


# A framework for selecting analytics tools to improve healthcare big data usefulness in developing countries

**Author:**Tiko Iyamu<sup>1</sup> **Affiliation:**

<sup>1</sup>Department of Information Technology, Faculty of Informatics and Design, Cape Peninsula University of Technology, Cape Town, South Africa

**Corresponding author:**

Tiko Iyamu,  
Connectvilla@yahoo.com

**Dates:**

Received: 13 June 2019

Accepted: 13 Dec. 2019

Published: 26 Mar. 2020

**How to cite this article:**

Iyamu, T., 2020, 'A framework for selecting analytics tools to improve healthcare big data usefulness in developing countries', *South African Journal of Information Management* 22(1), a1117. <https://doi.org/10.4102/sajim.v22i1.1117>

**Copyright:**

© 2020. The Authors.  
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

**Background:** In many developing countries including South Africa, there are challenges in understanding how the different networks of patients, diagnoses, and medical personnel are formed, as well as the types of big data that are generated. The challenges include the relationship and interaction that exist between the big data within the various networks. Some of the challenges manifest into different factors such as inaccuracy of data, inconsistency, incompleteness of data, and lack of cohesion. The trajectory of the challenge is the inability to select the most appropriate analytics tools for big data analysis.

**Objectives:** The objective of the study was to propose a solution that can be used to address the challenges in selecting analytics tools, to enhance big data usefulness for the improvement of healthcare services, particularly in developing countries.

**Method:** Literature of within 10 years of publication in the areas of big data, big data analytics and healthcare were gathered and used as data. The analysis of the data followed the hermeneutics approach within the interpretivist paradigm.

**Results:** From the analysis, factors that influence big data and analytics tools' usefulness were found, based on which a solution (framework) is proposed. The solution is intended to contribute to the works of information systems and technologies (IS/IT) personnel, health practitioners and academics.

**Conclusion:** The relationships that exist between actors, and the actors' interactions in the process of providing medical services contribute to the sources of big data. Therefore, it is necessary to holistically analyse healthcare big data from both technical and non-technical perspectives.

**Keywords:** big data; big data analytics; big data usefulness; healthcare; selection framework; South Africa.

## Introduction

Information and communication technology (ICT) is an enabler and a platform for technological activities, such as the transmission of data, which makes it necessary to briefly discuss it prior to focussing on big data (Yu, Lin & Liao 2017). Also, ICT offers tools and technologies that ultimately improve the quality of healthcare services (Busagala & Kawono 2016). However, it is important to highlight that it also has some challenges in spite of its benefits. The use of ICT and big data for healthcare improvement has encountered challenges regarding integration and the unavailability of sufficient infrastructure to maximise the benefits of big data (Abouzahra 2011; Mishra, Kalra & Choudary 2013).

Big data refer to data sets, which are obtained from different related or unrelated resources, and are characterised by volume, velocity and variety, also known as three Vs (Gandomi & Haider 2015; Schüll & Maslan 2018). Belle et al. (2015) deemed healthcare to be the prime example of how the three Vs are an essential aspect of the data it produces. Big data are known to create value, stated Watson (2014). However, that can only happen once it is analysed using data analytics tools. According to Elgendy and Elragal (2014), a simpler description of big data analytics is that it is the application of analytics techniques on big data.

In addition to the three Vs, the science of big data focusses on heterogeneity, which includes levels of granularity, media formats and complexity (Nanni et al. 2015). In an attempt to explore the value and usefulness of big data, heterogeneity poses a challenge in its analysis (Jagadish et al. 2014). Heterogeneity of the types of devices used and the nature of data generated are risks associated with big data (Marjani et al. 2017). Labrinidis and Jagadish (2012) argued that

**Read online:**

Scan this QR code with your smart phone or mobile device to read online.

heterogeneity hinders progress in the creation of value from data. This makes heterogeneity an important aspect of big data, especially in the area of integration (Micheni 2015).

The first specific objective of the study was to determine the factors that influence the use of big data analytics to improve healthcare service delivery in the South African environment. To meet this objective, it was necessary to understand the factors that influence data analytics, from a human perspective and a big data perspective. From the data analysis using the hermeneutics approach, factors of influence were revealed, as depicted in Figure 1.

This article is divided into eight main sections. It starts with introduction; then, the objective of the study is problematised. This is followed by a review of literature where big data, big data analytics and healthcare are combined. In the fourth section, the research methodology is discussed, and the next section is a discussion of the phenomenon being studied. The proposed solution is presented through a framework in the sixth section. Recommendation of the framework is presented in the seventh section. Lastly, a conclusion is drawn.

## Problematising the healthcare big data in a developing country

The main motivation of this study is in twofold: (1) majority of previous studies focus on the importance, challenges and opportunities of big data analytics (Shahbaz et al. 2019) and (2) very few studies focus on healthcare big data in developing countries (Luna et al. 2014; Malaka & Brown 2015). Purkayastha and Braa (2013) explained how in developing countries reliable diagnosis is increasingly a challenge and often medical practitioners reorder tests, which can be attributed to lagging analytics of healthcare big data.

In South Africa and many other developing countries, it is a challenge to bring patient's big data together within a facility or from different health programmes (Luna et al. 2014; Malaka & Brown 2015; Purkayastha & Braa 2013). This challenge is caused by lack of integration and analysis of a variety of healthcare big data to address impending problems (Kankanhalli et al. 2016). Thus, scalability is a fundamental challenge for big data analytics, as well as ontological extraction and semantic inference, to support innovative processes in the care for patients. The analytics tools also pose challenges to both scientists and IS/IT specialists in different ways (Nativi et al. 2015). From the academic front, big data analytics is found to be a disruptive innovation, which is reconfiguring how research is conducted, and has epistemological implications on data revolution (Kitchin 2014). From both empirical and experimental perspectives, it was revealed that big data present technical challenges to analytics tools, because of its volume, variety and velocity (Priyanka & Kulennavar 2014).

