



Journal of Population Therapeutics & Clinical Pharmacology

Research Article

DOI: 10.47750/jptcp.2022.974

The importance of big data for healthcare and its usage in clinical statistics of cardiovascular disease

Johanes Fernandes Andry¹, Hendy Tannady^{2,*}, Glisina Dwinoor Rembulan³, Antonius Rianto¹

¹Information Systems Department, Universitas Bunda Mulia, Jakarta, Indonesia

²Department of Management, Universitas Multimedia Nusantara, Banten, Indonesia

³Industrial Engineering Department, Universitas Bunda Mulia, Jakarta, Indonesia

***Corresponding author:** Hendy Tannady, Department of Management, Universitas Multimedia Nusantara, Banten, Indonesia. Email: hendy.tannady@umn.ac.id

Submitted: 21 August 2022. Accepted: 19 September 2022. Published: 26 November 2022.

ABSTRACT

In the era of technological trends, large statistics have been broadly carried out in diverse businesses, especially healthcare. An extensive amount of data has unfolded new gaps in fitness care. The immense facts in healthcare have the capability to improve healthcare to a higher level. Large records can correctly lessen healthcare problems such as the selection of the appropriate remedy, solution for healthcare, and enhancing the healthcare machine. There are six defining attributes in large data, namely, extent, range, speed, veracity, variability and complexity, and value. Massive information represents an expansion of possibilities that could enhance the performance of healthcare. The large data in healthcare should help in the advanced use of massive data analytics to gain valuable know-how. This large information analytics is used to get valuable facts from all types of sources in healthcare that may be used to take advantage of the data in order to make better choice in healthcare. The massive information analytics can enhance healthcare by discovering institutions and expertise styles and trends in scientific facts. Cardiovascular disorder datasets are massive data in healthcare, and they are used as part of facilitating the system of documenting scientific facts that must be analyzed to offer powerful answers to troubles in fitness care. This paper offers valuable statistics by using massive information analytics from clinical statistics of cardiovascular disease

J Popul Ther Clin Pharmacol Vol 29(4):e107–e115; 26 November 2022.

This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International License. ©2022 Andry JF et al.

to provide convincing answers for the troubles in healthcare and also to indicate how huge information is essential for healthcare.

Keywords: *big data analytics; cardiovascular disease; healthcare*

INTRODUCTION

As the world advances at a fast pace, virtual generation continues to develop. As it is one of the emerging international locations, Indonesia additionally faces the growth of digital generation innovation.¹ There are many facts available stating the wider application of technology in a corporation. There was rapid development of information and technology in various businesses including healthcare offerings.² Scientific facts are constantly explosive, and there are demanding situations for data management, storage, and processing.³ Efficient data acquisition, processing, and intake methodologies have been areas of top-notch interest for many years throughout business enterprises.⁴ Rich source of facts can enrich the knowledge of ailment mechanisms and better health care.⁵

Big data refers to huge and complex information sets which might be beyond the capability of conventional control systems to save, control, and manage.⁶ Big data applications represent numerous possibilities to enhance the performance of healthcare.⁷ Big data can efficiently lessen fitness care troubles including the selection of suitable remedies, improving the fitness care machine, and so forth.⁸ Big data in healthcare refers to massive and complex electronic fitness datasets that they are difficult to manipulate with traditional software, hardware, information management tools, and techniques.⁹

Big data creates demanding situations in information size, transfer, encryption, storage, evaluation, and visualization. Healthcare relies on medical information on the method of selection. Large records analytics may be used to reap treasured facts from all forms of resources which might be too massive, raw, or unstructured in healthcare.¹⁰

Big data analytics applications in healthcare take the benefit of extracting insights from statistics for higher decisions making through analyzing the significant amount of records, from specific information resources, and in diverse codecs.⁹ Big data analytics has the potential to improve healthcare by using institutions and understanding styles and traits within the scientific data.¹¹

