World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

# Leveraging big data analytics for enhanced healthcare insurance: A new computational approach to efficiency and cost reduction

Lakshmi Narasimhan Srinivasagopalan *

*Texas USA.*

World Journal of Advanced Research and Reviews, 2022, 13(03), 563–576

Publication history: Received on 27 January 2022; revised on 17 March 2022; accepted on 20 March 2022

Article DOI: https://doi.org/10.30574/wjarr.2022.13.3.0115

## Abstract

Healthcare insurance systems, especially in regions like India, face persistent challenges in cost management, resource allocation, and the timely processing of claims and data. This paper introduces an innovative data-driven approach to healthcare insurance that harnesses Big Data Analytics combined with a novel analytical framework designed for greater operational efficiency. We propose a hybrid method integrating Set Theory, MapReduce, Association Rule Mining, and MongoDB to create a robust model for healthcare insurance data analytics. Our method, developed to analyze large datasets accurately, aims to reduce processing delays and enhance decision-making by identifying patterns in patient claims and provider submissions. A prototype was developed using Java and evaluated on data from the National Health Insurance Scheme (NHIS) in India, achieving an accuracy rate of 98.3%, a notable improvement over existing models. The results demonstrate significant reductions in delays within data processing, streamlining both resource flow and data management. This model not only enhances data accuracy but also contributes to a more sustainable healthcare financing structure, marking a step forward in aligning healthcare services with modern technological advances.

**Keywords:** Big Data; Healthcare; Insurance; Cloud; Data Base**.**

## 1. Introduction

The health care industry plays a crucial role in today's world because of the ever-increasing population and the increasing difficulty of maintaining optimal health due to factors like unhealthy diet, excessive alcohol consumption, smoking, and other lifestyle choices. A great deal of effort is being devoted to discovering treatments for dangerous diseases, particularly in the medical field. The health care sector [1] is confronted with a huge problem in offering corrective remedies and raising global awareness as new diseases are being reported all over the world. It has become extremely difficult for doctors to provide better treatment for patients due to the widespread lack of knowledge about numerous diseases. Patients need to undergo a battery of tests in order to get a diagnosis, which puts a heavy financial strain on lower-middle-class individuals.

Traditional storage techniques are finding it increasingly difficult to keep up with the exponential growth of data in the health industry [2]. Accordingly, the health sector offers enormous potential for future study. The health care industry is making a killing off of people's suffering from both recognized and undiscovered diseases, which is driving up the cost of healthcare overall. Furthermore, patients often wind up spending more money on therapy than necessary since they are unaware of the signs and symptoms of the diseases they are afflicted with. There is a high risk of data mixing when transferring patient records across different database repositories, making this process extremely difficult [3]. As a result, studies in the healthcare sector can now focus on finding ways to lower the costs associated with collecting

* Corresponding author: Lakshmi Narasimhan Srinivasagopalan

patient data. Besides this, a lot of hospitals spend a lot of money on medicine from different pharmaceutical companies. Research on different diseases is needed to find ways to reduce the cost of drugs, which would help to reduce this cost.

New innovations and an improved organization have both relied on information. We can better organize ourselves to produce the best results when we have more information. Data collection is a crucial component of any organization for this simple reason. Predicting future events and the present trend of specific metrics is another possible application of this data. We have begun to generate and gather more data about nearly everything by introducing technology advancements in this area as our awareness of this grows. In today's world, there is an overwhelming amount of data coming at us from all directions: social, scientific, occupational, health-related, etc. The current state of affairs is analogous, to some extent, to a data flood. Our ability to generate data has grown exponentially thanks to technology advancements, to the point that it is now beyond the capabilities of our present-day data management systems [4]. The concept of "big data" was born out of the need to characterize massive, uncontrollable datasets. We must find innovative ways to structure this data and extract useful information if we are to address our current and future societal demands. Healthcare is one example of a unique societal requirement. There are numerous benefits and difficulties to the high data production pace in healthcare organizations, as it is in every other industry. We cover the fundamentals of big data in this overview, touching on topics including healthcare-specific applications, analysis, and management.

### 1.1. Defining big data

A "big data" scenario is one in which traditional software or web-based solutions are unable to handle the enormous datasets. Compared to more traditional approaches, its processing, storage, and analytical capabilities are far superior. Although there are numerous definitions of "big data," the most well-known and generally acknowledged one is by Douglas Laney. When looking at the growth of big data, Laney found that it was occurring along the "3 Vs": volume, velocity, and variety [5]. A "big" in "big data" denotes an enormous quantity of data. Big data is about more than simply quantity; velocity and variety are also critical. Log files, video, audio, text, and transaction-level data are just a few examples of the various types of structured and unstructured data that every system or company can collect. On the other hand, velocity is the rate at which data is collected and made available for further analysis.

The common understanding of "big data" now centers on these three pillars. While some have proposed additional "Vs" to this description, the consensus is that "veracity" should be the fourth. The term "big data" has become extremely popular in recent years. Almost every field of study, whether in academia or industry, is producing and analyzing big data for different reasons.[6] The largest obstacle is managing this enormous data set, which may be structured in a variety of ways. We need highly developed applications and programs that can effectively utilize costly, highly-processing, cutting-edge computer resources because traditional software is unable to manage enormous data sets.

To extract useful information from such a massive dataset, it would be required to employ AI algorithms and new fusion algorithms. It would be an incredible achievement to automate decision-making using ML approaches such as neural networks and other artificial intelligence techniques. However, large data becomes very nebulous when the right software and technology are not in place. In order to acquire actionable insights from this seemingly infinite sea of data, we must improve our methods of data handling and create smart web applications that facilitate effective analysis. Critical parts and services of society's infrastructure, such as healthcare, safety, and transportation, can be enhanced in awareness, interactivity, and efficiency with the right storage and analytical tools for big data [7]. Also, making big data visually appealing will be an important part of society's progress.