The relationships formed during the process of providing medical care contribute to big data. Therefore, it is important

to employ the most appropriate analytics tools for examining the relationship that exists between humans, that is, between medical personnel and patients on the one hand, and between humans and non-humans (data and medical apparatus in providing and receiving healthcare services) on the other hand. Exploring these relationships brings out the issues that can contribute to big data analytics. Gaining clarity on these issues helps in proposing a solution that would be suitable for healthcare. Most importantly, it helps to develop a solution that considers healthcare needs in the context of South Africa (Malaka & Brown 2015). In that, one of the critical technical challenges of big data analytics is the lack of capability to handle large-scale transactions from non-standard medical terminology in patient records (Purkayastha & Braa 2013).

## Literature review

Big data are collected from various sources, such as healthcare (Song & Ryu 2015), and national geographic conditions monitoring data and earth observation data (Li, Yao & Shao 2014). Big data are increasingly useful to scientists, health practitioners and the society in general (Shu 2016). To improve big data usefulness for healthcare services, it requires analytics (Abarda, Bentaleb & Mharzi 2017). Big data analytics refer to a collection of analytic techniques and technologies that have been specifically designed to analyse big data to inform decision-making (Kwon et al. 2014). Shahbaz et al. (2019) explain that healthcare organisations are lagging in the sophisticated use of big data analytics, in spite of the fact that the sector generates huge volume at high pace. Big data analytics engineer transformation data sets, from raw to refinement stages (Shu 2016). Innovation of big data from the perspective of developing countries can be understood as all the scientific, technological and healthcare activities (Micheni 2015).

Big data analytics are often considered as the process of examining large amounts of data from different sources and in different variations in order to gain an insight that can enable decision-making in real or near time (Sun & Reddy 2013). According to Kwon et al. (2014), big data analytics are the technologies and techniques that can be employed to analyse large-scale and complex data to improve a firm's performance. However, the employment of data analytics cannot be limited to just business; other sectors have to be considered as well. Big data analytics can be further described as a means of helping discover valuable decisions through understanding of data patterns and their relationships using machine-learning algorithms (Archenaa & Mary Anita 2015).

Big data analytics enable the capturing of insights from the data gathered from research, clinical care settings and operational settings to build evidence for improved care delivery as stated by Nambiar et al. (2013). Bottles, Begoli and Worley (2014) stated that studies have proven that the

analysis of big data can help uncover patterns and relations in healthcare, which are often new to health specialists. Earlier studies, such as that of Raghupathi and Raghupathi (2014), suggested that digitising big data through the act of integrating sources within a hospital network can help with accountability within an organisation and ultimately realise its benefits. Eswari, Sampath and Levanya (2015) stated that the analysis of big data not only helps in discovering patterns but also helps in predicting outcomes.

Data analytics enable the systematic review of existing medical information and inform sound decision-making and ultimately improve the efficiency of service of health professionals and facilities (Kavitha, Kannan & Kotteswaran 2016). From the perspective of the patient, data analytics can assist in providing patients with more accurate information that can help in decision-making through the analysis of their data (Sarkar 2017). The patient also benefits from analytics, based on care that is supported by a timelier diagnosis, as well as a more appropriate medication (Ganjir, Sakar & Kumar 2016).

Non-technical factors, such as resistance to change from employees, which manifest from lack of training and understanding, are a key challenge affecting big data analytics, especially in developing countries (Shahbaz et al. 2019). From a technical viewpoint, the integration of multiple sources of data sets brings about the challenge of increased volumes, increased velocity and increased variety of data (Purkayastha & Braa 2013). The challenges in big data analytics limit the full potential of healthcare big data analytics as the only way to yield its value is through thorough analysis (Sarkar 2017). The slow progress in the development of technology, which supports big data, especially in developing countries, is seen to be shocking as earlier predictions stated that the application of big data would be inevitable (Lee & Yoon 2017).

## Research methodology

A qualitative method and an interpretivist approach were employed in the study. The approach starts within the

premise that people's knowledge of reality and human actions are socially constructed (Walsham 2015). The interpretivist approach was preferred as the most suitable and used in the study because of three main reasons: (1) it shares a belief that the world is socially constructed, and these constructions are possible only because of the human ability to associate meanings with objects, events and interactions (Prasad 2017); (2) within the interpretivist approach, there is no objective reality that can be discovered by researchers and replicated by others, in contrast to the assumptions of positivist science (Walsham 2015); and (3) in interpretivism, reality is individually constructed, and there are as many realities as individuals (Scotland 2012). The belief from these reasons helps to gain an understanding of the factors that influence big data analytics for healthcare services in developing countries.

From an interpretivist angle, existing materials (peer-reviewed literature), which of course comprise findings from empirical studies, various views and opinions, were gathered and examined. Two main criteria were used in the collection of the materials: (1) the key areas of focus were big data in healthcare and healthcare challenges in developing countries, which formed the first part of the criteria; and (2) materials published within a 10-year period: 2009–2019 was used as the decade within which data were collected, the second part of the criteria. According to Iyamu, Nehemia-Maletzky and Shaanika (2016:), the:

[S]pread of historical perspectives, in terms of the consistency of the meaning that has been associated to the concepts, as well as the challenges and confusions that are caused overtime. (p. 171)

As presented in Tables 1–3, a total of 49 materials were gathered from academic databases, which include AIS, EBSCO, Google Scholar, IEEE and ProQuest. The keywords used in the search were big data, big data analytics, big data and health, and big data and healthcare in developing countries. As tabulated in tables, scope benefits, challenges and gaps were identified from the materials.

