storage. The security of biometric data stored on the cloud is a main challenge of biometric systems. Traditional encryption methods are not good in preserving the privacy of biometric data that are processed in the cloud. Homomorphic encryption is a promising encryption technology that allows processing over the encrypted data. Also, biometric template protection mechanisms that allow matching over the encrypted biometric data can be used to preserving the privacy of the biometric data processed in the cloud [21].

Very large-scale biometric systems are becoming mainstream in nationwide identity cards and mobile secure payment methods. The biometric systems have challenges that involve the effective managing of the complex life cycle and operations of identity information. The challenges lie in the "four Vs" of big data: the immense enrolment database size (*Volume*), rapid transaction response-time (*Velocity*), potentially noisy and fraudulent (*Veracity*), and multiple biometric identifiers (*Variety*). Biometric systems are possibly to emerge as among the most critical Big Data systems [8].

5. Big Data Analytics in Healthcare

5.1. Big Data and Biometrics in Healthcare

Biometrics such as fingerprint and signature has been used in the healthcare industry for identification, insurance, and management, etc. The healthcare industry and biometrics generate big data that is shown in Table 4. Big data in healthcare or medical service has increased greatly due to social media and networks (such as Twitter and Facebook), sensory or digital technology, mobile devices and smartphone apps, and personal health sensors or meters (generating digital data in real time) [7].

Table 4. Big Data Types in Healthcare [23]

Types	Description or Examples
Biometrics	• Genetics • Employee credentialing
Machine-to-machine	• Patient monitoring
Human generated	• Electronic medical records (EMR) • Call center voice recording
Big transaction data	• Claims records
Web and social media	• Twitter feeds • Clickstream data

Big Data analytical tools hold the promise to study outcomes of large-scale population-based longitudinal studies, develop predictive models, and predict trends for data generated from electronic medical and health records. The use of natural language processing plays an important role in the systematic analysis and indexing of semantic contents. Telemedicine involves connecting doctors and patients beyond the clinic. However, this communication has been expanded to new levels of interaction. This new feature has opened new possibilities of patient-to-patient communication regarding healthcare beyond the traditional doctor-to-patient paradigm. Many patients with chronic diseases such as cancer and diabetes are now using network to share experience with other patients, therefore providing another potential source of big data [24]. Table 5 shows a secure healthcare procedure with biometrics and data analysis.

5.2. Benefits and Applications of Big Data Analytics in Healthcare

The increasing use of Big Data analytics in healthcare offers benefits such as tailored healthcare solutions, better monitoring of health conditions, fewer mistakes, improved quality of care based on proactive early intervention, reduction in costs that are not reimbursed by insurance, and data sourced from multiple applications. Specifically, Big Data analytics helps reduce 30-day readmission rate of patients with congestive heart failure; leverage unstructured data to improve the quality of structured data for smoking status and drug & alcohol abuse; extract additional relevant clinical factors that are not found in structured data. For example, analytics teams used text analytics from discharge summaries, echocardiograms, patient histories, and doctors' notes to gather and analyze data [23,26]. Data enables the healthcare industry to cut costs and improve quality through [27]:

- Right living: Data can help patients to take an active role in their own health such as exercise, diet, and medication adherence.
- Right provider: Proven outcomes for patients based on data help match a provider's skills with the needs of a patient.
- Right care: Data can reduce medical errors and improve outcomes. Big Data tools help facilitate evidence-based care that is personalized to a specific patient.
- Right innovation: Healthcare providers can analyze a patient's history data, real-time data from monitors, clinical factors, and lifestyle choices, etc. to provide a holistic view of the patient and develop the most effective care plan.
- Right value: Cost-effective healthcare can be achieved through various methods such as patient-outcome reimbursement and eliminating fraud and abuse using big data.

Body sensor networks (BSNs), sometimes also called body area networks (BANs) or wireless body area networks, are designed to connect and operate sensors within, on, or at proximity to the human body. BSNs have been used in monitoring vital signs, gait patterns, daily activities, motor fluctuations in Parkinson's patients, and balance and fall, etc. Sensors and radio frequency (RF) technologies for BSNs are becoming mature and various products are commercially available. Both the volume and type of data that can be collected via BSNs have grown greatly and are now beyond the ability of commonly used software tools to process within a tolerable elapsed time. Big Data is an approach to mining the valuable real-time data. Since real-time decision support is in high demand in healthcare and medical service, BSNs are a crucial building block for future healthcare information systems. They can be used to collect health information and deliver health decisions or therapeutic treatments to the users. However, there are challenges in designing BSN nodes with antennas that operate efficiently around or inside the human body and processing the heterogeneous and growing amount of data on-node and in a distributed system for optimized performance and power consumption [28].