Healthcare studies have been growing exponentially in the past few years. In developed countries, the healthcare industry offers massive volumes of digital health information which includes cardiovascular disease.¹² Big data analytics can be used to achieve treasured facts from large and complicated datasets along with cardiovascular disease to enhance scientific treatment and healthcare.¹⁰ Cardiovascular ailment is one of the services in healthcare and is used as a component to facilitate the method of documenting scientific facts.¹³ Structured healthcare data are a vital useful resource for healthcare informatics research in predictive modeling. Massive scientific records have a large analytical capability that may be used to provide effective solutions for problems in the healthcare domain.¹⁴

LITERATURE REVIEW

Big Data

Big data typically refers to the considerable volumes of facts that the old information tools and practices are not prepared to address and presents extraordinary possibilities to strengthen technology and inform aid control via records-intensive approaches, and massive records technologies are permitting new varieties of activism environment

within the procedure.^{15,16} Big data has six defining attributes, which are volume, range, velocity, veracity, variability and complexity, and value. The term quantity represents the importance of the information; variety is the structural heterogeneity in a record set; velocity is the fee of producing facts; veracity is the unreliability inherent in records resources; complexity represents the version in statistics flow price; and fee measures the facts extracted from historic incident statistics units for highest quality manipulate choice.¹⁷

Big data has the functions that are high dimensional, heterogeneous, complex, unstructured, incomplete, and noisy which makes it possible to collect precious statistics.¹⁸ The main resources of big data in healthcare are administrative databases, scientific databases, electronic fitness file information, and laboratory statistics systems information that can improve healthcare by using institutions and knowledge styles within the scientific statistics.^{19,20}

Big data analytics

Massive facts analytics describe the procedure of accumulating, organizing and studying large statistics to discover styles, unknown correlations, market trends, personal alternatives, and other valuable facts that could not be analyzed with traditional tools.²¹ Big statistics analytics is a hard and fast technology and technique that requires new types of integration to discover hidden values from huge statistics which can be distinct from the standard ones, greatly complicated, and of a big scale.²²

The commonplace varieties of big facts analytics include predictive, diagnostic, descriptive, and prescriptive analytics to extract exceptional styles of information for one-of-a-kind purposes.²³ Huge records analytics in healthcare is hard and fast type of methodologies, tactics, frameworks, and technology which are used to transform facts into meaningful as well as useful records. These sets of information are used to make the more powerful selection methods for healthcare.²⁴

Big healthcare data

Massive healthcare information includes huge collections of statistics from numerous healthcare foundations observed through storing, handling, reading, visualizing, and turning in facts for effective decision-making. Huge healthcare information may be dependent, semi-structured, or unstructured²² and can be received from number one sources (medical selection aid structures, electronic fitness records, etc.) and secondary assets (laboratories, coverage organizations authorities sources, pharmacies, and so forth.).²⁵

Collection, agency, annotation, storage, and distribution of big data are vital activities in biomedical, medical, and translational discovery strategies. Massive healthcare records have the capacity to improve diagnostic signs and symptoms, predict epidemics, advantage precious understanding, avoid preventable sicknesses, reduce the fee of healthcare, and enhance the excellence of healthcare.²⁶

METHODS

Big data analytics refers to the strategy of analyzing a large extent of statistics accrued from numerous sources in an unstructured, semi-dependent, or based form with the aid of the use of special analytical strategies.²⁷ The common styles of large facts analytics are predictive, diagnostic, descriptive, and prescriptive that are implemented to extract extraordinary styles of know-how or insights from huge facts that can be used for exclusive functions relying upon the application area.²³ Figure 1²⁸ indicates the stages of the study.

Big records analytics includes stringent analytical methodologies and gear which comprise correlations, cluster analysis, filtering, selection bushes, Bayesian evaluation, neural community evaluation, regression evaluation, and textural evaluation.²⁹ Table 1³⁰ indicates the method of huge information analytics that consists of numerous strategies.