Within the context of developing countries, the use of big data in providing healthcare services is influenced by its

**TABLE 1:** Big data and healthcare.

Object of focus	Description	References
Scope and benefits	Based on its potential benefits, the focus has been on big data in health informatics; new epistemologies and paradigm shifts.	Kitchin (2014); Kumar and Singh (2017); Li, Yao and Shao (2014).
	Big data are employed for a secure healthcare system, which include conceptual design, and big data as an e-health service.	Liu and Park (2014); Sarkar (2017).
	Increasingly, big data are employed for healthcare services, such as a divided latent class analysis for big data.	Abarda, Bentaleb and Mharzi (2017); Ganjir, Sarkar and Kumar (2016); Sun and Reddy (2013).
	Research directions on the adoption, usage and impact of the Internet of Things (IoT) through the use of big data analytics. Predictive methodology for diabetic data analysis in big data.	Eswari, Sampath and Lavanya (2015); Riggins and Wamba (2015).
Challenges and gaps	There are challenges in the use of big data for healthcare services, which include integration, scalability and complexity of heterogeneity.	Jagadish et al. (2014); Nativi et al. (2015).
	The big data analytics are a disruptive innovation that has reshaped research focusses and challenge semantic inference to support innovative activities and processes in providing care for patients.	Hilbert (2016); Labrinidis and Jagadish (2012).
	Other challenges in the use of big data in biomedicine and health come from data sources, infrastructure and analytics tools.	Lee and Yoon (2017); Peek, Holmes and Sun (2014).

Note: For more information, see the full reference list of the article.

**TABLE 2:** Big data analytics and healthcare.

Object of focus	Description	References
Scope and benefits	The concept of big data analytics is gaining presence in both academic (health informatics) and business (healthcare) domains, including government health environments.	Archenea and Mary Anita (2015); Nanni et al. (2015); Elgendy and Elragal (2014); Priyanka and Kulennavar (2014).
	Beyond the hype, promise and potential, the concept of big data analytics has been used to focus on the analysis of risks and integration of data sets within the healthcare environment.	Gandomi and Haider (2015); Kwon, Lee and Shin (2014); Raghupathi and Raghupathi (2014).
	The most common analytics tools, which include diagnostic, descriptive, predictive and prescriptive, are being adopted for healthcare services. Tutorial: big data analytics: concepts, technologies and applications.	Schüll and Maslan (2018); Shu (2016); Watson (2014).
Challenges and gaps	Integration of systems to access healthcare big data. This is attributable to lack of architecture that is specific to healthcare big data.	Bare Bhakti and Kini (2017); Marjani et al. (2017).
	Lack of understanding of the pros and cons of big data analytics in its adoption and use for healthcare services. This challenge is influenced by the uniqueness of health-related tasks, such as cardiovascular care.	Bottles, Begoli and Worley (2014); Rumsfeld, Joynt and Maddox (2016).
	Another set of challenges in the use of big data analytics in healthcare includes the moderating role of resistance to change.	Kankanhalli et al. (2016); Nambiar et al. (2013); Shahbaz et al. (2019).

Note: For more information, see the full reference list of the article.

**TABLE 3:** Healthcare in developing countries.

Object of focus	Description	References
Scope and benefits	Bangladesh and Saudi Arabia are different from African countries such as Bangladesh, Saudi Arabia, South Africa, South Korea and Tanzania.	Malaka and Brown (2015); Moktadir et al. (2019); Song and Ryu (2015).
	In many developing countries, the use of big data analytics has been explored from various angles, such as cloud-based solution, innovation and diffusion.	Belle et al. (2015); Micheni (2015); Mishra, Kalra and Choudhary (2013).
Challenges and gaps	Adoption of big data analytics tools in providing healthcare services in developing countries has been challenging, particularly in the areas of integration and cloud-based innovation.	Alaboudi et al. (2016); Busagala and Kawono (2013); Kavitha, Kannan and Kotteswaran (2016).
	There are very few studies that focus on big data analytics for healthcare services from the perspective of developing countries. This is a gap that makes many facilities and developing countries sceptical in their attempts to adopt the concept.	Abouzahra (2011); Luna et al. (2014); Malaka and Brown (2015); Purkayastha and Braa (2013).

Note: For more information, see the full reference list of the article.

premise, on the one hand (Luna et al. 2014). On the other hand, the use of the analytics tools manifests in some of the challenges that are experienced in providing care to patients (Li et al., 2014). Some of the existing literature in these contexts, big data and big data analytics within healthcare, from the perspective of developing countries, are briefly described in Tables 1–3.

The practice of healthcare heavily relies on patient's data sets in facilitating the services that practitioners provide (Mathew & Pillai 2015). This triggered the phenomenon being studied in the areas of big data and healthcare as briefly described in Table 1.

Table 2 provides a brief description of the disparities as well as challenges and gaps in the adoption and use of big data analytics tools for healthcare services (Purkayastha & Braa 2013). This includes the integration of healthcare big data and systems.

From the perspective of developing countries, a brief description of the benefits and challenges of big data analytics in healthcare that are described in literature are provided in Table 3. Apparently, the use of big data analytics for healthcare services in developing countries has become synonymous with numerous problems, challenges, obstacles and pitfalls, which has prompted studies, such as those that are presented in Table 3.

In the analysis of the data, the hermeneutics approach was followed, because it aims to unveil the concealed messages in

the text through subjective reasoning (Kalaga 2015). The approach was employed by reading through each of the data sets (literature), forward and backward, and in circle. This helps to gain better understanding and to reference particular culture and historical time (Schmidt 2016). The approach was thus useful in identifying the factors that influence big data analytics for healthcare services, which focusses on developing countries. Another important factor is that, from a hermeneutics perspective, human meanings are often not expressed directly but are embedded in artefacts by their creators, and it can be known through interpreting these artefacts (Yanow & Schwartz-Shea 2015). To gain an understanding of the meanings in the data, a close relationship with the text was required, hence the forward and backward approach. In summary, the hermeneutics explores how people read, understand and handle texts within contexts (Thiselton 2009).

## Discussion of analytics tools

Big data analytics are used as a solution for healthcare systems in many countries including developing countries (Song & Ryu 2015). The four most common types of data analytics tools are predictive, prescriptive, descriptive and diagnostic (Shao et al. 2014). In the context of healthcare, Raghupathi and Raghupathi (2014) argued that predictive analytics are used to anticipate risk through analysis of historical health data and patterns. According to Rumsfeld et al. (2016), prescriptive analytics are used to support medical decisions on individual cases by assessing the risk and benefits of the available solutions. The descriptive analytics



provide a summary of the past and the present data, which can be used to inform healthcare decisions (Mathew & Pillai 2015), while the diagnostic analytics focus on finding solutions or answers as to why certain occurrences happen the way they do (Shao et al. 2014).

