Research Area: Big Data Analytics in Healthcare

Hope Antony Makoni, Wellington Gwavava

Abstract: The study performs a review to synthesize prior research on the applicability of big data analytics in healthcare. It examines several studies and presents them in a comprehensive framework. This review paper is looking at the use of big data analytics in the healthcare. While the use of big data analytics in still in its infancy stage in Zimbabwe, companies are beginning to realize the importance but it has surprisingly not moved much in the healthcare where opportunities have risen, for example the outbreak of cholera in Zimbabwe could have used big data analytics and now the Corona virus pandemic is an area where big data analytics can be used. This review paper will however look mainly into outside Zimbabwe papers where big data analytics have been in use for some time. It has become a topic of special interest for the past 2 decades or so because of great potential that is hidden in it. In the healthcare industry, various sources for big data include hospital records, medical records of patients, results of medical examinations, financial payments, medication administered and devices that are part of the internet of things. There are various challenges associated with each step of handling big data analysis.

Keywords: Rogue access point, wireless security, service set identifier, beacon frame

1. Introduction

Information has been key to a better organization and new developments. The more information we have, the more optimally we can organise ourselves to deliver the best outcomes. Data collection is very important in every organization and we can use this data for the prediction of current trends of certain parameters and future events. The technological advances have helped us in generating more and more data even to a level where it has become unmanageable with currently available technologies. In this review, we discuss about the basics of big data including its management, analysis and future prospect in healthcare sector.

Big data analytics can be conceptualised as the analysis of detailed, dynamic, low-cost, massive and varied data sets to deliver sophisticated solutions.

Big data
This represents large amounts of data that is unmanageable using traditional software or internet-based platforms. It surpasses the traditionally used amount of storage, processing and analytical power. The characteristics of big data are defined by the 3 Vs which some scholars have extended to 7. These are:

- **Volume (size)**-the data is large in amount
- **Variety (complexity)**-big data includes structured, semi-structured and unstructured data in different formats, such as text, image, audio, sensor data among others.
- **Velocity (speed)**-big data handles high rates of data inflow and processes the data in real-time.
- **Veracity (quality)**-big data accumulates detailed data that is exhaustive in scope.
- **Value (knowledge)**-big data offers in-depth information about a topic of discussion.
- **Variability (flexibility)**-big data provides for the constantly changing nature of data by offering extensibility (the addition of new data fields) and scalability.
- **Valence (connectedness)**-big data connects common fields to co-join different data sets.

Big data in healthcare
Healthcare is a multi-dimensional system established with the sole aim for the prevention, diagnosis and treatment of health-related issues or impairments in human beings. The major components of a healthcare system are the health professionals (physicians or nurses), health facilities (clinics, hospitals for delivering medicines and other diagnosis or treatment technologies) and a financing institution supporting the former two. The health professionals belong to various health sectors like dentistry, medicine, midwifery, nursing, psychology, physiotherapy among others. Healthcare is required at several levels depending on the urgency of situation. Professionals serve it as the first point of consultation (for primary care), acute care requiring skilled professionals (secondary care) advanced medical investigation and treatment (tertiary care) and highly uncommon diagnostic surgical procedures (quaternary care). At all these levels, the health professionals are responsible for different kinds of information such as patient’s medical history (diagnosis and prescriptions related data), medical and clinical data (like data from imaging and laboratory examinations) and other private or personal medical data.