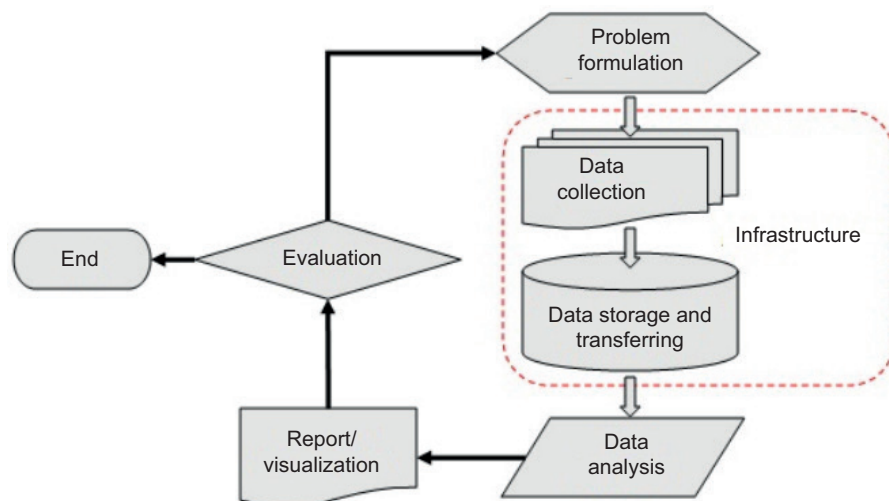


FIG 1. Research stages.

TABLE 1. Big data analytics process.

BDA step	Critical question	Epistemological challenge	Possible guidance
Acquisition	What records do I need? What kinds of datasets which might be used to decide?	Sampling	Observe statistics summarization, graphical representation, measurement discount (e.g. PCA), and outlier detection
Pre-processing	How the datasets may be represented and processed without falsification or understanding loss?	Quality of data	Make sure multi-expert and multidisciplinary participation in data reduction and choice hint and observe all levels of extraction, transformation, loading, and merging for completeness, correctness, and consistency
Analytics	Which techniques to use and what rules govern conclusions from those datasets?	Knowledge discovery	Map the constructs of analytics for theoretical ideas and expand or practice framework for preference of techniques consisting of mining, system mastering, records, or models
Interpretation	The way to interpret such conclusions?	Interpretability and reliability of estimation	Develop or practice the theoretical framework for interpreting the result

RESULTS AND DISCUSSION

Big amounts of records, pushed through document keeping, regulatory compliance and necessities, and patient care are generated using

healthcare. Healthcare used massive statistics analytics to investigate data to get valuable facts to improve healthcare performance. Large information analytics can help early detection of disorder, correct prediction of disorder, identification of

deviation from the healthy kingdom, and detection of fraud.

RapidMiner is a software program that we use to investigate facts with algorithms to get useful statistics for healthcare. We analyzed the records using a classification technique and a choice tree set of rules to classify cardiovascular sickness which may be used as predictive and prescriptive analytics. Predictive evaluation predicts what might take place at the end and prescriptive evaluation recommends actions that may be taken to act on those results.

We used cardiovascular sickness datasets that had been accumulated at the moment of scientific examination in this research. There are three styles of capabilities on these datasets specifically objective (actual statistics), examination (results of the clinical exam), and subjective (statistics given by the patient). This dataset has 70,000 rows of patient records and 12 attributes of facts for patients and the outcomes of the clinical exam. Following are the attributes of the cardiovascular disease datasets:

1. Age: authentic records of approximate patient's age (in days).
2. Height: facts about the patient's height (in cm).
3. Weight: factual records of approximately affected person's weight (in kg).
4. Gender: authentic facts of approximately affected person's gender (1 is a female and a couple of is a man).
5. Systolic blood stress: effects of clinical examination from patient's systolic blood stress (in mmHg).
6. Diastolic blood pressure: consequences of scientific examination from affected person's diastolic blood pressure (in mmHg).
7. LDL cholesterol: outcomes of scientific examination from affected person's LDL cholesterol (1: ordinary; 2: above every day; three: nicely above every day).
8. Glucose: effects of clinical exam from patient's glucose (1: normal; 2: above ordinary; 3: properly above regular).
9. Smoking: statistics are given via the patient about smoking (zero: non-smoker; 1: smoker).
10. Alcohol consumption: information is given via the affected person's approximate alcohol intake (0: no longer an alcohol drinker; 1: alcohol drinker).
11. Physical hobby: records given through the affected person about physical activity (zero: no bodily interest; 1: have physical pastime).
12. Presence or absence of cardiovascular disorder: records from analytics of the use of a decision tree set of rules about the presence of cardiovascular sickness (0: absence; 1: presence).

