AIM INTERNATIONAL JOURNAL Publisher
Journal of Next-Gen Engineering Systems, 2025; 2(01): 29-47
DoI: 10.70937/jnes.v2i01.51

BIG DATA ANALYTICS IN HEALTHCARE: TOOLS, TECHNIQUES, AND APPLICATIONS - A SYSTEMATIC REVIEW

Mst Shamima Akter [1]
[1]Master in Management Information Systems and MBA (Dual), College of Business, Lamar University, Texas, USA
Corresponding Email: makter2@lamar.edu

Rafiqul Islam [2]
[2]Master in Management Information System, College of Business, Lamar University, Texas, USA
Email: rislam9@lamar.edu

Md Atiqur Rahman Khan [3]
[3]Master in Management Information System, College of Business, Lamar University, Texas, USA
Email: mkhan35@lamar.edu

Shaharima Juthi [4]
[4]Master in Management Information Systems, College of Business, Lamar University, Texas, USA
Email: sjuthi@lamar.edu

Keywords

Big Data Analytics
Healthcare
Predictive Modeling
Precision Medicine
Data Integration

ABSTRACT

Big Data Analytics (BDA) has emerged as a transformative force in healthcare, offering innovative solutions to analyze large and complex datasets for actionable insights. This systematic review, encompassing 142 peer-reviewed studies published between 2010 and 2024, explores the tools, techniques, and applications of BDA in healthcare. The findings reveal the critical role of BDA in enhancing clinical decision-making, optimizing hospital workflows, and advancing medical research. Key applications such as predictive analytics for disease prevention, real-time monitoring through IoT integration, and precision medicine through genomic analysis are highlighted. Tools like Hadoop, Spark, and TensorFlow, combined with advanced techniques such as machine learning and natural language processing, have been pivotal in transforming healthcare data into actionable knowledge. However, the review also identifies significant challenges, including data integration issues, algorithmic bias, and ethical concerns related to patient privacy and data security. By addressing these barriers, BDA has the potential to revolutionize healthcare delivery, providing more personalized, efficient, and equitable care. This study provides a comprehensive understanding of the current state of BDA in healthcare, its limitations, and its promising future applications, offering valuable insights for researchers, policymakers, and healthcare practitioners.

1 INTRODUCTION

Big Data Analytics (BDA) has emerged as a transformative force in healthcare, offering unprecedented opportunities to analyze large and complex datasets for actionable insights (Abera et al., 2014). As healthcare systems generate massive amounts of data from electronic health records (EHRs), wearable devices, genomic studies, and other sources, the adoption of BDA has become crucial to improving patient outcomes and operational efficiencies. Studies emphasize that BDA allows for the integration and analysis of diverse data types, enabling a more holistic understanding of patient health and disease

Received: 28th November, 2024 **Accepted**: 12th January, 2025 **Published**: 14th January, 2025

management (Abouelmehdi et al., 2018; Bahga & Madisetti, 2013; Belle et al., 2015). With healthcare costs rising globally, BDA serves as a vital tool to optimize resource allocation and minimize inefficiencies (Jee & Kim, 2013). The tools used in BDA play a critical role in processing vast amounts of healthcare data, from traditional database management systems to advanced machine learning algorithms. Apache Hadoop, Spark, and NoSQL databases are widely used for data storage and processing, while machine learning techniques facilitate predictive analytics and pattern recognition (Gamache et al., 2018). These tools enable healthcare organizations to process both structured and unstructured data, such as patient records, imaging data, and sensor outputs, at an unprecedented scale and speed (Young et al., 2014). Moreover, the integration of artificial intelligence (AI) technologies with BDA tools has amplified the capacity to deliver personalized healthcare interventions (Belle et al., 2015; Willems et al., 2019).

Techniques employed in BDA for healthcare are diverse, spanning statistical modeling, machine learning, natural language processing, and data visualization (Jee & Kim, 2013). Techniques like clustering, classification, and regression analysis are frequently applied to predict patient risk factors and optimize treatment (Dede et al., 2016). For example, clustering algorithms are used to group patients with similar health profiles for targeted interventions, while natural language processing helps extract meaningful insights from clinical notes and medical literature (Huang et al., 2016). Data visualization techniques further support clinicians and decision-makers by presenting complex data in intuitive formats, facilitating faster and more informed decision-making (Swan, 2013).