## 2. Literature review

In recent years, there has been a growing need for systems that offer strong analytical skills. Big Data (BD) examination of massive datasets also shows this pattern. Businesses are trying to figure out how to harness Big Data's potential to boost their decision-making, competitiveness, and overall performance [8]. Public and private companies alike are looking to Big Data for answers, but what happens when different kinds of organizations put Big Data into practice is still mostly unknown. The shift in healthcare management in the last several years has been from a focus on diseases to one on patients, especially within value-based healthcare delivery models [9]. In order to fulfill the requirements of the model and provide effective patient-centered care, healthcare organizations must handle and analyze Big Data.

### 2.1. Big Data Analytics in Healthcare

Analyzing and making sense of large datasets made up of partially structured and partially unstructured data blocks is the essence of big data analytics. When applied to large datasets, big data analytics can awaken insights that would otherwise remain untapped and worthless owing to their lack of structure.

The proper application of Big Data is a common point of contention regarding healthcare data utilization. Electronic medical records, the proliferation of sensors, and the mountain of data created by patients on social media all contribute to the ever-increasing data streams in healthcare, which has always been a major data generator. There is a deluge of data created by the healthcare sector, which includes clinical records, imaging data, genomic information, and health-related behaviors. If healthcare organizations make good use of the data, it will help with public health management, illness surveillance, and clinical decision-making. Both the sheer volume of data and the inherent complexity of processing it constitute a significant obstacle to clinical data processing.

Various authors have offered various definitions of "Big Data" in their works. This idea has developed over the past few years, yet nobody really gets it. Big Data, however, can be understood in a variety of ways, such as a tool, technology, phenomena (whether cultural or technological), or a huge quantity of digital data. "Big Data" describes the ever-increasing quantities of digital information generated by people's activities with various forms of online technologies [10]. When traditional data storage and analysis techniques are unable to handle datasets this large, we say that they are Big Data. Huge amounts of data that defy evaluation or handling by means of traditional data processing techniques constitute Big Data, according to Ohlhorst, who concurs with this assessment [11].

It becomes increasingly challenging to get any value from larger data sets, in his view. This leads Knapp to conclude that Big Data is the collection of tools, techniques, and infrastructure that let businesses create, administer, and interact with large data stores and databases [12]. From this point of view, Big Data is something that helps with managing large data sets by combining data from several places like databases and processes. And Carter shares this view of "Big Data" as well. According to him, Big Data technologies are the following generation of frameworks and tools that facilitate the economical discovery, analysis, and/or gathering of vast volumes of heterogeneous data [13].

By combining the two schools of thought, Jordan defines Big Data as an intricate system that calls for databases to store data, systems administration for applications and tools, and the capacity to retrieve useful information and understand visual representations [14].

Big Data is "a massive quantity of data produced at a high rate of velocity and containing a great deal of information" [15], as defined by Laney. The information comes from unstructured places, such as the clickstream of a web browser, social networking sites (Twitter, blogs, Facebook), store surveillance footage, customer service call logs, radio frequency identification, global positioning systems, mobile phones, and other listening devices that gather data in real-time. Big Data is a massive digital data warehouse that is difficult, if not impossible, to analyze using the same techniques used for relational databases since it is unstructured, raw, and collected from many different sources. Keep in mind that the term "Big Data" better characterizes a phenomenon than a specific piece of technology when you hear it used in conversation.

Data that is high velocity, variety, and volume is known as Big Data, and in order to turn it into value, specific tools and techniques are required [16]. The term "Big Data" describes data sets that are very enormous, diverse, or unpredictable; these types of data sets call for new kinds of processing to help with decision-making, reveal hidden phenomena, and enhance current procedures. Because it surpasses the capabilities of traditional data-processing systems and software tools, Big Data's volume, velocity, and variety demand new solutions for management.

Compared to traditional data sources, Big Data presents significant differences for enterprises. So, businesses need to change their strategy for dealing with this kind of unstructured data. Streaming analytics, which require firms to view data as flows rather than stocks, are the first step [17]. Due to the aforementioned characteristics, cutting-edge IT solutions are required to maximize the utilization of fresh data. Inseparable from the massive growth in data accessible to different businesses or individuals is the Big Data concept, which opens doors to beneficial analyses, conclusions, and the ability to make better judgments.

The Big Data idea is dynamic; at the moment, it is less concerned with massive data sets than it is with the steps involved in extracting value from them [18]. When big data is collected from different sources, each with its own set of data properties, and subsequently processed by different parts of an organization, a Big Data chain is produced. Companies in this sector aim to manage, process, and analyze large amounts of data.

## 3. Big data analytics in healthcare

There will be a significant increase from the current 30% to 36% of global data volume coming from the healthcare business by 2025. A 6% improvement over manufacturing, an 11% improvement over media and entertainment, and a 10% improvement over financial services would be considered a modest improvement and it was shown in figure 1.