Electronic health records
EHRs have introduced many advantages for handling modern healthcare related data. Below, we describe some of the characteristic advantages of using EHRs. The first advantage of EHRs is that healthcare professionals have an improved access to the entire medical history of a patient. The information includes medical diagnoses, prescriptions, data related to known allergies, demographics, clinical narratives, and the results obtained from various laboratory tests. The recognition and treatment of medical conditions thus is time efficient due to a reduction in the lag time of previous test results. With time we have observed a significant decrease in the redundant and additional examinations, lost orders and ambiguities caused by illegible handwriting, and an improved care coordination between multiple healthcare providers. Overcoming such logistical errors has led to reduction in the number of drug allergies by reducing errors in medication dose and frequency. Healthcare professionals have also found access over web based and electronic platforms to improve their medical practices significantly using automatic reminders and prompts regarding vaccinations, abnormal laboratory results,
cancer screening, and other periodic checkups. There would be a greater continuity of care and timely interventions by facilitating communication among multiple healthcare providers and patients. They can be associated to electronic authorization and immediate insurance approvals due to less paperwork. EHRs enable faster data retrieval and facilitate reporting of key healthcare quality indicators to the organizations, and also improve public health surveillance by immediate reporting of disease outbreaks. EHRs also provide relevant data regarding the quality of care for the beneficiaries of employee health insurance programs and can help control the increasing costs of health insurance benefits. Finally, EHRs can reduce or absolutely eliminate delays and confusion in the billing and claims management area. The EHRs and internet together help provide access to millions of health-related medical information critical for patient life. EHRs have introduced many advantages for handling modern healthcare related data. Below, we describe some of the characteristic advantages of using EHRs. The first advantage of EHRs is that healthcare professionals have an improved access to the entire medical history of a patient. The information includes medical diagnoses, prescriptions, data related to known allergies, demographics, clinical narratives, and the results obtained from various laboratory tests. The recognition and treatment of medical conditions thus is time efficient due to a reduction in the lag time of previous test results. With time we have observed a significant decrease in the redundant and additional examinations, lost orders and ambiguities caused by illegible handwriting, and an improved care coordination between multiple healthcare providers. Overcoming such logistical errors has led to reduction in the number of drug allergies by reducing errors in medication dose and frequency. Healthcare professionals have also found access over web based and electronic platforms to improve their medical practices significantly using automatic reminders and prompts regarding vaccinations, abnormal laboratory results, cancer screening, and other periodic checkups. There would be a greater continuity of care and timely interventions by facilitating communication among multiple healthcare providers and patients. They can be associated to electronic authorization and immediate insurance approvals due to less paperwork. EHRs enable faster data retrieval and facilitate reporting of key healthcare quality indicators to the organizations, and also improve public health surveillance by immediate reporting of disease outbreaks. EHRs also provide relevant data regarding the quality of care for the beneficiaries of employee health insurance programs and can help control the increasing costs of health insurance benefits. Finally, EHRs can reduce or absolutely eliminate delays and confusion in the billing and claims management area. The EHRs and internet together help provide access to millions of health-related medical information critical for patient life.

Opportunities for Big Data analytics in healthcare
The applications of descriptive, predictive and prescriptive analytical techniques when using big data offer opportunities to enhance the quality of various aspects of healthcare. The literature proposed different opportunities offered by big data analytics in healthcare sector. Such as the following:

a) Medical diagnosis – a data driven diagnosis may detect diseases at early stage and reduce complications during the treatment.
b) Community healthcare-authorities may take preventive steps against the predicted risks of chronic disease among a population and contagious disease outbreaks.
c) Hospital monitoring-real-time monitoring of hospitals can help government authorities ensure optimal service quality.
d) Patient quality-customised patient care facilitated by the big data analytics has the potential to provide rapid relief and reduce readmission rates in hospital.

Challenges of Big Data Analytics
The application of big data analytics to healthcare may face various challenges. Common challenges in this area include:

a) Initial investment-the deployment of the requisites to leverage the benefits of big data incurs huge initial costs for organizations providing health care.
b) Quality of data-the lack of trained personnel and resistance to change in organisational routines may affect the quality of big data accumulated by the organisation.
c) Quality of insights-the poor quality of heterogenous biomedical data has the potential drawback of yielding inadequate insights and misleading suggestions.
d) Privacy and security-scholars warn about the privacy and security concerns of patients regarding exposure to unauthorised data access during intersystem exchanges.

2. Methodology
The review paper is comprised of three sequential processes namely planning the review, performing the review and presenting the review.

Planning the review
The review focused on four databases namely Scopus, web of science PsycINFO and PubMed. These databases are reportedly the most important sources for studies related to medical health informatics. Studies meeting the eligibility
criteria (namely, the inclusion and exclusion criteria) were assessed for quality and robustness.

**Performing the review**

do to determine the appropriate keywords, a search was performed on Google Scholar with the phrase 'big data analytics in healthcare.' The most commonly related terms were identified from the first 100 search results (Khanra, Dhir, and Mäntymäki 2020). We identified from the co-occurrence of keywords that the term 'predictive analytics' was frequently used to refer to 'big data analytics,' following an approach adapted by Khanra, Dhir, and Mäntymäki (2020). Raghupathi and Raghupathi (2014) highlighted applications of big data analytics in healthcare, including analysis of patient profiles with predictive modelling to identify suitable treatments, prediction of outcomes of different treatments, and perception of the most clinically and cost-effective treatments for the patient.

