International Journal of Big Data and Analytics in Healthcare
Volume 8 • Issue 1

Data Visualization of Big Data for Predictive and Descriptive Analytics for Stroke, COVID-19, and Diabetes

Richard S. Segall, Arkansas State University, USA*
https://orcid.org/0000-0001-7627-2609

Soichiro Takashashi, Arkansas State University, USA

ABSTRACT

Visualization of big data is crucial for meaningful interpretations and especially for healthcare. Brief discussions are made for big data, background for healthcare, and recent work in big data analytics for healthcare. This research pertains to different levels of big data: 5,110 vs. 101,766 vs. 320,200 vs. 1 million data values. Data visualizations and predictive analytics are presented of big data for selected diseases of stroke with 5,110 data values, diabetes with 101,766 data values, and two COVID-19 studies: one with 320,200 data values and another with 1 million data values. Data visualizations are generated for these diseases with big data using Tableau. For stroke patients, an investigation was performed to determine how different living environments affect relationship between strokes. The data visualizations for diabetes showed impact of insulin use yielded reduced hospital stays. Data visualizations for COVID-19 provided temporal trends in confirmed cases, mortality, and recovery rates for 2020-2023. Conclusions and future directions of research are presented.

KEYWORDS

Big Data, Big Data Analytics, COVID-19, Descriptive Analytics, Diabetes, Predictive Analytics, Stroke, Visualization

INTRODUCTION

Big data was originally defined as the collection of datasets whose volume, velocity, or variety is so large that storing, managing, processing, and analyzing the data using traditional databases and data processing tools is complex (Bahga & Madisetti, 2016). According to an estimate in 2017 by IBM, 2.5 quintillion bytes of data is created daily, and 90% of the data in the world today was created in the last two years alone (Perry, 2017). Miele and Shockley (2013) authored a twenty-page IBM Executive Report titled "Analytics: The Real-World Use of Big Data" which is one of the first detailed discussions of big data.

In 2012, the United States (U.S.) government committed $200 million in big data research and development investment (The White House, 2012). Big data applications are estimated to be worth

DOI: 10.4018/IJBDAH.331996
*Corresponding Author

300 billion dollars for the U.S. healthcare industry and 250 billion euros for Europe's public sector administration (Manyika et al., 2011). So, what is big data? The numerical definition of big data is evolving with technological development. A dynamic definition is data that exceeds the capacity of commonly used hardware and software tools to capture, store, and analyze within a tolerable elapsed time is big data (Franks, 2012). Clegg (2017) authored a book on how the information revolution of big data is transforming our lives.

While Segall (2020a) discussed the crucial question "what is big data?," Segall (2020b) discussed open-source software for big data. Segall and Niu (2020) wrote an entire book on open-source software for the statistical analysis of big data. Segall and Cook (2018) completed a two-volume handbook of big data storage and visualization techniques.

This paper specifically addresses big data for three selected diseases of strokes, diabetes, and COVID-19 for databases at different levels of big data (5,110 vs. 101,766 vs. 320,200 vs. 1 million data values).

RESEARCH BACKGROUND AND MOTIVATIONS

Yee et al. (2020) discussed the implications of big data on healthcare and its future steps with uses for clinical decision-making, research and development, population health and surveillance, detecting fraud, prediction capabilities, Google Trends, and preventive measures. They referred to Chen et al. (2016), describing a cognitive computing tool developed by IBM. The tool, named Watson, has been applied to big data challenges in life sciences research by integrating and analyzing big data that includes medical literature, patents, genomics, and chemical and pharmacological data. Chen et al. (2016) specifically discussed the application of IBM Watson to explore big data for cancer kinases.

Healthcare applications that have used big data include those for cancer research, disease detection, and population health. Big data has changed how researchers understand diseases, providing access to patient information, trends, and patterns that were not accessible before. Companies that use big data in healthcare applications include:

- Cancer research carried out by Tempus in Chicago, Illinois (USA) and Flatiron Health in New York City (USA)
- Early disease detection by Pieces in Irving, Texas (USA) and Prognos in New York City (USA)
- Population health research conducted by Amitech in Creve Coeur, Missouri (USA), Linguamatics in Marlborough, Massachusetts, and Socially Determined in Washington, DC (USA). (Schroer, 2023)

Pramanik et al. (2022) provided a comprehensive overview of healthcare big data that extends the traditional 5 V's to 10 V's for healthcare big data: Volume, Velocity, Variety, Veracity, Validity, Viability, Volatility, Vulnerability, Visualization, and Value. Each of these are defined as below in Table 1 where the traditional 5 V's are listed as the first five.

Hogue and Bao (2016) discussed the challenges of big data in health care and analytics tools such as MapReduce, Spark, and Storm. Patil and Vohra (2021) edited a handbook of research that focuses on healthcare applications of data science and analytics.

Varatharajan et al. (2020) discussed an in-depth study of big data analytics in the healthcare industry by analyzing healthcare applications using machine learning. Varatharajan et al. (2020) discussed the need for big data analytics in the healthcare ecosystem with personal health records, electronic health records, health information exchange, and national and international health analytics.

Varatharajan et al. (2020) compared the biggest challenges in big data analytics of data complexity, data volumes, performance, skills, data velocity, and cost for Boston, San Francisco, and Chicago.

Related Content

Artificial Intelligence Framework for Opinion Mining of Netizen Readers' Reviews of Arundhati Roy's The God of Small Things

Pantea Keikhosrokiani, Moussa Pourya Asl, Kah Em Chuand Nur Ain Nasuha Anuar (2023). *Handbook of Research on Artificial Intelligence Applications in Literary Works and Social Media (pp. 68-92).*
www.irma-international.org/chapter/artificial-intelligence-framework-for-opinion-mining-of-netizen-readers-reviews-of-arundhati-roys-the-god-of-small-things/317156