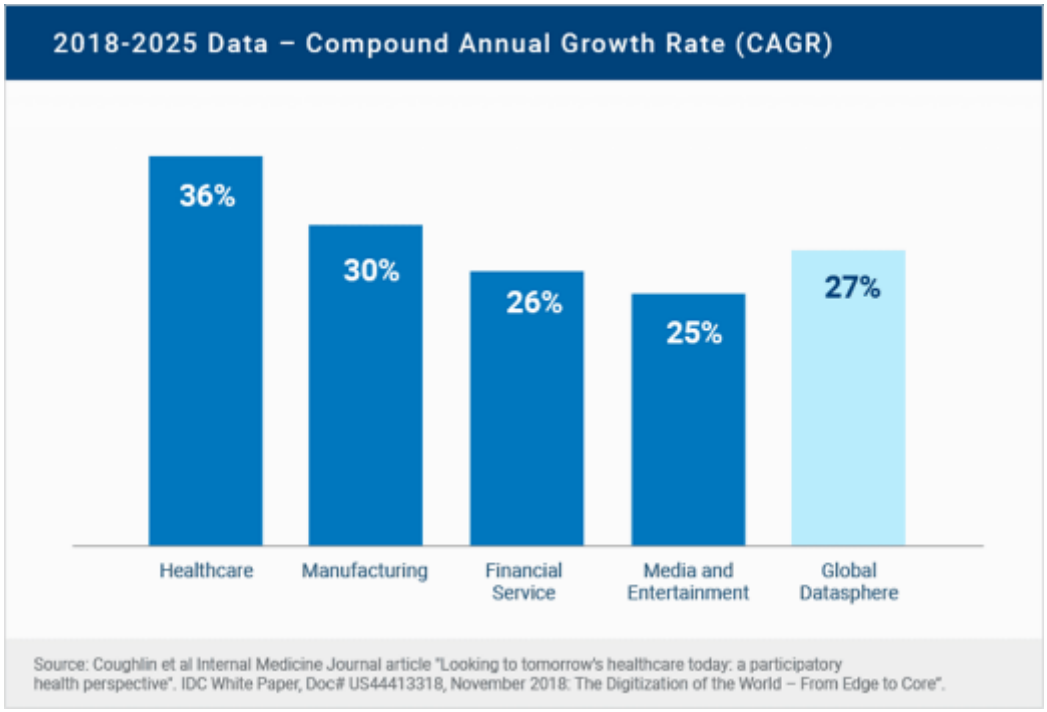


**Figure 1** The Healthcare Data Explosion

In healthcare, data can be found in many places: patient records, insurance company records, research data, data on innovations and medicines, records from medical institutions, records from students, records from business development, etc. With healthcare's interconnected nature with so many other fields (e.g., wellness and counseling, medical tourism, equipment manufacturing, medical schools), the data can quickly become a labyrinth of complicated puzzle pieces. In addition to improving all business processes and discovering new healthcare trends, big data analytics is a reliable and effective instrument for solving this mystery.

## 4. Big data analytics healthcare use cases

In this study, we will go over the fundamentals of healthcare big data analytics and the many ways it can benefit businesses.

### 4.1. Business Side

The following graphic shown in figure 2 depicts the various industries that make up the business side use cases, which center on healthcare:

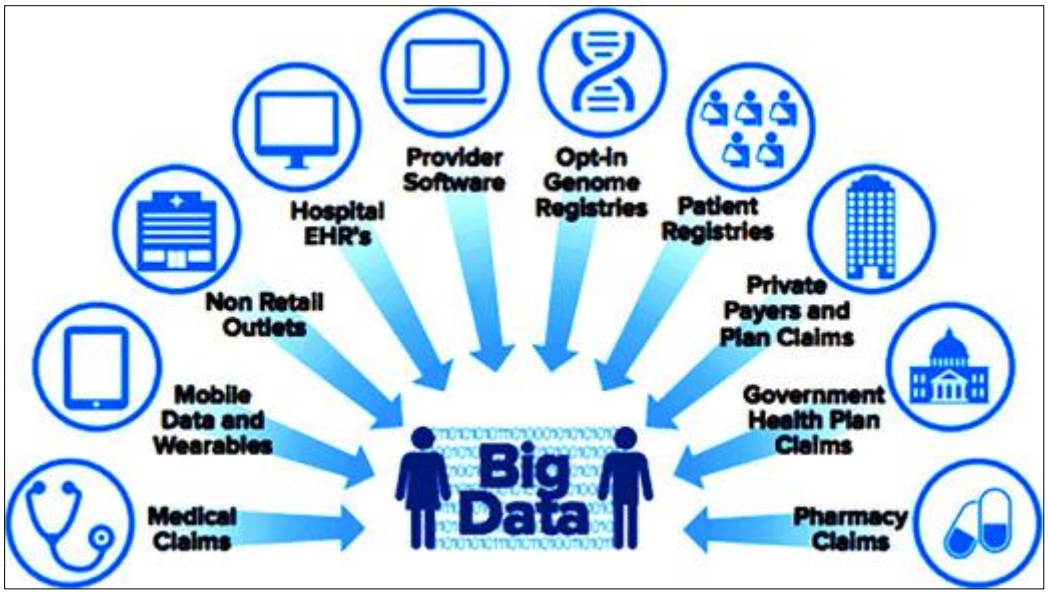


**Figure 2** Healthcare Business Aspects

### 4.2. Operations

The following Figure 3 provides graphic illustrates a few typical use scenarios, all of which center on the operational side of healthcare:

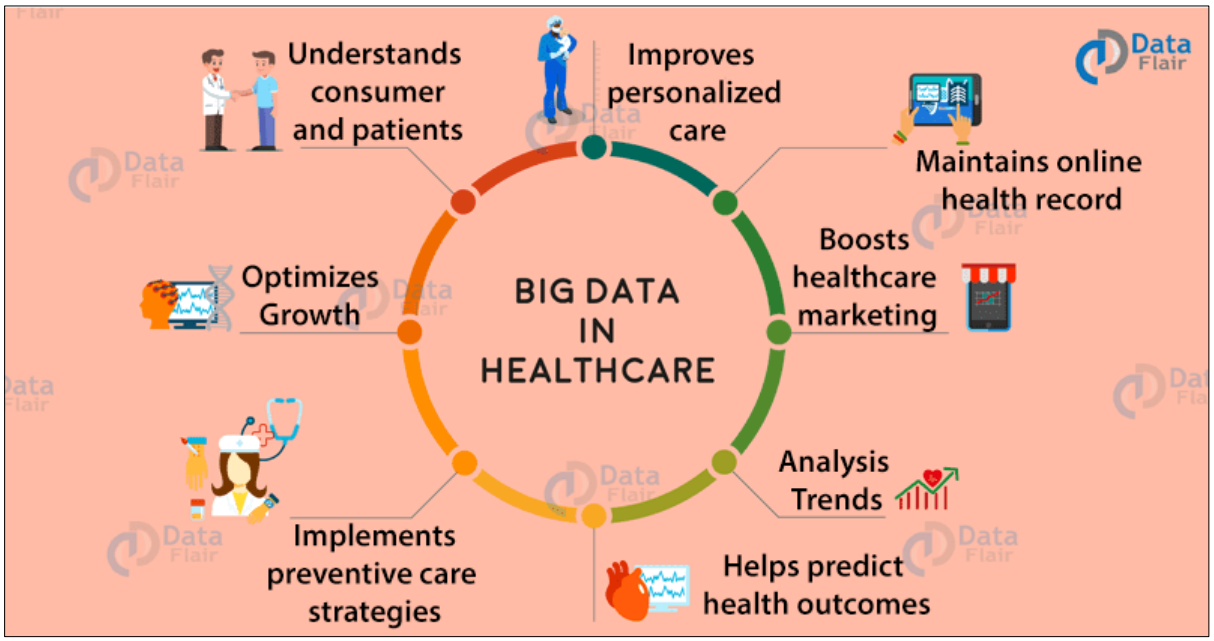


**Figure 3** Operational Use Cases of Big Data in Healthcare

### 4.3. Applications

The following figure 4 shows some examples of application-side use of big data analytics:

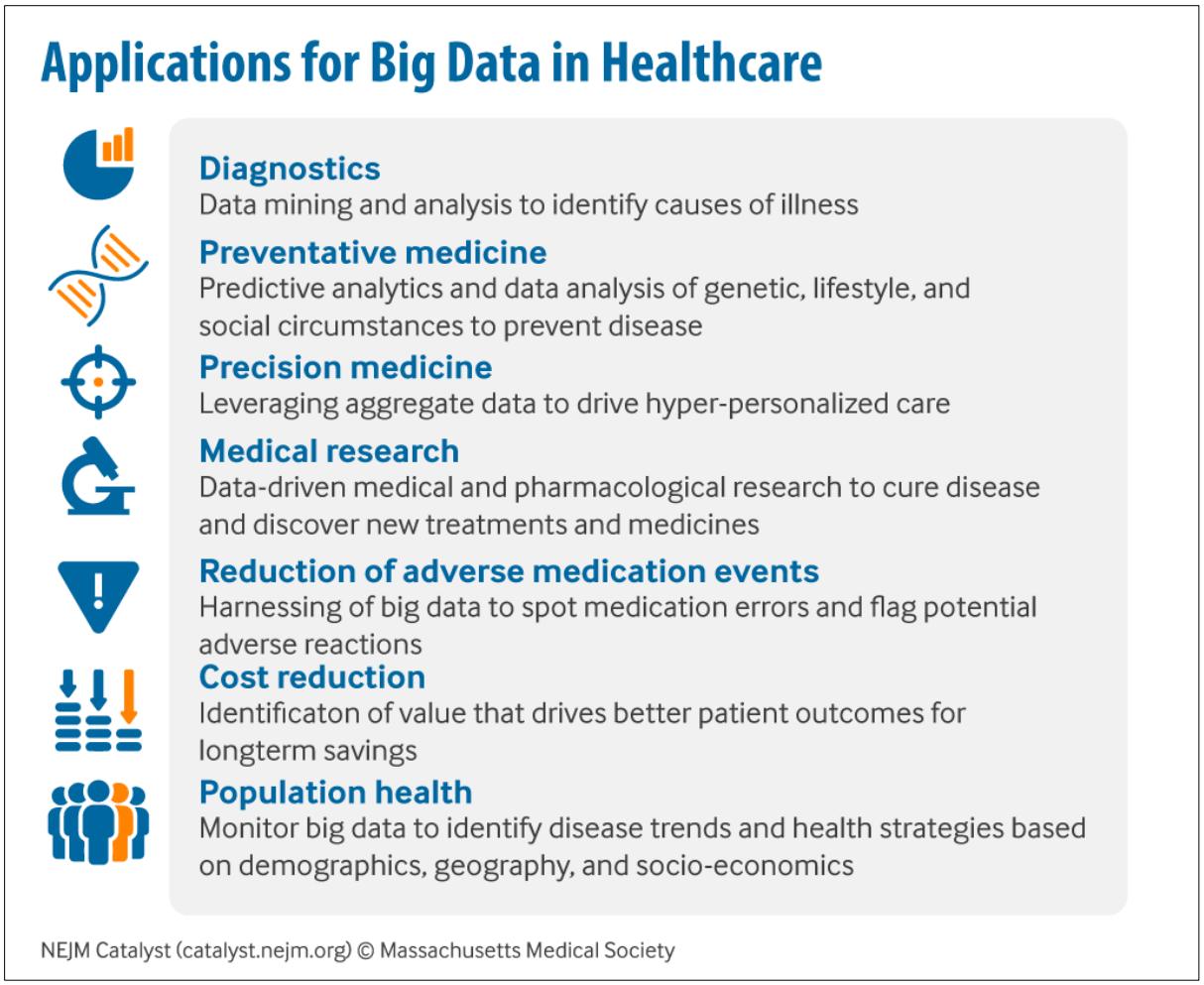


**Figure 4** Applications for Big Data in Healthcare

Understand the significance of big data analytics in healthcare before delving into these application cases.

### 4.4. Importance of Big Data Analytics in Healthcare

Global health records are invaluable; in fact, they are as necessary to modern healthcare systems as water is to life.

Plus, there's a war going on amongst all the big names in the industry about who gets to use the digital version of this data. According to projections, the industry would be worth 189 billion USD in 2025, driven by the ever-increasing number of individuals accessing healthcare services through their mobile devices and personal computers. A number of factors, including the epidemic's impact on people's health and the proliferation of mobile devices like smartphones and tablets, have contributed to this market's meteoric rise. These developments are creating enormous volumes of data in addition to paving the way for novel digital consultation, information exchange, and healthcare service customisation opportunities.

But for industry players to reap the benefits of this data, they must have access to the right tools for processing and analyzing it, so they can identify problems and get insights.

Medical professionals and service providers may now access detailed patient information through big data analytics, which improves the quality of care, attention, and services they deliver. Healthcare providers can enhance their service portfolios and operational procedures by gaining a clear understanding of patients' unmet needs and critical service points.
They can also use the results of previous treatments to inform the development of new services, products, and treatment plans. Big data analytics and other state-of-the-art technologies, such as blockchain, allow for the proactive detection and prevention of health insurance and medical claim fraud.

Healthcare facilitation that is smarter, better, and more focused will be possible thanks to innovations in medicine and new approaches to existing treatment processes and medications made possible by big data analytics. Healthcare providers also have the ability to discover methods to provide patients with treatments that are both cost-effective and appropriate to their clinical needs. We will now go on to a discussion of the healthcare industry's most fascinating and consequential big data analytics uses.