*e*Health encompasses a diverse range of technologies, including Big Data, software, digital networking, mobile connectivity, and smart infrastructure, etc. eHealth solutions such as Big Data analytics are being increasingly identified and adopted to enable analysis and improve the 'speed to insight' from research initiatives. Major opportunities to enhance eHealth enabling tools will focus on [29]:

- Improving the availability of data and analytics tools to support research and learning.
- Facilitating the translation of research evidence and information into policy and practicing the relationship between translational research and personalized medicine.
- Undertaking research on the factors that have great return on investment for attracting and retaining healthcare professionals.

Table 5. A Complete Secure Solution in Healthcare from Data Collection, Analysis, to Decision [25]

Incoming Data	Stream, Acquire	Organize /Discover	Analyze, Visualize	Decide
• Biometrics • EMR • Clinical • Pharma- ceutical • Metabolic • Social • Lifestyle /History • Activity /Claims Cost	• Correlate biometrics data across devices • Correlate Omics w/ EMR data • Track treat- ments, notes • Track outcomes • Track social media, payments	• Model protocols • Model normal behavior • Model outcomes	• Populate metrics • Discover unknowns • Highlight outliers • Determine risk • Visualize data • Report	• Diagnose • Predict • Measure • Treat • Prevent

Big Data technologies will open new opportunities and enable breakthroughs in healthcare. Patient-generated data from IoT devices such as weighing scales and blood pressure monitors are providing critical information about the day-to-day lifestyle characteristics of an individual. Mobile Health (mHealth) apps have the potential to personalize interventions. mHealth technologies exploit contextual information which is the key to personal and precision medicine [14].

Precision medicine is a field of medicine that considers individual differences in people's genes, microbiomes, family history, environments, and lifestyles to make diagnostic and therapeutic strategies precisely tailored to individual patients. Personalized medicine means a selection of treatment that is best suited for an individual. Precision medicine focuses on classifying individuals into subpopulations that differ in their susceptibility to a disease and in the biology and/or prognosis of diseases. Precision medicine is an approach to integrating clinical and molecular information to understand the biological basis of diseases. It is more research-oriented whereas personalized medicine is more clinical practice-oriented. Precision medicine and personalized medicine have much overlap that they are often used interchangeably in practice. Omics data, mobile Internet real-time data and electronic health record (EHR) data are the top three areas for Big Data in medical research. Precision medicine will use all the three kinds of big data [30].

IBM has helped healthcare providers in predicting patient health risks using predictive analytics, identifying crises by analyzing real-time data as it streams from monitoring equipment, and improving healthcare outcomes by providing timely and meaningful insights to care providers. There are endless exciting opportunities for Big Data in the future of healthcare: genome sequencing, innovative smart devices, accessing data anywhere, point-of-care decision-making, and reducing readmissions [27].

5.3. Challenges of Big Data Analytics in Healthcare

Big Data challenges in health include speed (velocity in data generation), data heterogeneity, and variety of data, etc. In healthcare, data heterogeneity and variety arise because of linking the diverse range of biomedical data sources. *Volume*, *Velocity*, and *Variety* are the three Vs of the characteristics of big data. *Veracity* is important for big data because personal health records may contain typographical errors, abbreviations, and cryptic notes [24]. As for challenges regarding big data and population health, data protection regulation makes it difficult to analyze data from different healthcare providers and services because many population health records are unstructured text. Also, there is interoperability, data quality, and data integration limitations among existing systems that are not dynamically scalable to manage and maintain big data structures [14].