In the context of healthcare, big data analytics can be used to solve the complexities that reside within information systems that are used to host and manage patients' data sets (Bare Bhakti & Kini 2017; Hemon & Williams 2014). Mancini (2014) explained how the use of analytics has the potential for enhancing the provision of quality treatment, better surveying of public health, as well as improving responses to, and mitigation against diseases that may affect patients. In spite of these identified challenges, the use of analytics for patients' data is not a one-way affair, and it has its own challenges as well. In healthcare, lack of integration is listed as a challenge brought on by different types and sources of data sets (Abouzahra 2011; Lee & Yoon 2017). However, there are no specific challenges that are of a standard type. In most developing countries, there are often different types of challenges, and hence, unique solutions are relatively required (Alaboudi et al. 2016).

Other challenges of big data analytics include creating efficient and strong analytics methods that are essential for healthcare services (Peek et al. 2014). According to Kumar and Singh (2017), the challenges start from the choice of big data analytics platform and the functionalities in terms of criteria, such as scalability. The integration of big data analytics with current healthcare processes and practices is another challenge, which has been highlighted by Lee and Yoon (2017), in that it is not easy to get them to co-exist and function appropriately within health facilities. The traditional systems no longer suffice for big data as stated by Bare Bhakti and Kini (2017), and this has resulted in issues such as the inability to conduct decision-making in real time, which ultimately challenges predictive analytics. The challenges in big data analytics limit the potential of healthcare big data in providing services. This is because the analytics tools seem to be the only way to maximise value and usefulness from patient's big data through its use for analysis (Sarkar 2017).

The outcomes from the analytics tools lie in their applications (Priyanka & Kulennavar 2014). This makes selection and use of the analytics critical, if it is to help in addressing functions such as clinical decision support, personalisation of healthcare activities, public health, operationalisation of processes and policies implementation. The criticality of these functions makes it even more crucial to be more detailed in assessing the existing systems, because majority of them focus on the same or similar solutions, which include to store, find, analyse, visualise and secure data sets. Some of the most common analytics tools are MapReduce, Hadoop, STORM, Tableau, Apache Hadoop, Apache Hive, Memcached, Cloudera, Hue and Splunk (Chang et al. 2016; Liu & Park 2014). Although the existing

solutions seem to hold promise, healthcare big data still encounter challenges (Rumsfeld et al. 2016).

Heterogeneity extends to network, within which big data exist. According to Law (1992), networks are materially heterogeneous, and agents, texts and devices that are subsequently generated form part of the network. It became crucial in examining relationships in which actors participate and influence the shape of the heterogeneous networks (Dwiartama & Rosin 2014). Heterogeneous entities such as people and data contribute to forming networks (Horowitz 2011). Materials join together to generate data and reproduce themselves (Law 1992). Examples of reproduction of big data include: digital closed-circuit television (CCTV), recording of retail purchases and healthcare historical records (Micheni 2015)

In the course of health activities, the networks become heterogeneous, which also increases the levels of security, making it more difficult for analytics tools to produce useful and purposeful data sets from the analysis (Archenea & Mary Anita 2015). In addition, the heterogeneity of data imposes new requirements from source viewpoint (Marjani et al. 2017), which can also be challenging as the practitioners attempt to trace the origins.

## Big data analytics framework

Following the hermeneutics approach, integration, structure, skill, availability of data, requirements, data sets, appropriate apparatus, external organisations, integrity and translation were identified as the main entities (actors) that influence the usefulness of healthcare big data for service delivery. The identification of the actors (factors) results from two primary qualifications: (1) the frequency of each factor, which was

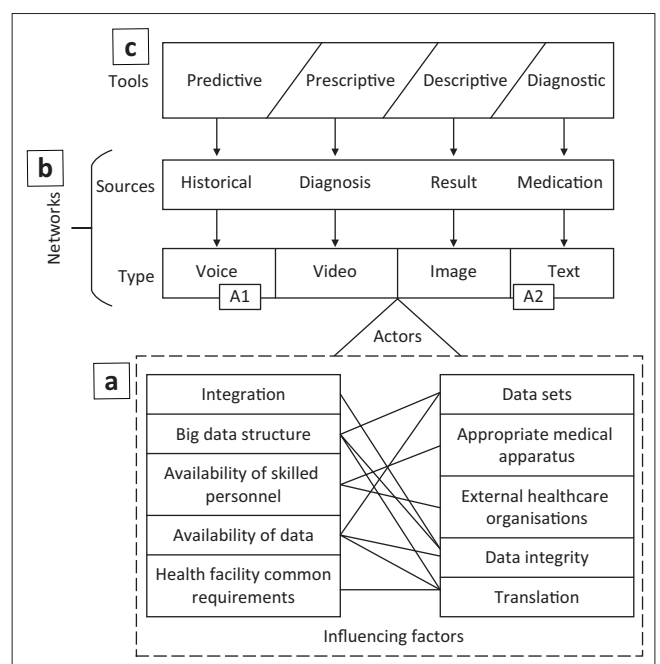


FIGURE 1: Big data analytics framework.

based on the number of articles that the factor has appeared in at the time of this study; and (2) which of the factors co-occur. As shown in Figure 1, the actors are interrelated and interconnected. The actors are influenced by the types and sources of big data, classified as networks. Based on the networks, analytics tools can be appropriately selected, aimed at enhancing the usefulness of healthcare big data. The actors, networks and tools are grouped into categories (levels) A, B and C, respectively, which together form the framework as shown in Figure 1.

The framework is proposed as a solution, which can be used in addressing the challenges of big data, towards improving healthcare service delivery in a health facility. The framework is a bottom-up approach. This means that from the actors (A1 and A2), data are generated and grouped into categories of networks (B). Based on the networks (groupings), analytics tools (C) are selected and applied.