Within the preprocessing section, we remodel raw data right into a beneficial and efficient layout. We explore our cardiovascular disease datasets for cleansing of facts. In this phase, we identify lacking attributes and blank fields, cleaning or change lacking values, replica or incorrect statistics, and inconsistent facts. We examine the facts for completeness, correctness, and consistency. Intricate statistics that have not been recognized and analyzed can produce deceptive effects. This section is crucial to produce accurate outcomes with the aid of processing the analyzed facts.

After preprocessing the segment, we put the information into RapidMiner to start the analytics. Before analyzing the information, we set the characteristic type based on the facts type. We have to set the attribute kind effectively to be able to produce correct results. We examine the information on the usage of a decision tree set of rules in RapidMiner to do analytics. Choice tree set of rules is used to categorize the presence or absence of cardiovascular ailment. First, we gather the information using a choice tree algorithm to generate the guidelines and selection tree and then analyze the effects.


```

ap_hi > 138.500
| ap_lo > 5
| | height > 190.500: 1 {0= 0, 1= 20}
| | height ≤ 190.500
| | | weight > 168.500: 1 {0= 0, 1= 9}
| | | weight ≤ 168.500
| | | | weight > 166.500: 0 {0= 1, 1= 1}
| | | | weight ≤ 166.500
| | | | | weight > 158.500: 1 {0= 0, 1= 9}
| | | | | weight ≤ 158.500
| | | | | | weight > 155.500
| | | | | | | age > 19339: 1 {0= 0, 1= 2}
| | | | | | | age ≤ 19339: 0 {0= 3, 1= 0}
| | | | | | | weight ≤ 155.500: 1 {0= 3130, 1= 16225}
| ap_lo ≤ 5: 0 {0= 8, 1= 3}
ap_hi ≤ 138.500
| age > 23681: 1 {0= 0, 1= 5}
| age ≤ 23681
| | height > 58
| | | weight > 21.500
| | | | ap_hi > -130
| | | | | ap_hi > 129.500: 1 {0= 3719, 1= 5541}
| | | | | ap_hi ≤ 129.500
| | | | | | age > 22269
| | | | | | | weight > 38.500
| | | | | | | | height > 191.500: 0 {0= 3, 1= 0}
| | | | | | | | height ≤ 191.500: 1 {0= 1853, 1= 2382}
| | | | | | | | weight ≤ 38.500: 0 {0= 4, 1= 0}
| | | | | | | | age ≤ 22269: 0 {0= 26300, 1= 10776}
| | | | | | | | | ap_hi ≤ -130: 1 {0= 0, 1= 2}
| | | | | | | | | weight ≤ 21.500: 1 {0= 0, 1= 2}
| | | | | | | | | height ≤ 58: 1 {0= 0, 1= 2}
    
```

FIG 2. Decision tree rules.

The decision tree algorithm offers a choice tree to find the classification rules from the facts. In that decision tree, the foundation node or predictor is ap_hi (systolic blood pressure), internal nodes are different attributes that carry the information and the leaf node is the aerobic (Presence or Absence of cardiovascular disease). The result of the decision tree from cardiovascular ailment is used to explain or understand the result from classification based on the alternative attributes such as the root node and inner nodes that determine the presence or absence of cardiovascular ailment because of the leaf nodes.

The gain of a decision tree may be defined as rules or descriptions from a decision tree. The rules

TABLE 2. Result of performance testing.