Raghupathi and Raghupathi (2014) highlighted applications of big data analytics in healthcare, including analysis of patient profiles with predictive modelling to identify suitable treatments, prediction of outcomes of different treatments, and perception of the most clinically and cost-effective treatments for the patient. Similarly, we identified that the term 'health management' frequently represents different components.

**Presenting the review**
The oldest study included in the sample was published around 2013 hence the topic under discussion is, arguably, fairly new to the literature of ‘healthcare’ in the extant literature. Among the major components of health management are the clinical diagnosis, clinical research, prediction of disease transmission, preventive healthcare, health insurance, and healthcare service delivery (Kamble et al. 2019).

3. **Findings**

Applications of big data analytics in healthcare

**Health awareness**

This theme involves different facets of general awareness of the holistic health and well-being of patients. For instance, prior studies on health awareness discussed health insurance, living environment and sports behaviour among other topics.

**Healthcare ecosystem**

This theme captures the dynamic relationship among stakeholders in the healthcare ecosystem in managing hardware resources, device networks, data warehousing and other facilities for the benefit of big data analytics.

**Hospital management**

The theme involves the practices of hospital management, such as medication assignment, outpatient management and pre-admission testing. Agnihotri et al (2015) suggested that process innovation and efficient scheduling are keys to addressing bottlenecks in healthcare service delivery.

**Specific medical conditions**

This theme includes studies that discuss specific medical conditions such as cancer, diabetes developed a structural degradation modelling framework for sparse data sets and reported that the framework works.

**Technology aspects**

This theme captures the application of technology to meet the responsibilities of healthcare through e-health services. It encompasses related issues, such as disease surveillance and alert systems developed an algorithm based on a reduced variable neighbourhood search to improve the functioning of e-health networks.

Value delivered by big data analytics in healthcare

The aptness of big data analytics to add significant value to healthcare became evident after an analysis of the findings of the studies under review.

**Conceptual evolution**

This theme encapsulates the contributions of big data analytics in introducing new concepts in healthcare, for instance, platform-as-a-service, healthcare-as-a-service and crowd sourced e-health networks. Boudhir, Ben Ahmed, and Soumaya (2017) identified the potential of big data in conjunction with cloud computing to provide the platform-as-a-service model. Jindal et al. (2018) conceptualised the healthcare-as-a-service model using BDA in cloud computing. Shao et al. (2016) developed an algorithm to solve the data-congestion problem in crowd sourced e-Health networks.

**Data governance**

This theme captures the legal and ethical concerns regarding the usage and security of data in healthcare, for example, access rights management (Zaragoza, Kim, and Chung 2017), the security of patient data (Patil and Seshadri 2014), and trade-offs between healthcare efficiency and privacy risk (Li et al.2015). Zaragoza, Kim, and Chung (2017) proposed improving healthcare data storage through encryption and access rights management. Patil and Seshadri (2014) attributed paramount importance to the security and privacy of patient data as BDA transforms healthcare. Li et al. (2015) concluded that the trade-offs between healthcare efficiency and privacy risk need to be balanced.

**Decision support**

This theme acknowledges how BDA has improved decision-making processes in healthcare organisations with evidence-based decisions (Moutselos, Kyriazis, and Maglogiannis 2018) and faster decision making (Navaz et al.2018); it also shows how improved design produced better public health policy at the national level (Christensen et al.2018). Moutselos et al. (2018) suggested that system trust could be preserved by a cloud design where tools are decoupled from data stores and interfaces. Navaz et al. (2018) proposed that efficient analytics could optimise the handling, processing, and analysis of health data from mobile devices. Christensen et al. (2019) identified that the duration of device use influences the physical activity level of patients.
Disease prediction
This theme captures the ways of predicting serious medical conditions in patients by using the efficient application of BDA, for example, in the prediction of disease (Moreira et al. 2018), the identification of disease patterns (De Silva et al. 2015), and disease-based monitoring systems (Bravo et al. 2018). Moreira et al. (2018) identified that an artificial neural network-based approach is an excellent predictor for gestational diabetes mellitus. De Silva et al. (2015) suggested that BDA can effectively identify disease patterns in patients. Bravo et al. (2018) reported that m-Health services could be used to facilitate continuous monitoring of patients.