Table 6. Some Strengths, Weaknesses, Opportunities and Threats of Big Data Technologies for the Health Economics and the General Public

Aspects	For the Field of Health Economics	For the General Public
Strengths	• Open datasets help people from different disciplines analyze data and bring new perspectives for more open collaboration • More robust long-term outcomes data are available for economic models; more cost data in silos are available that represent the whole patient journey. • Availability of more individual-level data help understand the effects of drugs at the individual level, changes in outcomes and value to individuals.	• Interactions arising from uncommon drug combinations. • Better measured individual preferences and more precise treatments. • More tailored and better drugs, greater efficiency. • Behavioral data from sensors and the 'Internet of Things' make communities healthier. • Giving healthcare providers and patients access to relevant, well presented, and high- quality data, helping them to make better decisions.
Weaknesses	• It is costly to store and manipulate big data.	• People do not feel safe if monitored by such as biomonitoring systems.
Opportunities	• Possibility of predicting mechanisms and side effects of new drugs through Big Data. • Potential to further personalize medicines by informing how drugs and other treatments work in individual patients.	• Better predictions of new development in diseases mean quicker responses and better prevention of diseases. • Individuals understand more about their genetic predisposition to diseases.
Threats	• The health economics might suffer a backlash if people feel that their data are misused or they are over-monitored. • Health economics as a discipline may need to develop new skills for Big Data applications.	• Risk of companies using big data to collude at the expense of customers. • Risk of data being lost or stolen. • Ethical risks in excessive genetic screening.

Big data in healthcare also raises challenges in privacy, security, data ownership, and governance. Personal data are often generated and stored at a centralized location. They are often distributed over various servers and networks. Data must be linked to the right person to ensure correct diagnosis and treatment. The collected data about an individual must be uniquely tagged with an identifier. Furthermore, data security should be ensured at all levels of healthcare systems, especially at the sensor level at which the data is collected [24]. There are risks in using big data such as loss of patient confidentiality or misuse of data by insurers or other companies. Table 6 [26] shows some strengths, weaknesses, opportunities and threats of Big Data technologies in healthcare.

6. Conclusion

Big Data analytics can unearth valuable information. The integration of Big Data analytics and cloud computing leads to the service model AaaS High-performance computing can support Big Data analytics. Deep learning has the potential in big data and enables the analysis of very big and very complex data such as mages or videos, text data, or other unstructured data in healthcare. The security and privacy problems of big data are due to more and more accessibility and result from the collection, storage, manipulation, and retention processes of the data. There are challenges of Big Data infrastructures, data analytics, and SaaS in data security and privacy.

Biometrics can be used to secure cloud storage. Cloud computing can take advantage of biometrics to improve the security of the cloud and develop new service models such as (BioAaaS). Cloud's unbounded computational resources help improve the performance of biometrics systems' processes. The privacy and security of biometric data stored on the cloud is a main challenge. Biometric template protection mechanisms can be used to preserving the privacy of the biometric data processed in the cloud. Biometric systems are possibly to emerge as among the most critical Big Data systems.

Biometrics has been used in healthcare for identification, insurance, and management, etc. The healthcare industry and biometrics generate big data. BSNs are one of crucial building blocks for healthcare information systems. eHealth and IoT-based *m*Health are important to personalized medicine and precision medicine that use kinds of big data. There are big opportunities for Big Data analytics in healthcare. However, big data in healthcare also raises challenges including stream data with high velocity, data heterogeneity and variety, privacy, data ownership, governance, and data security, etc.

References

[1] M. O'Leary and S. Rahman, "Driverless cars, drones and DNA: how to build trust in the data age," White Paper, Herbert Smith Freehills, Australia, May 16, 2017.

[2] Gantz and D. Reinsel, "The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east," IDC iView: IDC Analyze the future 2007, no. 2012, PP. 1-16.

[3] K.E. Britton, "IoT Big Data: consumer wearables data privacy and security," ABA Section of Intellectual Property Law. vol. 8, no. 2, 2015, pp. 1-8.

[4] M.E. Milakovich, "Anticipatory Government: Integrating Big Data for Smaller Government." Internet, Politics, Policy 2012: Big Data, Big Challenges, 2012.

[5] A.R Manu, V.K. Agrawal, and K.N.B. Murthy. "A user identification technique to access big data using cloud services," International Journal of Computer Applications, vol. 91, no. 1, 2014.

[6] B. Thuraisingham, "Big data security and privacy," In Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, ACM, 2015, pp. 279-280.