The first level, level 'A', consists of the main factors that influence the usefulness of healthcare big data. The factors (actors) are divided into two parts: A1 (technical) and A2 (non-technical). This level is intended to assist healthcare practitioners to gain knowledge and an understanding of the importance of: (1) factors of influence and (2) how the factors are interrelated or interconnected. Level 'B' helps to identify, examine and understand the networks, which consist of historical records, diagnoses, results and medications. This level enables identification of the factors that influence the selection of analytics tools as well as the analysis of patients' big data. The last level, level 'C', comprises analytics tools, from which selection can be made for healthcare big data analysis. The discussion that follows should be read within the framework (Figure 1), to gain a better understanding about the framework and how it can possibly be applied.

### **Influencing factors (level 'A')**

The factors that influence big data analytics for healthcare services are revealed in level 'A' of Figure 1. The factors are classified as actors. In this study, actors are anything that has the ability to make a difference (Callon 1986). The factors were grouped into two categories, A1 and A2, which are data science/IT (technical) and health facility (non-technical), respectively. The influencing factors are both human and non-human actors because each of them has the ability to make a difference. The factors in the 'A1' were mapped against those in 'A2'. This was purposely done to establish the relationship between factors, and begin to understand how they directly or indirectly influence the healthcare service delivery. By indication, this means that the factors influence both healthcare practitioners and the IT unit, which draws their relationship and necessitates alignment between the units. On the one hand, the health practitioners solicit for support from the IT unit to enable their activities. On the other hand, the unit requires information through interaction in order to provide support and enable the processes and activities concerning healthcare.

The mapping helps in that many medical practitioners are often limited in terms of the insight they can gain from the classification of patient's big data, which guides their analysis of the data sets towards providing services. Currently, as revealed from this study, many medical practitioners continue to employ traditional methods of analysis, which does not necessarily categorise the patient's big data, and require manual application. This is in spite of the fact that the traditional methods are time-consuming, are less effective and produce less accurate results, which significantly contribute to fatality in some South African healthcare facilities.

The existence of these factors holds both negative and positive connotations and effects. This is primarily because each of the factors has the capability to make a difference of either enabling or constraining the use of big data in providing healthcare services. However, through acknowledging these influencing factors, they could drive health facilities towards proper selection and use of big data analytics tools. As presented in Figure 1, the influencing factors can potentially support and enable the health facilities towards many benefits, through the following three main ways: (1) put the types of data into perspective, (2) shape their sources of data and (3) select the most suitable data analytics tool.

With due consideration to these influencing factors, health facilities would be able to identify the types of data they accumulate daily and help put them into perspective. This way, health facilities would be able to identify the structured and unstructured types of data by grouping it into big data. Moreover, instead of disregarding unstructured data sets in their analysis, unstructured data sets would also have a role in patient's treatment and ultimately contribute to improving the standard of health provision. By putting their data types into perspective, health facilities would also be able to shape the data sources, through which knowledge is gained. The knowledge enables traceability of the sources or origins of patients' data sets. This helps in having standard source of reference, when conducting health activities, which lead to analysis as well as providing care to patient's health conditions. Knowledge about data types and sources would be useful to advise health facilities on the suitability of analytics tools choice. Thus, decision can be substantiated in that data types and their sources are known and traceable. This assists in formulating requirements and provides clarity in the selection and use of big data analytics tools for healthcare big data, in order to improve service delivery.

### **Networks (level 'B')**

As shown in the framework (Figure 1), there are two main categories of networks, namely, source of data and type of data. These networks are discussed next.

#### **Source of data**

The sources of big data within healthcare facilities are a part of the different networks that exist in the environment (Iyamu

& Mgudlwa 2018). From the perspective of healthcare facilities, each source of patient's big data consists of actors with aligned interests. The sources of big data as proposed in the framework include historical data, diagnoses, results and medication. Each of these sources of big data has groups of interested actors for the purposes of bettering patients' care. For instance, historical data are used to inform decisions made on patient's health condition. This means that diagnoses, and results from tests and medications are based on the patient's medical history. Through analysis of each patient's medical history, health practitioners are able to gain an insight on individual cases, which enables better decision-making.

### Types of data

Similar to other developing countries, there are also different types of data increasingly accumulated within the South African healthcare facilities. A group of data of the same type forms a network (Mgudlwa & Iyamu 2018). The existing networks are voice, video, image and text data. Other actors of each of the networks include the contributors (patients) of the data, extractors (medical practitioners) of the data, managers of the data, support and enablers of the data (IT/IS specialists), and those that make use of the data.

The interest of the actors comes from their involvement from different angles such as: (1) scheduling of medical appointments with patients; (2) consultations with medical personnel; (3) medical tests and treatment; (4) the use of various tools and medical apparatus; and (5) IT infrastructure and systems that are used to store, manage and retrieve the types of data.

### Analytics tools (level 'C')

The four most common types of data analytics tools are predictive, prescriptive, descriptive and diagnostic (Shao et al. 2014). In the context of healthcare, Raghupathi and Raghupathi (2014) stated that predictive analytics are used to anticipate risk through the analysis of historical health data and patterns. According to Rumsfeld et al. (2016), prescriptive analytics are used to support medical decisions on individual cases by assessing the risk and benefits of the available solutions. Mathew and Pillai (2015) stated that the descriptive analytics provide summary of the past and present data, which are used to inform healthcare decisions. Shao et al. (2014) stated that diagnostic analytics help in finding out why certain things are happening.

Each of these tools has the capability of adding value to the activities of healthcare, but from different perspectives. Therefore, there should be criteria for selecting the appropriateness of the tools. The choice of tool is determined by the healthcare facility's need from the existing big data. By establishing what they intend to use the big data for, the organisation is able to narrow down

what tool is best suitable for their goals. Thus, if an inappropriate or less appropriate tool is selected, there will be risks and challenges during analysis. This potentially results in incorrect diagnoses, medications and/or counselling.