Accuracy: 72.17%			
	True 0	True 1	Class precision
Pred. 0	26319	10780	70.94%
Pred. 1	8702	24199	73.55%
Class recall	75.15%	69.18%	

or descriptions from a decision tree is an if-else statement. Figure 2 shows the rules from the decision tree.

These rules are generated from the selection tree beginning from the foundation node or predictor till the leaf node. Those policies deliver a clean analytical view of the result from the selection tree. We will understand all the manners from the selection tree using those regulations.

After the selection tree is generated from the cardiovascular disease datasets using RapidMiner, we additionally check for the performance. We carry out performance testing to determine whether the analyzed information is accurate or not. Table 2 shows the performance testing result.

The overall performance testing indicates the extent of accuracy, class precision, and sophistication. We acquired 70.17% for the level of accuracy in the category. The elegance precision for prediction 0 or Absence is 70.9% and the magnificence precision for prediction 1 or Presence is 70.6%. For the class do not forget, we acquired 75.15% for actual 0 or Absence and 69.2% for real 1 or Presence. The measured degree of accuracy, precision, and do not forget achieved high performance. This means that a class progressed the efficiency and effectiveness of cardiovascular disease.

Big data analytics with category method and the usage of choice tree algorithm for the cardiovascular disorder can improve the performance and effectiveness. The selection tree presents the cause of the hidden sample from the records so that we can apprehend the records that we get from the facts using a choice tree algorithm. We also examined

the effects from a category using a selection tree as predictive or prescriptive analytics. Healthcare can predict an affected person who has a cardiovascular disorder and provide preventive care to the affected person. Big facts in healthcare are vital, right here is the opportunity for healthcare that implemented big information:

1. *Improved preventive care* – The use of analytics in healthcare scientific records can improve prevention for the affected person. With big facts analytics, healthcare can seize, analyze, and examine affected person signs. Healthcare is advanced with preventive care that can treat the patient properly and prevent or postpone the illness and disorder. As in the research, we classify cardiovascular sickness so that healthcare can understand the right remedy for the patient to be more powerful and effective.
2. *Improved diagnostic symptoms* – By doing large statistics analytics, healthcare has advanced with a thorough diagnosis of symptoms in patients. Diagnostic signs are a system to decide patients with the disease. Improved diagnostic signs and symptoms are gathered from the hidden sample of patients' facts. With advanced diagnostic signs and symptoms, healthcare can diagnose the patients with greater efficacy. Within the studies, we classify cardiovascular disorder having numerous attributes to decide the sickness so that healthcare can diagnose the affected person with greater accuracy from their signs.
3. *Reducing healthcare cost* – Big data can assist in reducing the fee of imparting clinical treatment. Massive statistics analytics for healthcare can carry valuable data to improve their gadget through discovering associations, knowledge patterns, and developments within scientific records.

With analyzed information, their improved scientific remedy can examine and diagnose the patient to a greater extent. With this effective and efficient healthcare system, a patient pays less and gets an accurate remedy than the normal scientific treatment. Further, the presence and clarity of presidential guidelines and laws become necessary to reduce uncertainty or dangers related to the usage of databases and shield the purchaser's hobby.¹

CONCLUSIONS

In this research, we used category techniques with a decision tree algorithm for cardiovascular disease datasets. The outcomes from this research are the decision tree and guidelines that classify the presence or absence of cardiovascular disease. We additionally take a look at the overall performance to decide whether the analyzed information is accurate or not. We acquired 72.17% for the level of accuracy in type. The elegance precision for prediction 0 or Absence is 70.94% and the class precision for prediction 1 or Presence is 73.55%. For the class recall, we obtained 75.15% for true 0 or Absence and 69.18% for true 1 or Presence. The analyzed information can be used for predictive and prescriptive analytics. Big data analytics with classification methods using decision tree algorithms for cardiovascular disease can improve the efficiency and effectiveness of healthcare. Healthcare can treat a patient who has cardiovascular sickness and offers preventive care to the patient. With the help of big facts, healthcare has the opportunity to improve better healthcare, which includes advanced preventive care, progressed diagnostic signs, and decreased healthcare prices. In this research, classification methods using the choice tree algorithm can be used for other datasets and can be advanced by combining or evaluating the use of different category algorithms to get better effects.