Applications of BDA in healthcare are vast, ranging from population health management to precision medicine. Predictive modeling techniques have been employed to forecast patient readmission rates, identify high-risk patients, and improve hospital workflow efficiency (Jacofsky, 2017). In the realm of precision medicine, BDA is instrumental in analyzing genetic data to develop tailored treatments for individual patients, thereby enhancing treatment efficacy and reducing adverse effects (Porche, 2014). Furthermore, real-time analytics in BDA has facilitated the monitoring of patient vitals and chronic disease management through wearable devices and remote patient monitoring systems (Huang et al., 2016). Despite the advancements, the implementation of BDA in healthcare is not without challenges. Issues such as data heterogeneity, scalability, and data security significantly impact the effectiveness of BDA solutions (Jee & Kim, 2013; Willems et al., 2019). Data heterogeneity, arising from the integration of structured and unstructured data from diverse sources, often complicates the development of interoperable systems (Porche, 2014). Scalability concerns emerge as healthcare data grows exponentially, demanding

Figure 1: Bigdata Application in Healthcare

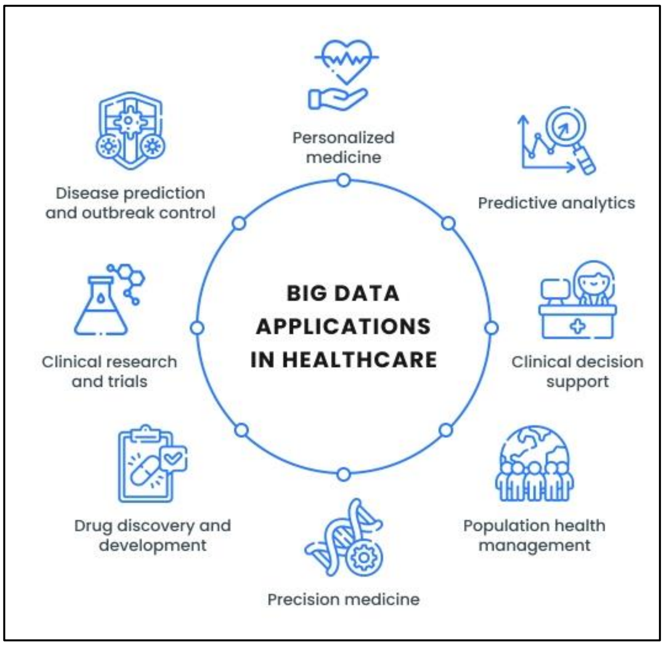
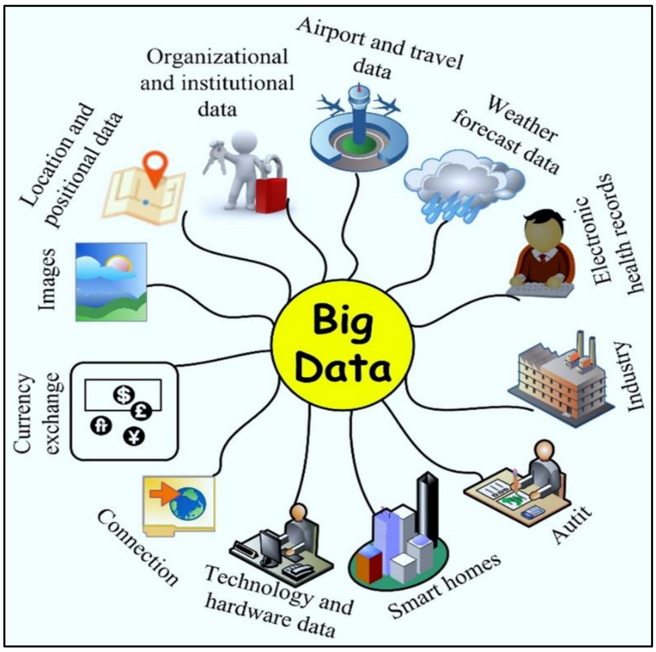


Figure 2: Applications and Data Sources in Big Data Analytics for Healthcare



Source: Khan et al. (2022).