## Validation of the framework

Although there exist some empirically validated artefacts, none can be used as a sole basis for research purposes, as they did not focus on developing countries. This makes this study an early exploratory study in the context of developing countries. As a result, it was difficult to validate the framework because of its uniqueness, 'healthcare big data analytics' in the context of developing countries. Thus, the influencing factors were used.

Although the data used in this study were based on literature, the framework was validated to ensure that the influencing factors (actors) are still applicable in the contexts of both developing countries and healthcare. This was against the existing model using the influencing factors, which include integration, structure, skill set, availability, data set, data integrity and translation. The IS frameworks can be validated through different ways, such as through case descriptions, framework or components of the framework (Mueller et al. 2010). Belle et al. (2015) argued that validation of framework can be objective or subjective. In this study, the influencing factors were used in validating the study, by employing a subjective approach.

Integration of patients' data sets, which include their risk profile, can be used to improve healthcare services (Elgendy & Elragal 2014). Availability of patients' big data is critical primarily for the purpose of achieving clinical predictive analytics (Bates et al. 2014). This helps to ensure ethical standards and manage privacy concerns. Data sets are most purposeful through predictive and diagnostics analytics tools, in developing decision for patient's healthcare (Wang & Hajli 2017). Belle et al. (2015) explained that skill sets are crucially essential in the use of big data analytics tools for decision-making and achieving healthcare solutions. Based on the validation, this study is suitable for gaining a better understanding in selecting analytics tools to improve the usefulness of big data in healthcare in developing countries.

## In practice

The factors revealed in this study are intended to contribute to the South African and other developing countries' healthcare environment from the following angles: in the development or review of policies, rules and regulations towards addressing some of the challenges that have been encountered in healthcare over the years.

In practice, the implementation of the framework requires formulation of templates for each of the levels, as depicted in Figure 1. The templates should consist of critical success

factors that are environment- and context-based. This is to ensure that the implementation of the framework appropriately guides the selection and use of analytics tools, making healthcare big data more useful and ultimately improving service delivery. Only then, the framework can help bridge that the gap created by lack of framework:

1. enlightening practitioners on the factors that influence big data analytics in the South African healthcare environment
2. being a step-by-step guide on factors considered prior to selecting a data analytics tool
3. improving the quality of services through the use of big data analytics.

In addition, prior to implementation of the framework, big data users (healthcare practitioners) have to be educated on the basics of big data and analytics. The framework can therefore further guide them, on step-by-step basis, in putting that knowledge into practice.

## Conclusion

This study helps to address some of the challenges that are encountered in the healthcare environment from an IS research viewpoint. A solution is proposed through a framework to provide a clearer understanding of the factors to consider prior to selecting and using big data analytics tools. This is primarily to increase the usefulness of the big data for healthcare services, particularly in developing countries. This study highlights that successful selection and implementation of big data analytics tools requires knowledge of components stated within the framework. Through the framework and the influencing factors, the study adds to academia in IS and health sciences' understanding of the use and roles of big data in the South African healthcare environment. In addition, the study can be of importance and benefit to the academics mainly because of its empirical nature.

This study addresses an area of big data analytics and healthcare in developing countries, which has not previously been explored. The framework can therefore be beneficially explored further in order to create artefact for validation credibility for future studies, in the areas of big data analytics, and healthcare big data in developing countries. Based on the analysis, findings and the interpretation, further research on this study is recommended. As the framework has not yet been applied, future studies could focus on the application of this framework on a healthcare-based case study. In addition, the use of different theories is encouraged.

## Acknowledgements

The author would like to thank the members of his research forum and the Department of Information Technology at Cape Peninsula University of Technology, Cape Town, for the assistance in conducting the research.

## Competing interests

The author has declared that no competing interest exists.

## Author's contributions

T.I. is the sole author of this research article.

## Ethical consideration

This article followed all ethical standards for a research without direct contact with human or animal subjects.

## Funding information

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

## References

- Abarda, A., Bentaleb, Y. & Mharzi, H., 2017, 'A divided latent class analysis for big data', *Procedia Computer Science* 110(2017), 428–433. <https://doi.org/10.1016/j.procs.2017.06.111>
- Abouzahra, M., 2011, 'Causes of failure in healthcare IT projects', in *3rd international conference on advanced management science*, vol. 19, November 4–6. IACSIT Press, Singapore.
- Alaboudi, A., Atkins, A., Sharp, B., Balkhair, A., Alzahrani, M. & Sunbul, T., 2016, 'Barriers and challenges in adopting Saudi telemedicine network: The perceptions of decision makers of healthcare facilities in Saudi Arabia', *Journal of Infection and Public Health* 9(6), 725–733. <https://doi.org/10.1016/j.jiph.2016.09.001>
- Archenaa, J. & Mary Anita, E.A., 2015, 'A survey of big data analytics in healthcare and government', *Procedia Computer Science* 50(1), 408–413. <https://doi.org/10.1016/j.procs.2015.04.021>
- Bare Bhakti, C. & Kini, S.N., 2017, 'Challenges and opportunities of big data analytics', *International Journal of Engineering Science* 7(6), 13713–13715.
- Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A. & Escobar, G., 2014, 'Big data in health care: Using analytics to identify and manage high-risk and high-cost patients', *Health Affairs* 33(7), 1123–1131. <https://doi.org/10.1377/hlthaff.2014.0041>
- Belle, A., Thiagarajan, R., Sorousmehri, S.M.R., Navidi, F., Beard, D.A. & Najarian, K., 2015, *Big data analytics in healthcare*, Hindawi Publishing Corporation, London, pp. 1–16.
- Bottles, K., Begoli, E. & Worley, B., 2014, 'Understanding the pros and cons of big data analytics', *Physician Executive* 40(4), 6–12.
- Busagala, L.S. & Kawono, G.C., 2013, 'Perceptions and adoption of information and communication technology for healthcare services in Tanzania', *International Journal of Computing & ICT Research* 7(1), 12–21.
- Callon, M., 1986, 'Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay', in J. Law (ed.), *Power, action & belief: A new sociology of knowledge?*, pp. 196–229, Routledge, London.
- Chang, B., Tsai, H., Tsai, C., Kuo, C. & Chen, C., 2016, 'Integration and optimization of multiple big data processing platforms', *Engineering Computations* 33(6), 1680–1704. <https://doi.org/10.1108/EC-08-2015-0247>
- Dwiartama, A. & Rosin, C., 2014, 'Exploring agency beyond humans: The compatibility of Actor-Network Theory (ANT) and resilience thinking', *Ecology and Society* 19(3), 28. <http://doi.org/10.5751/ES-06805-190328>
- Elgendy, N. & Elragal, A., 2014, 'Big data analytics: A literature review paper', in *Industrial conference on data mining*, Springer International Publishing, St Petersburg, Russia, July 16–20, 2014, pp. 214–227.