ETHICAL APPROVAL

All procedures were conducted with allowance and followed the regulation of all parties involved as objects in this research, and also with the permission of related universities.

ACKNOWLEDGMENTS

We thank those who have helped us in implementing this research; especially, Universitas Bunda Mulia and Universitas Multimedia Nusantara and those involved in this research directly or indirectly.

CONFLICTS OF INTEREST

The authors do not have any conflicts of interest in the present work.

AUTHORS' CONTRIBUTION

Johanes Fernandes Andry, Hendy Tannady, Glisina Dwinoor Rembulan, and Antonius Rianto participated in the design, data collection, figures, and analysis with similar roles.

REFERENCES

1. Kurniasari F, Putri FP, Firmansyah A. The role of financial technology to increase financial inclusion in Indonesia. *Int J Data Netw Sci.* 2021;5:391–400. <https://doi.org/10.5267/j.ijdns.2021.5.004>
2. Kurniasari F, Hamid NA, Qinghui C. The effect of perceived usefulness, perceived ease of use, trust, attitude and satisfaction into continuance of intention in using alipay. *Management & Account Review.* 2020;19(2):131–150. <http://dx.doi.org/10.24191/mar.v19i2.1190>
3. Zhang Y, Qiu M, Tsai CW, Hassan MM, Alamri A. Health-CPS: Healthcare cyber-physical system assisted by cloud and big data. *IEEE Syst J.* 2017;11(1):88–95. <https://doi.org/10.1109/JSYST.2015.2460747>
4. Schultz T. Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle. *Bull Am Soc Inf Sci Technol.* 2013;39(5):34–40; <https://doi.org/10.1002/bult.2013.1720390508>
5. Chawla NV, Davis DA. Bringing big data to personalized healthcare: A patient-centered framework. *J Gen Intern Med.* 2013;28(3):660–5. <https://doi.org/10.1007/s11606-013-2455-8>
6. Nambiar R, Bhardwaj R, Sethi A, Vargheese R. A look at challenges and opportunities of Big Data analytics in healthcare. *Proc - 2013 IEEE Int Conf Big Data, Big Data 2013.* 2013; pp. 17–22. <https://doi.org/10.1109/BigData.2013.6691753>
7. Zillner S, Neururer S. Technology roadmap development for big data healthcare applications. *KI – Kunstl Intelligenz.* 2015;29(2):131–41. <https://doi.org/10.1007/s13218-014-0335-y>
8. Jee K, Kim GH. Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system. *Healthc Inform Res.* 2013;19(2):79–85. <https://doi.org/10.4258/hir.2013.19.2.79>
9. Madyatmadja ED, Rianto A, Andry JF, Tannady H, Chakir A. Analysis of big data in healthcare using decision tree algorithm. *1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), IEEE Conference Proceeding, Indonesia, 2021.*
10. Wang L, Alexander CA. Big data analytics in healthcare systems. *Int J Math Eng Manag Sci.* 2018;4(1):269–76. <https://doi.org/10.1109/ICoAC44903.2018.8939061>
11. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. *Heal Inf Sci Syst.* 2014;2(1):1–10. <https://doi.org/10.1186/2047-2501-2-3>
12. Srinivasan U, Arunasalam B. Leveraging big data analytics to reduce healthcare costs. *IT Prof.* 2013; 15(6):21–8. <https://doi.org/10.1109/MITP.2013.55>
13. Prasad A, Prasad S. Imaginative geography, neoliberal globalization, and colonial distinctions: Docile and dangerous bodies in medical transcription outsourcing. *Cult Geogr.* 2012;19(3):349–64. <https://doi.org/10.1177/1474474012445734>
14. Bouhriz M, Chaoui H. Big data privacy in healthcare moroccan context. *Procedia Comput Sci.* 2015; 63:575–80. <https://doi.org/10.1016/j.procs.2015.08.387>
15. Hasan SS, Zhang Y, Chu X, and Teng Y. The role of big data in China's sustainable forest management.