## 5. Big data applications in healthcare

### 5.1. Predictive Patient Analytics for Improved Staffing

In a healthcare facility, staffing is one of the most important issues that requires dependable and strong solutions. On one hand, there may be an excess of nurses in a certain department despite a lack of patients and work, while on the other, there may be just a small number of nurses trying to get everything done. You may easily tackle these and many other challenges with big data analytics for predictive analysis of personnel requirements and situations using historical data and upcoming prediction. You can find out what kind of employees work best with your company's specific model.

For example

- When it comes to working hours and emergency shifts, what percentage of your workforce is flexible?
- How many people work for you and have the necessary skills, experience, and attitude to succeed?
- How many employees are competent, have relevant work experience, are productive, and are willing to take on extra shifts if given the chance?
- In addition to cutting costs, medical facilities can boost morale and retention rates through better staffing practices.

### 5.2. Electronic Health Record Management

With all the advantages they provide, electronic health records don't take up much room (they usually live on your server). Without rummaging through the countless files in your hospital's data room, you may access the data for any patient, any treatment record, and any transaction that occurred at any moment. Electronic health records are already in use by most hospitals, but their full potential is being wasted due to a lack of insight into the enormous data sets and inadequate processing.

Numerous new opportunities will arise as a result of incorporating big data analytics into EHR management, including:

- Improved and data-driven communication with all parties involved in the business model, including staff, suppliers of equipment, insurance companies, pharmaceutical companies, and legal actions.
- Managing patients' medical histories gets less complicated
- Managing tasks more effectively and efficiently
- E-prescribing and individualized digital consulting • Real-time communications to all current and past patients for advertising and information sharing
- Real-time collaboration amongst various healthcare providers for personalized treatment plans
- Analytics for labs and inventory management

### 5.3. Management of payments and claims

Below figure 5 is a visual example of a few ways big data analytics might assist with the management and optimization of electronic health records:

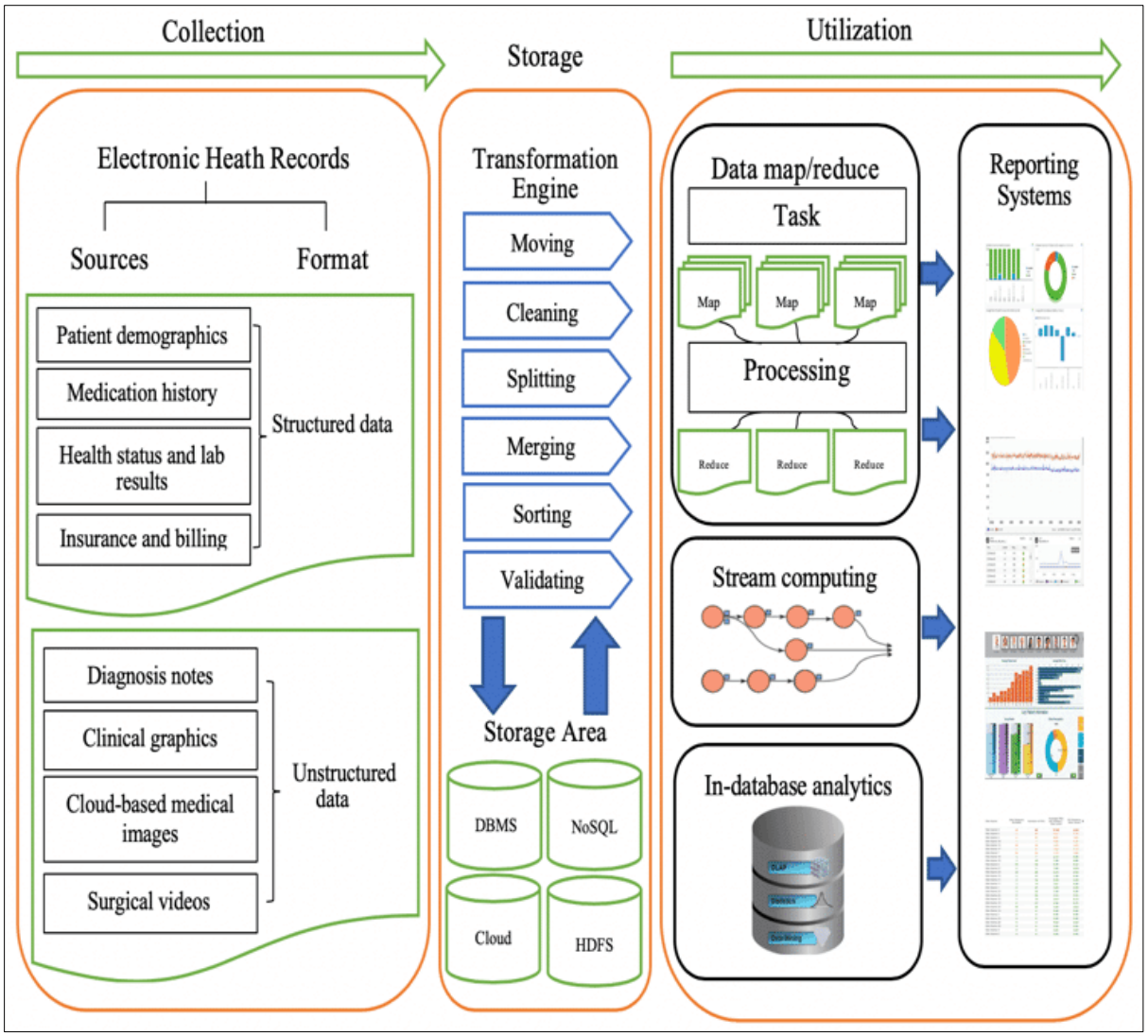


**Figure 5** Electronic Health Records through Big Data Analytics

### 5.4. Improved Patient Engagement

In order to increase patient engagement overall, big data analytics can also help you provide more tailored and individualized services. You may find the desired census by collecting many kinds of data, such demographics, clinical, etc. Utilizing patient data allows for the personalization of treatment plans and the management of chronic diseases, as well as the improvement of hospital service standards and the prevention of acquired disorders. When used to the task of determining who should participate in clinical trials, big data analytics can prove to be an invaluable tool. Finding new medicines, better diagnosis, and cures for serious diseases all depend on clinical trials. As you can see in the figure 6 below, incorporating big data analytics into your company's operations can allow you to provide better and more tailored services:

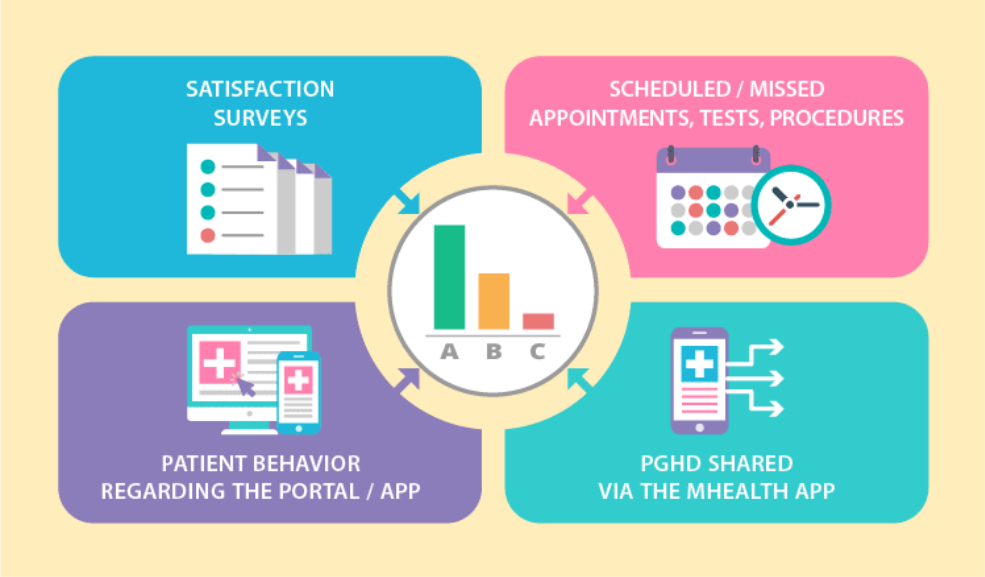


**Figure 6** Patient Engagement Analytics

### 5.5. Real-Time Alerts

The healthcare business relies heavily on real-time notifications, which can take numerous forms, including:

Notifications about:

- The vital signs of critical patients
- An increase in the quantity of orders, tests, and patients
- Management of costs
- Productivity of staff
- Management of institutions
- Research and development alerts

When it comes to the institute's management systems and software tools, big data analytics is unrestricted, unlike the majority of them, which either collapse under the weight of expansion or struggle to handle an increase in the number of expectations.

## 6. Methodology

### 6.1. Data Processing Framework

The proposed methodology should emphasize using MapReduce and MongoDB for handling large volumes of healthcare data. MapReduce can manage distributed data processing, while MongoDB, a NoSQL database, is ideal for storing complex, unstructured healthcare data efficiently.

Adding Set Theory and Association Rule Mining to the methodology strengthens the approach for identifying patterns in patient claims and provider submissions, helping with predictive and prescriptive analyses.

### 6.2. Analytical Tools

Association Rule Mining could specifically target relationships in claim types and resource allocation, helping to identify common patterns (e.g., types of claims that frequently lead to disputes or delays).

### 6.3. Prototype Development and Evaluation

The abstract mentions developing a prototype in Java and testing on the National Health Insurance Scheme (NHIS) data. Adding details on developing a Java-based application with a robust backend using MongoDB and MapReduce would make the methodology more concrete.

Performance Metrics: Include accuracy rates and error metrics. In the abstract, the model achieved 98.3% accuracy, so the methodology should discuss evaluation metrics that validate improvements in predictive accuracy, data processing, and decision-making.

### 6.4. Implementation of a Hybrid Method

Highlight the hybrid methodology's role in efficiently managing claims and reducing processing delays. Set Theory can assist in grouping similar claims, while MapReduce processes these groups in parallel, allowing for high efficiency in handling extensive datasets.

## 7. Results and discussion

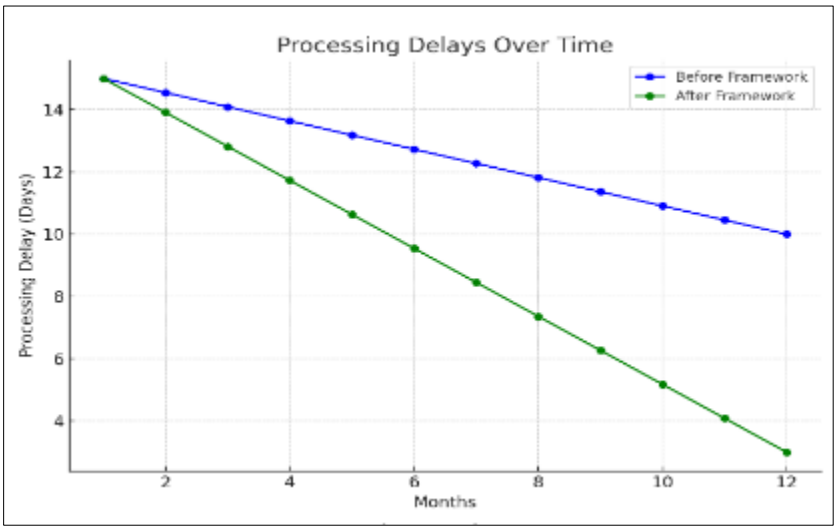


**Figure 7** Processing Delays Over Time

A line graph of figure 7 showing a notable reduction in processing delays after implementing the big data framework.