Journal of Next-Gen Engineering Systems
DoI: 10.70937/jnes.v2i01.51

advanced computing infrastructures to manage and analyze the data efficiently (Shamli & Sathiyabhama, 2016). Moreover, maintaining patient privacy and adhering to stringent data protection regulations remain critical concerns for healthcare providers and policymakers (Willems et al., 2019). Another significant focus in the application of BDA in healthcare is the ethical implications of data use. Studies highlight concerns regarding informed consent, potential biases in data interpretation, and the equitable distribution of data-driven healthcare benefits (Dede et al., 2016). Ethical frameworks are necessary to guide the responsible use of BDA while ensuring patient trust and regulatory compliance. Additionally, the development of frameworks to assess the quality and reliability of big data sources is essential to maintain the integrity of healthcare analytics (Gamache et al., 2018). As healthcare systems increasingly rely on BDA for decision-making, addressing these ethical concerns is crucial for sustainable implementation (Willems et al., 2019). The primary objective of this systematic review is to explore the tools, techniques, and applications of Big Data Analytics (BDA) in the healthcare sector, emphasizing their role in enhancing patient care, optimizing operational efficiencies, and addressing critical challenges. Specifically, this review aims to identify the most commonly utilized tools and methodologies for managing large healthcare datasets, analyze their effectiveness in various applications such as predictive modeling and precision medicine, and evaluate their impact on healthcare delivery. By synthesizing evidence from recent studies, this review also seeks to uncover the challenges and limitations faced in implementing BDA, including issues related to data security, ethical considerations, and system interoperability. Through this objective-driven approach, the study aspires to provide a comprehensive understanding of how BDA can be effectively harnessed to improve healthcare outcomes and inform decision-making for healthcare providers, researchers, and policymakers.