- For Econ Rev. 2019;1(1):96–105. <https://doi.org/10.1108/fer-04-2019-0013>
16. Madyatmadja ED, Liliana L, Andry JF, Tannady H. Risk analysis of human resource information systems using Cobit 5. *J Theor Appl Inf Technol*. 2020;98(21):3357–67. <http://www.jatit.org/volumes/Vol98No21/4Vol98No21.pdf>
 17. Ajayi A, Oyedele L, Delgado JMD, Akanbi L, Bilal M, Akinade O, Olawale O. Big data platform for health and safety accident prediction. *World J Sci Technol Sustain Dev*. 2019;16(1):2–21. <https://doi.org/10.1108/wjstsd-05-2018-0042>
 18. Tsai CW, Lai CF, Chao HC, Vasilakos AV. Big data analytics: A survey. *J Big Data*. 2015;2(1):1–32. <https://doi.org/10.1186/s40537-015-0030-3>
 19. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. *Int J Med Inform*. 2018;114(March):57–65. <https://doi.org/10.1016/j.ijmedinf.2018.03.013>
 20. Madyatmadja ED, Marvell JF, Andry, Tannady H, Chakir A. Implementation of big data in hospital using cluster analytics. International Conference on Information Management and Technology (ICIMTech), IEEE Conference Proceeding, Jakarta, Indonesia, 2021.
 21. Hariri RH, Fredericks EM, Bowers KM. Uncertainty in big data analytics: Survey, opportunities, and challenges. *J Big Data*. 2019;6(1):1–16. <https://doi.org/10.1186/s40537-019-0206-3>
 22. Riahi Y, Riahi S. Big data and big data analytics: Concepts, types and technologies. *Int J Res Eng*. 2018;5(9):524–8. <https://doi.org/10.21276/ijre.2018.5.9.5>
 23. Bibri SE, Krogstie J. The core enabling technologies of big data analytics and context-aware computing for smart sustainable cities: A review and synthesis. *J Big Data*. 2017;4(1):1–50. <https://doi.org/10.1186/s40537-017-0091-6>
 24. Patel S, Patel A. A big data revolution in health care sector: Opportunities, challenges and technological advancements. *Int J Inf Sci Tech*. 2016;6(1/2):155–62. <https://doi.org/10.5121/ijist.2016.6216>
 25. SA S. Big data in healthcare management: A review of literature. *Am J Theor Appl Bus*. 2018;4(2):57. <https://doi.org/10.11648/j.ajtab.20180402.14>
 26. Soleimani-Roozbahani F, Rajabzadeh Ghatari A, Radfar R. Knowledge discovery from a more than a decade studies on healthcare Big Data systems: A scientometrics study. *J Big Data*. 2019;6(1):1–15. <https://doi.org/10.1186/s40537-018-0167-y>
 27. Shabbir MQ, Gardezi SBW. Application of big data analytics and organizational performance: The mediating role of knowledge management practices. *J Big Data*. 2020;7(1):1–17. <https://doi.org/10.1186/s40537-020-00317-6>
 28. Acharjya DP, Kauser Ahmed P. A survey on big data analytics: Challenges, open research issues and tools. *Int J Adv Comput Sci Appl*. 2016;7(2):511–18. <https://doi.org/10.26438/ijcse/v6i6.12381244>
 29. Sarkar BK. Big data for secure healthcare system: A conceptual design. *Complex Intell Syst*. 2017;3(2):133–51. <https://doi.org/10.1007/s40747-017-0040-1>
 30. Elragal A, Klischewski R. Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *J Big Data*. 2017;4(1):1–20. <https://doi.org/10.1186/s40537-017-0079-2>