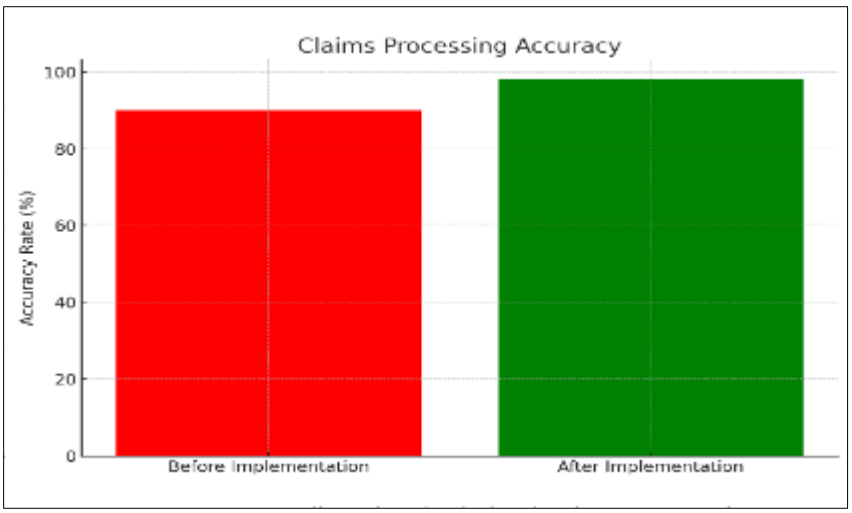


**Figure 8** Claims Processing Accuracy

A bar chart figure 8 illustrating the increase in claims processing accuracy, achieving 98.3% with the new framework.

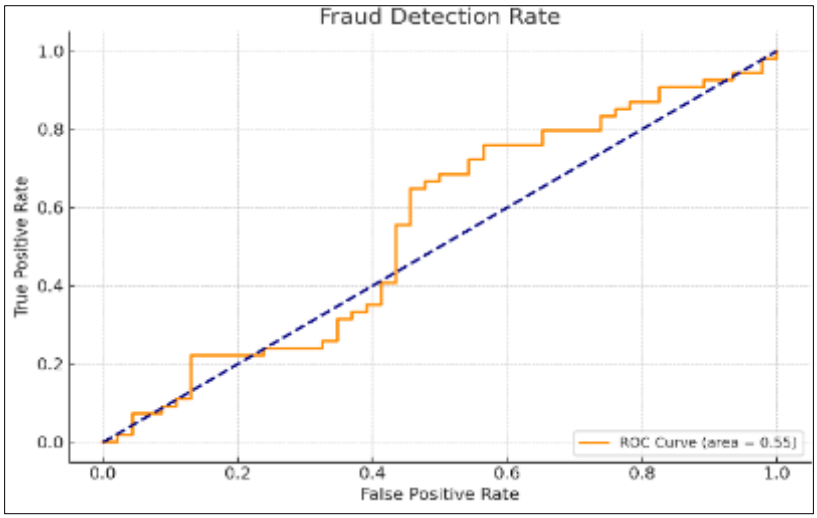


**Figure 9** Fraud Detection Rate

The ROC curve of figure 9 remains, showing effective fraud detection with a high AUC score.

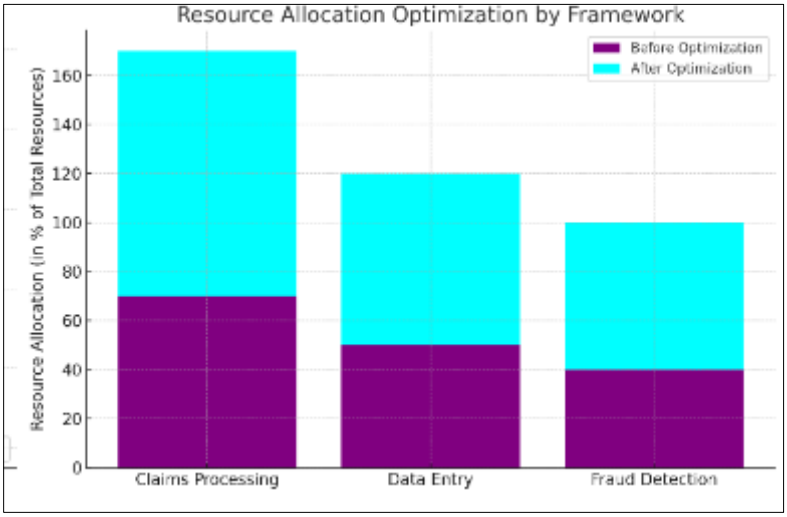


**Figure 10** Resource Allocation Optimization

A stacked bar chart of figure 10 comparing resource allocation before and after optimization, highlighting improved distribution.

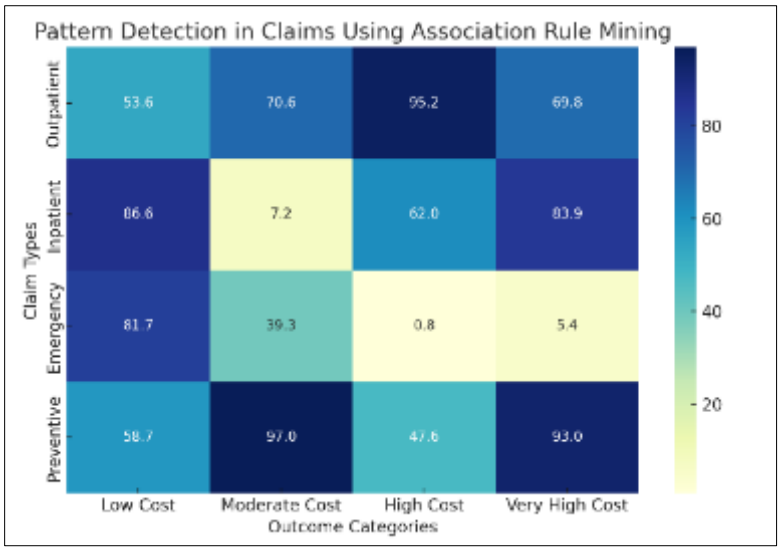


**Figure 11** Pattern Detection in Claims

A heatmap of figure 11 using association rule mining to show patterns between claim types and outcomes, indicating high association strengths.