- Eswari, T., Sampath, P. & Lavanya, S., 2015, 'Predictive methodology for diabetic data analysis in big data', *Procedia Computer Science* 50(2015), 203–208. <https://doi.org/10.1016/j.procs.2015.04.069>
- Gandomi, A. & Haider, M., 2015, 'Beyond the hype: Big data concepts, methods, and analytics', *International Journal of Information Management* 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Ganjir, V., Sarkar, B.K. & Kumar, R.R., 2016, 'Big data analytics for healthcare', *International Journal of Research in Engineering, Technology and Science* 2(6), 1–6. <https://doi.org/10.1155/2015/370194>
- Hemon, R. & Williams, P.A., 2014, 'Big data in healthcare: What is it used for?' in *The Proceedings of the 3rd Australian eHealth Informatics and Security Conference*, Perth, Western Australia, 1–3 December.
- Hilbert, M., 2016, 'Big data for development: A review of promises and challenges', *Development Policy Review* 34(1), 135–174. <https://doi.org/10.1111/dpr.12142>
- Horowitz, L.S., 2012, 'Translation alignment: Actor-network theory, resistance, and the power dynamics of alliance in New Caledonia', *Antipode* 44(3), 806–827.
- Iyamu, T. & Mgodlwa, S., 2018, 'Transformation of healthcare big data through the lens of actor network theory', *International Journal of Healthcare Management* 11(3), 182–192. <https://doi.org/10.1080/20479700.2017.1397340>
- Iyamu, T., Nehemia-Maletzky, M. & Shaanika I., 2016, 'The overlapping nature of business analysis and business architecture: What we need to know', *The Electronic Journal Information Systems Evaluation* 19(3), 169–179.
- Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R. et al., 2014, 'Big data and its technical challenges', *Communications of the ACM* 57(7), 86–94. <https://doi.org/10.1145/2611567>
- Kalaga, T., 2015, *Literary hermeneutics: From methodology to ontology*, Cambridge Scholars, Cambridge.
- Kankanhalli, A., Hahn, J., Tan, S. & Gao, G., 2016, 'Big data and analytics in healthcare: Introduction to the special section', *Information Systems Frontiers* 18(2), 233–235. <https://doi.org/10.1007/s10796-016-9641-2>
- Kavitha, R., Kannan, E. & Kotteswaran, S., 2016, 'Implementation of cloud based Electronic Health Record (EHR) for Indian healthcare needs', *Indian Journal of Science and Technology* 9(3), 1–5. <https://doi.org/10.17485/ijst/2016/v9i3/86391>
- Kitchin, R., 2014, 'Big Data, new epistemologies and paradigm shifts', *Big Data & Society* 1(1), 1–12. <https://doi.org/10.1177/2053951714528481>
- Kumar, H. & Singh, N., 2017, 'Review paper on Big Data in healthcare informatics', *International Research Journal of Engineering and Technology* 4(2), 197–201.
- Kwon, O., Lee, N. & Shin, B., 2014, 'Data usage experience and acquisition intention of big data analytics', *International Journal of Information Management Data Quality Management* 34(2014), 387–389.
- Labrinidis, A. & Jagadish, H.V., 2012, 'Challenges and opportunities with big data', *Proceedings of the VLDB Endowment* 5(12), 2032–2033. <https://doi.org/10.14778/2367502.2367572>
- Law, J., 1992, 'Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity', *Systems Practice* 5(4), 379–393. <https://doi.org/10.1007/BF01059830>
- Lee, C.H. & Yoon, H.J., 2017, 'Medical big data: Promise and challenges', *Kidney Research and Clinical Practice* 36(1), 3–11. <https://doi.org/10.23876/j.krcp.2017.36.1.3>
- Li, D.R., Yao, Y. & Shao, Z.F., 2014 'Big Data in the Smart City', *Geomatics and Information Science of Wuhan University* 39(6), 630–640.
- Liu, W. & Park, E.K., 2014, 'Big data as an e-health service', in *2014 IEEE International Conference on Computing, Networking and Communications (ICNC)*, pp. 982–988.
- Luna, D.R., Mayan, J.C., García, M.J., Almerares, A.A. & Househ, M., 2014, 'Challenges and potential solutions for big data implementations in developing countries', *Yearbook of Medical Informatics* 23(1), 36–41. <https://doi.org/10.15265/IY-2014-0012>
- Malaka, I. & Brown, I., 2015, 'Challenges to the organisational adoption of big data analytics: A case study in the South African telecommunications industry', in *Proceedings of the 2015 annual research conference on South African institute of computer scientists and information technologists*, ACM, Stellenbosch, September 28–30, 2015, pp. 27.
- Mancini, M., 2014, 'Exploiting big data for improving healthcare services', *Journal of e-Learning and Knowledge Society* 10(2), 23–33.
- Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I.A.T., Siddiqa, A. et al., 2017, 'Big IoT data analytics: Architecture, opportunities, and open research challenges', *IEEE Access* 5, 5247–5261. <https://doi.org/10.1109/ACCESS.2017.2689040>
- Mathew, P.S. & Pillai, A.S., 2015, 'Big data solutions in healthcare: Problems and perspectives', in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, IEEE, Coimbatore, India, March 19–20, 2015, pp. 1–6.
- Mgodlwa, S. & Iyamu, T., 2018, 'Integration of social media with healthcare big data for improved service delivery', *South African Journal of Information Management* 20(1), 1–8. <https://doi.org/10.4102/sajim.v20i1.894>