Strategy formulation
This theme discusses how BDA can aid healthcare organisations to formulate sustainable business strategies, for example, capability development (Austin and Kusumoto 2016), resource allocation (Gowsalya, Krushitha, and Valliyammal 2014), and profit enhancement for healthcare organisations (Wang, Kung, and Byrd 2018). Austin and Kusumoto (2016) confirmed that the application of BDA has the potential to improve healthcare services. Gowsalya et al. (2014) highlighted how data-driven decision making facilitates efficient resource allocation in healthcare. Wang, Kung, and Byrd (2018) suggested that BDA could deliver business benefits to healthcare organisations that follow strategic approaches.

Technology development
This theme involves advancement in technology to discover novel benefits of BDA in healthcare, for example, in embedded intelligent technologies (Gopal et al. 2019), cloud-based infrastructure (Jindal et al. 2018), and parallel execution of accelerated kernels (Koliogeorgi et al. 2017). Gopal et al. (2019) suggested that applications of embedded intelligent technologies will be crucial for healthcare organisations. Comprehensive framework for the use of biog data analytics in healthcare. Insights from the review helped develop a comprehensive framework comprising six important components of big data analytics. The components that exhibit a degree of interconnectedness

Medical records
Medical records act as building blocks of historical data in healthcare, particularly for sourcing patient data. These are commonly obtained from diagnostic reports (Amirian et al. 2017; De Silva et al. 2015), hospital registers (Austin and Kusumoto 2016; Babar et al. 2016), and patients’ history (Ozminkowski et al. 2015; Zaragoza, Kim, and Chung 2017). Medical records are often available in electronic formats, such as electronic medical records containing a patient’s treatment history from past visits to a doctor. The electronic health record may comprise of comprehensive information about a patient, such as the patient’s records from multiple doctors, medical history, and medications, among other information, for a longer-term.

Sensor data
Increased adoption of newer technologies makes real-time data from sensors in electronic devices available to healthcare functions. These data are often accumulated from health devices (Gopal et al. 2019; Ma et al. 2018), Internet of Things (IoT) devices (Bravo et al. 2018; George, Chacko, and Kurien 2019), and smartphone applications (Navaz et al. 2018; Wu et al. 2017). In general, body temperature sensors, blood oxygen sensors, and electrocardiogram sensors, among other sensors are attached to a patient’s body to monitor health parameters continuously.

Ethical aspects
Big data in healthcare must be accumulated with permission from appropriate stakeholders. Among the issues raised by the use of BDA in healthcare are data privacy (Chen et al. 2017; Hadi et al. 2019), security concerns (Koliogeorgi et al. 2017; Patil and Seshadri 2014), and surveillance (Cheng, Kuo, and Zhou 2018; Sabharwal, Gupta, and Thirunavukkarasu 2016). Therefore, it is critical for authorities accumulating big data to ensure that legal and ethical guidelines preserve data integrity. For instance, collection, storage, and sharing of personally identifiable information in medical records need to comply with the Health Insurance Portability and Accountability Act in the United States.

Technology integration
Integrated technologies in healthcare play a holistic role in the accumulation of big data and delivering the benefits of BDA to appropriate beneficiaries. The literature on BDA in healthcare attributes great importance to the use of peripheral support technologies, such as big data platforms (Gowsalya, Krushitha, and Valliyammal 2014; Manogaran et al. 2018), cloud storage (Chehade and Liu 2019; Kuo et al. 2015), and smartphone-based interfaces (Li et al. 2015; Wu et al. 2016). Therefore, healthcare administrators often view the integration of state-of-the-art technologies as a critical part of organisational value chains.

Hospital management
Applications of BDA have the potential to derive reliable insights for specific beneficiaries in healthcare, including hospital administrators, doctors, and nurses. BDA can serve to help hospital administrators in resource allocation (Agnihotri et al., 2015; Jindal et al. 2018), doctors in patient profiling (Narayan and Greco 2016; Lin et al. 2017), and nurses in providing disease-specific patient facilitation (Boudhir, Ben Ahmed, and Soumaya 2017; Moreira et al. 2018). For instance, hospital management may dynamically allocate resources for treating the Covid-19 patients by using BDA-based insights from data on confirmed cases, population density, demographics, and migration flow.