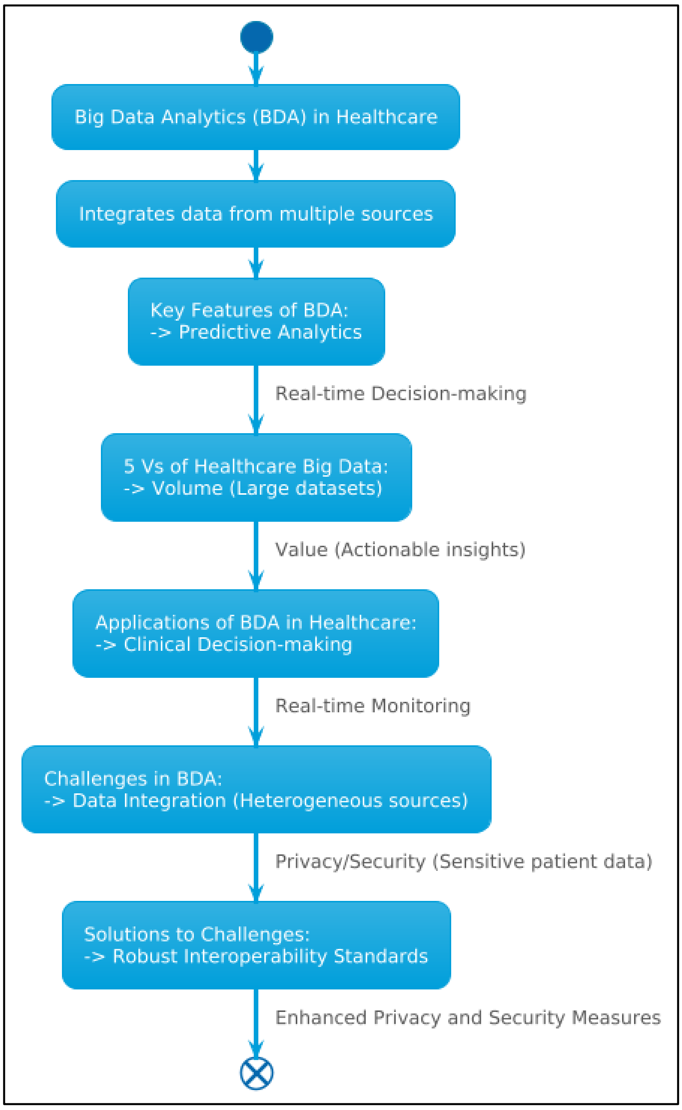
2 LITERATURE REVIEW

The literature review explores the evolving landscape of Big Data Analytics (BDA) in healthcare, synthesizing key findings from previous studies to provide a comprehensive understanding of the tools, techniques, and applications shaping this field. This section is structured to highlight the pivotal contributions of BDA in various healthcare domains, examining methodologies, implementation challenges, and ethical considerations. The review also underscores the transformative potential of BDA in improving patient care, operational efficiency, and clinical decision-making. By organizing the literature into thematic categories, this section aims to provide a clear and systematic overview of the state of research in BDA for healthcare.

2.1 Overview of Big Data Analytics in Healthcare

Big Data Analytics (BDA) in healthcare is defined as the process of examining large, diverse datasets to uncover hidden patterns, correlations, and insights that can significantly improve patient care and operational efficiency (Bahga & Madisetti, 2013). BDA integrates data from multiple sources, such as electronic health records (EHRs), wearable devices, and genomic studies, into actionable insights (Bainbridge, 2019). The scope of BDA encompasses predictive analytics, disease prevention, personalized medicine, and real-time decision-making (Berger & Doban, 2014). Additionally, it facilitates improved resource management and reduces healthcare costs by optimizing workflows (Berndt et al., 2001). Studies emphasize the potential of BDA to address complex healthcare challenges by leveraging advanced analytical tools and techniques, offering a more comprehensive understanding of health systems and patient outcomes (Sharp, 2011; Ward et al., 2014). Moreover, healthcare big data exhibits distinct characteristics, often summarized as the five Vs: Volume, Velocity, Variety, Veracity, and Value. Volume refers to the vast quantities of data generated daily, including structured data from EHRs and unstructured data from imaging and clinical notes (Abera et al., 2014). Velocity addresses the speed at which healthcare data is generated and processed, particularly in real-time applications like wearable devices and remote patient monitoring (Son et al., 2010). Variety underscores the diversity of healthcare data, encompassing structured, semi-structured, and unstructured formats such as text, images, and sensor data (Ward et al., 2014). Veracity pertains to data accuracy and reliability, a critical factor in clinical decision-making (Buntin et al., 2011). Value highlights the actionable insights derived from analyzing this data, such as predicting disease progression or optimizing hospital operations (Corsi et al., 2020).

Figure 3: Big Data Analytics (BDA) in Healthcare



The integration of big data analytics into healthcare is transforming clinical decision-making and population health management. For instance, predictive models based on BDA have been used to anticipate patient readmissions and assess risk factors for chronic diseases (Kumar & Singh, 2019). Additionally, BDA enables precision medicine by analyzing genomic and proteomic data, tailoring treatments to individual patient profiles (Belle et al., 2015). Studies also highlight the role of real-time analytics in enhancing patient monitoring through wearable technologies and IoT devices, allowing for timely interventions (Belle et al., 2015; Kumar & Singh, 2019). These applications underscore the potential of BDA to revolutionize healthcare by improving diagnosis accuracy and treatment efficacy (Abera et al., 2014; Mohammed et al., 2014). Despite its transformative potential, BDA in healthcare faces challenges such as data integration, scalability, and privacy concerns. Data integration is complicated by the heterogeneity of healthcare data sources, which requires robust interoperability standards to ensure seamless data exchange (Lamarche-Vadel et al., 2014). Scalability remains a pressing issue, with the exponential growth of healthcare data necessitating advanced storage and processing infrastructures (Mohr et al., 2013). Privacy and security concerns are paramount, given the sensitive nature of patient data and the increasing incidence of data breaches in healthcare systems (Bressan et al., 2012). Addressing these challenges is essential for maximizing the benefits of BDA and ensuring its sustainable adoption in healthcare (Corsi et al., 2020).