- Micheni, E.M., 2015, 'Diffusion of big data and analytics in developing countries', *The International Journal of Engineering and Science* 4(8), 44–50.
- Mishra, S., Kalra, A. & Choudhary, K., 2013, 'Influence of information and communication technology in health sectors', *International Journal of Soft Computing and Engineering* 3(5), 66–68.
- Moktadir, M.A., Ali, S.M., Paul, S.K. & Shukla, N., 2019, 'Barriers to big data analytics in manufacturing supply chains: A case study from Bangladesh', *Computers & Industrial Engineering* 128, 1063–1075. <https://doi.org/10.1016/j.cie.2018.04.013>
- Mueller, B., Viering, G., Legner, C. & Riempp, G., 2010, 'Understanding the economic potential of service-oriented architecture', *Journal of Management Information Systems* 26(4), 145–180. <https://doi.org/10.2753/MIS0742-1222260406>
- Nambiar, R., Bhardwaj, R., Sethi, A. & Vargheese, R., 2013, 'A look at challenges and opportunities of big data analytics in healthcare', in *International Conference on Big Data, IEEE*, October 2013, pp. 17–22.
- Nanni, M., Thanos, C., Giannotti, F. & Rauber, A., 2015, 'Big Data Analytics: Towards a European research agenda', *ERCIM NEWS* 100, 9–10.
- Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M. & Ochiai, O., 2015, 'Big data challenges in building the global earth observation system of systems', *Environmental Modelling & Software* 68(2015), 1–26. <https://doi.org/10.1016/j.envsoft.2015.01.017>
- Peek, N., Holmes, J.H. & Sun, J., 2014, 'Technical challenges for big data in biomedicine and health: Data sources, infrastructure, and analytics', *Yearbook of Medical Informatics* 9(1), 42–47. <https://doi.org/10.15265/IY-2014-0018>
- Prasad, P., 2017, *Crafting qualitative research: Beyond positivist traditions*, Routledge, New York.
- Priyanka, K. & Kulennavar, N., 2014, 'A survey on big data analytics in health care', *International Journal of Computer Science and Information Technologies* 5(4), 5865–5868.
- Purkayastha, S. & Braa, J., 2013, 'Big data analytics for developing countries – Using the cloud for operational BI IN health', *The Electronic Journal of Information Systems in Developing Countries* 59(1), 1–17. <https://doi.org/10.1002/j.1681-4835.2013.tb00420.x>
- Raghupathi, W. & Raghupathi, V., 2014, 'Big data analytics in healthcare: Promise and potential', *Health Information Science and Systems* 2(1), 1–10. <https://doi.org/10.1186/2047-2501-2-3>
- Riggins, F.J. & Wamba, S.F., 2015, 'Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics', in *2015 48th Hawaii International conference on system sciences*, IEEE, Kaula, January 5–8, 2015, pp. 1531–1540.
- Rumsfeld, J.S., Joynet, K.E. & Maddox, T.M., 2016, 'Big data analytics to improve cardiovascular care: Promise and challenges', *Nature Reviews Cardiology* 13(6), 350. <https://doi.org/10.1038/nrcardio.2016.42>
- Sarkar, B.K., 2017, 'Big data for secure healthcare system: A conceptual design', *Complex & Intelligent Systems* 3(2), 133–151. <https://doi.org/10.1007/s40747-017-0040-1>
- Schmidt, L., 2016, *Understanding hermeneutics*, Routledge, London.
- Schüll, A. & Maslan, N., 2018, 'On the adoption of big data analytics: Interdependencies of contextual factors', in *ICEIS 1*, Funchal, March 21–24, 2018, pp. 425–431.
- Scotland, J., 2012, 'Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms', *English Language Teaching* 5(9), 1–8. <https://doi.org/10.5539/elt.v5n9p9>
- Shahbaz, M., Gao, C., Zhai, L., Shahzad, F. & Hu, Y., 2019, 'Investigating the adoption of big data analytics in healthcare: The moderating role of resistance to change', *Journal of Big Data* 6(1), 6. <https://doi.org/10.1186/s40537-019-0170-y>
- Shu, H., 2016, 'Big data analytics: Six techniques', *Geo-spatial Information Science* 19(2), 119–128. <https://doi.org/10.1080/10095020.2016.1182307>
- Song, T.M. & Ryu, S., 2015, 'Big data analysis framework for healthcare and social sectors in Korea', *Healthcare Informatics Research* 21(1), 3–9. <https://doi.org/10.4258/hir.2015.21.1.3>
- Sun, J. & Reddy, C.K., 2013, 'Big data analytics for healthcare', in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 1525–1525.
- Thiselton, A., 2009, *Hermeneutics: An introduction*, William Eerdmans, Cambridge.
- Walsham, G., 2015, *Interpreting information systems in organizations*, Cambridge Press, London.
- Wang, Y. & Hajli, N., 2017, 'Exploring the path to big data analytics success in healthcare', *Journal of Business Research* 70(C), 287–299. <https://doi.org/10.1016/j.jbusres.2016.08.002>
- Watson, H.J., 2014, 'Tutorial: Big data analytics: Concepts, technologies, and applications', *Communications of the Association for Information Systems* 34(1), 1247–1268. <https://doi.org/10.17705/1CAIS.03462>
- Yanow, D. & Schwartz-Shea, P., 2015, *Interpretation and method: Empirical research methods and the interpretive turn*, Routledge, New York.
- Yu, T.K., Lin, M.L. & Liao, Y.K., 2017, 'Understanding factors influencing information communication technology adoption behavior: The moderators of information literacy and digital skills', *Computers in Human Behaviour* 71(2017), 196–208. <https://doi.org/10.1016/j.chb.2017.02.005>