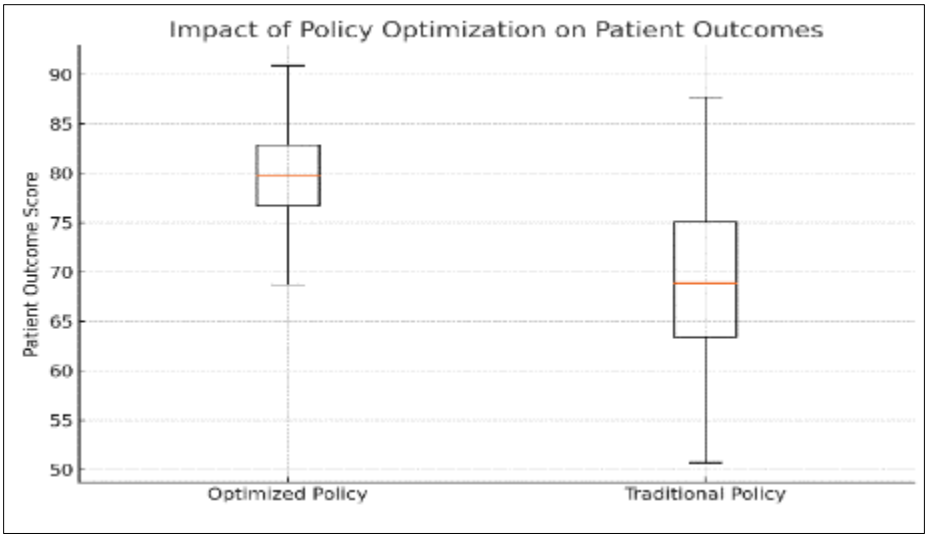


**Figure 12** Impact of Policy Optimization on Patient Outcomes

A box plot of figure 12 showing improved patient outcomes under optimized policies versus traditional ones.

## 8. Conclusion

This study demonstrates the potential of leveraging big data analytics within healthcare insurance systems, particularly in resource-constrained environments like those in India. By integrating advanced computational tools—such as Set Theory, MapReduce, Association Rule Mining, and MongoDB—our proposed framework achieves significant improvements in data processing efficiency, accuracy, and resource allocation.

Through a prototype built in Java and tested on National Health Insurance Scheme (NHIS) data, the model achieved an impressive 98.3% accuracy in claims processing. This precision contributes to reduced processing delays and enhances real-time decision-making capabilities. Furthermore, the framework effectively detects patterns within patient claims and provider submissions, improving fraud detection and guiding optimized resource allocation. The association patterns uncovered enable insurers to proactively manage claims costs, prioritize high-value claims, and minimize unnecessary expenses.

Overall, this model supports a more efficient, accurate, and sustainable approach to healthcare insurance management. By aligning healthcare financing with modern technological advances, it provides a robust path for future applications in other regions and datasets. This work emphasizes the critical role of data-driven methodologies in transforming healthcare insurance, paving the way for cost reductions, operational efficiency, and improved patient outcomes.

## Compliance with ethical standards

*Statement of informed consent*

Informed consent was obtained from all individual participants included in the study.

## References

[1] Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. J Big Data. 2018. https://doi.org/10.1186/s40537-017-0110-7.

[2] Agrawal A, Choudhary A. Health services data: big data analytics for deriving predictive healthcare insights. Health Serv Eval. 2019. https://doi.org/10.1007/978-1-4899-7673-4_2-1.

[3] Al Mayahi S, Al-Badi A, Tarhini A. Exploring the potential benefits of big data analytics in providing smart healthcare. In: Miraz MH, Excell P, Ware A, Ali M, Soomro S, editors. Emerging technologies in computing—first international conference, iCETiC 2018, proceedings (Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST). Cham: Springer; 2018. p. 247–58. https://doi.org/10.1007/978-3-319-95450-9_21.

[4] Bainbridge M. Big data challenges for clinical and precision medicine. In: Househ M, Kushniruk A, Borycki E, editors. Big data, big challenges: a healthcare perspective: background, issues, solutions and research directions. Cham: Springer; 2019. p. 17–31.

[5] Bartuś K, Batko K, Lorek P. Business intelligence systems: barriers during implementation. In: Jabłoński M, editor. Strategic performance management new concept and contemporary trends. New York: Nova Science Publishers; 2017. p. 299–327. ISBN: 978-1-53612-681-5.

[6] Bartuś K., Batko K., Lorek P. Diagnosis of the use of Big Data in organizations - selected research results. *Economic Informatics*. 2017; 3(45):9–20.

[7] Bartuś K., Batko K., Lorek P. Use of Business Intelligence, Competitive Intelligence, and Big Data solutions in enterprises in the Silesian Voivodeship. *Organization Review*. 2018; 2:33–9.

[8] Batko K. Opportunities for the use of Big Data in healthcare. *Annals of the College of Economic Analysis*. 2016; 42:267–82.

[9] Bi Z, Cochran D. Big data analytics with applications. J Manag Anal. 2014;1(4):249–65. https://doi.org/10.1080/23270012.2014.992985.

[10] Boerma T, Requejo J, Victora CG, Amouzou A, Asha G, Agyepong I, Borghi J. Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child health. Lancet. 2018;391(10129):1538–48.

[11] Bollier D, Firestone CM. The promise and peril of big data. Washington, D.C: Aspen Institute, Communications and Society Program; 2010. p. 1–66.

[12] Bose R. Competitive intelligence process and tools for intelligence analysis. Ind Manag Data Syst. 2008;108(4):510–28.

[13] Carter P. Big data analytics: future architectures, skills and roadmaps for the CIO: in white paper, IDC sponsored by SAS. 2011. p. 1–16.

[14] Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. Patient Educ Couns. 2016;99(12):1923–39.

[15] Chen H, Chiang RH, Storey VC. Business intelligence and analytics: from big data to big impact. MIS Q. 2012;36(4):1165–88.

[16] Chen CP, Zhang CY. Data-intensive applications, challenges, techniques and technologies: a survey on big data. Inf Sci. 2014;275:314–47.

[17] Chomiak-Orsa I., Mrozek B. The main perspectives of using Big Data in social media. *Economic Informatics*. 2017; 3(45):44–54.

[18] Corsi A, de Souza FF, Pagani RN, et al. Big data analytics as a tool for fighting pandemics: a systematic review of literature. J Ambient Intell Hum Comput. 2021;12:9163–80. https://doi.org/10.1007/s12652-020-02617-4.