Customised Care
Patients are often key beneficiaries of insights derived from BDA. BDA can aid in providing personalised care to patients by controlling medications (Christensen et al. 2018; Wang, Kung, and Byrd 2018), predicting diseases (Tseeng et al. 2017; Wang, Kung, and Byrd 2018), and supervising patients (Jin et al., 2016; Shao et al. 2016). The rapid spread of Covid-19 has threatened to overwhelm health systems across the world, forcing hospitals to defer scheduled surgeries and treatments for an unknown period. A BDA-enabled smartphone application may conduct a personalised risk assessment of patients awaiting surgeries, provide suggestions to address minor health complications, and
prioritise patients based on the urgency in the requirement of medical attention from doctors.

The framework posits that the primary sources of big data in the context of healthcare are medical records and sensor data. Big data are accumulated from healthcare sources through the use of integrated technologies. BDA extracts pertinent insights from the accumulated data. The accumulation and utilisation of healthcare data to derive such insights are subject to certain ethical considerations. The derived insights are useful to hospital management and are used in delivering customised care to patients. Integrated technologies in healthcare play critical roles in converting insights from BDA into actions

4. Conclusion and Recommendations

The current study intended to address applicability of big data analytics in the healthcare. It seems that it can play an important role in improving the healthcare ecosystem. The prior literature on the application of big data analytics in healthcare focused on healthcare ecosystem, hospital management practices and specific medical conditions healthcare service delivery through technology use.

Some limitations include studies not available in the four mentioned databases, some journal articles or conferences not available in English. Future research is invited to overcome these limitations. There has also been argument on the confidentiality and privacy of patients and their records, but access can only be allowed after some kind of authorisation.

It is unfortunate that there is not much we have done as a country (or developing countries some would say) to embrace the technological advancement of the western world. This could have helped us a lot especially in this pandemic in terms of contact tracing for given areas. Records of those vaccinated and not vaccinated could be easily availed across the country from one platform. Subsequently those carrying out statistics for people dying with Covid while vaccinated or not can be easily interpreted.

Big Data Analytics in healthcare: promise and potential

5. Advantages of big data in healthcare

By digitizing, combining and effectively using big data, healthcare organizations ranging from single-physician offices and multi-provider groups to large hospital networks and accountable care organizations stand to realise significant benefits. Potential benefits include:

1) Detecting diseases at early stages when they can be treated more easily and effectively.
2) Managing specific individual and population health.
3) Detecting health care fraud more quickly and efficiently.
4) Numerous questions can be addressed with big data analytics. Certain developments or outcomes may be predicted and / or estimated based on vast amounts of historical data such as:
   a) Length of stay of patients in a hospital,
   b) Patients who will choose elective surgery,
   c) Patients who will likely not benefit from surgery,
   d) Patients at risk for medical complications,
   e) Causal factors of illness,
   f) Possible co-morbid conditions

Mckinsey estimates that the big data analytics can enable more than $300 billion in savings per year in U. S. Mackinsey believes big data help reduce waste and inefficiency in the following three areas:

5) Clinical operations-comparative effectiveness research to determine more clinically relevant and cost-effective ways to diagnose and treat patients.
6) Research and development-predictive modelling to lower attrition and produce a leaner, fast, more targeted R&D pipeline in drugs. Statistical tools and algorithms to improve clinical trial design and patient recruitment and speeding new treatments to market. Analysing clinical trials and patients’ records to identify follow-on indications and discover adverse effects before products reach the market.
7) Public health
   a) Analysing disease patterns and tracking disease outbreaks and transmission to improve public health surveillance and speed response.
   b) Faster development of more accurately targeted vaccines, for example, choosing the annual influenza strains.
   c) Turning large amounts of data into actionable information that can be used to identify needs, provide services and predict and prevent crises for the benefit of the population.

References

in the era of big data. ” Journal of Clinical Medicine 8 (3): 316. doi: 10.3390/jcm8030316. [Crossref], [Web of Science ®], [Google Scholar]


Advanced Scholar

Park Harerimana

Elmirghani

Gu

Chennai

International

Gowsalya

Medicine

on

In

“George

George

ijinfomgt.2019.05.003. [Crossref], [Web of Science ®], [Google Scholar]


networks. " IEEE Access 4: 9174–9183. doi: 10.1109/ACCESS.2016.2646058. [Crossref], [Web of Science ©], [Google Scholar]