[7] Y. Liang and A. Kelemen, "Big Data Science and its Applications in Healthcare and Medical Research: Challenges and Opportunities," Austin Biometrics and Biostatistics, vol. 3, no. 1, 2016, pp. 1-9.

[8] S. Sreeja, A.K. Sangeetha "No science no humans, no new technologies no changes: "Big Data a Great Revolution"," International Journal of Computer Science and Information Technologies, vol. 6, no. 4, 2015, pp. 3269-3274.

[9] S. Patel and A. Patel, "A big data revolution in health care sector: opportunities, challenges and technological advancements," International Journal of Information Sciences and Techniques (IJIST), vol.6, no.1/2, March 2016.

[10] J.W. Crampton, "Collect it all: National security, big data and governance," GeoJournal, 2014.

[11] C. Munoz, S. Megan, and D. Patil. "Big data: A report on algorithmic systems, opportunity, and civil rights," Executive Office of the President. The White House, USA, 2016.

[12] G. George, E.C. Osinga, D. Lavie, and B.A. Scott. "Big data and data science methods for management research." Academy of Management Journal, VOL. 59, no. 5, 2016, PP. 1493-1507.

[13] H. Sahu, S. Shrma, and S. Gondhalakar. "A brief overview on data mining survey." International Journal of Computer Technology and Electronics Engineering (IJCTEE), vol. 1, no. 3, 2011, pp. 2227-1899.

[14] A. Heinrich, A. Lojo, F. Xu, et al. "Big data technologies in healthcare: Needs, opportunities and challenges," White Paper, Big Data Value Association, TF7 Healthcare Subgroup, 21 December, 2016.

[15] C. Schmitt, "Security and privacy in the era of big data." ARENCI White Paper Series, ARENCI/National Consortium for Data Science, 2013.

[16] E. Bertino, and M. Kantarcioglu. "Big Data–Security with Privacy." The NSF Workshop, University of Texas at Dallas, USA, September 16-17, 2014.

[17] van der Sloot, Bart, D. Broeders, and E. Schrijvers, Eds. Exploring the Boundaries of Big Data, Amsterdam University Press, 2016.

[18] S. Yu, "Big privacy: Challenges and opportunities of privacy study in the age of big data." IEEE access 4, 2016, PP. 2751-2763.

[19] T.F. Dapp, and V. Heine. "Big Data. The untamed force." Dtsch. Bank Res. May 5, 2014.

[20] A. Miller, D. Byles, J. Dowd, et al. "Current and future uses of biometric data and technologies," House of Commons Science and Technology Committee, Sixth Report of Session 2014–15, London, UK, 7 March 2015.

[21] A.A. Albahdal, and E. B. Terrance, "Problems and promises of using the cloud and biometrics." In Information Technology: New Generations (ITNG), 2014 11th International Conference on, IEEE, 2014, pp. 293-300.

[22] P. Peer, J. Bule, J.Z. Gros, and V. Struc. "Building cloud-based biometric services." Informatica, vol. 37, no. 2, 2013, 115.

[23] S. Soares, "Big data governance." Information Asset, LLC, 2013.

[24] J. Andreu-Perez, C.C. Poon, R.D. Merrifield, S.T. Wong, & G.Z. Yang, "Big data for health," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, 2015, pp.1193-1208.

[25] E. Kleiman, "Big Data and Advanced Spatial Analytics," Whrkshop, APAC Technology Product Releases & Management Programs, Oracle, 2012.

[26] B. Collins, "Big Data and health economics: strengths, weaknesses, opportunities and threats." PharmacoEconomics, vol. 34, no. 2, 2016, 101.

[27] A.K. Roy, "Impact of Big Data Analytics on Healthcare and Society," Journal of Biometrics & Biostatistics, vol. 7, no. 3, 2016, pp.1-7.

[28] C.C.Y. Poon, P.L.L.Benny, 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, 2015, pp. 4-16.

[29] J. Skinner, et al., “eHealth Strategy for NSW Health 2016-2026,” White Paper, NSW Health, 23 May, 2016.

[30] X.D. Zhang, “Precision Medicine, Personalized Medicine, Omics and Big Data: Concepts and Relationships,” J Pharmacogenomics Pharmacoproteomics, vol. 6, no. 1, 2015, 1000e144, pp. 1-2.