2.2 Hadoop, Spark, and NoSQL Databases

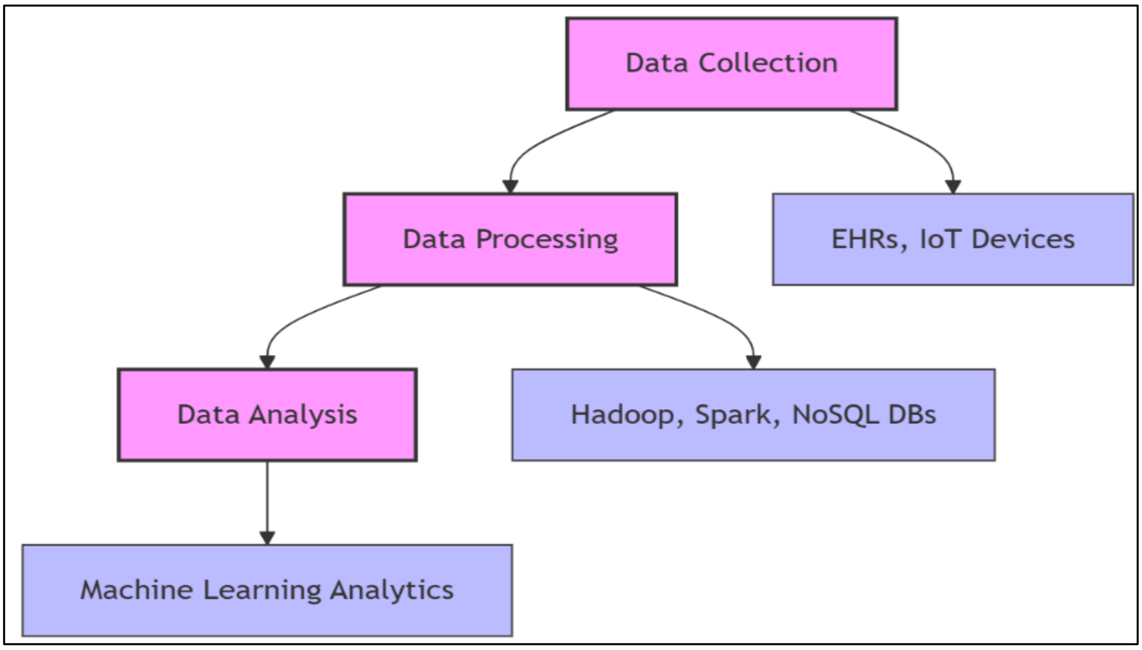
Hadoop, a widely recognized framework for big data processing, has become foundational in healthcare data analytics due to its scalability and ability to handle vast amounts of structured and unstructured data (Jach et al., 2015). It employs a distributed storage system, Hadoop Distributed File System (HDFS), which ensures data redundancy and accessibility across multiple nodes (Mohammed et al., 2014). Studies show that Hadoop's MapReduce programming model is particularly effective in processing large-scale healthcare datasets, such as electronic health records (EHRs) and genomic data, enabling rapid and parallel computation (Rangarajan et al., 2018). This capability has been leveraged to analyze patient trends, forecast healthcare demands, and facilitate disease outbreak prediction (Kumar & Singh, 2019). Apache Spark, an advanced big data processing engine, offers significant improvements over Hadoop in terms of speed and in-memory processing. Spark's ability to perform real-time analytics is instrumental in healthcare applications, such as wearable device data monitoring and patient vital tracking (Saraladevi et al., 2015). Unlike Hadoop, which writes data to disk between computational steps, Spark processes data in memory, resulting in faster computation for iterative machine learning algorithms and large-scale data transformations (Patel & Sharma, 2014). Its real-time streaming capabilities are essential for managing continuous data streams from IoT-enabled healthcare devices and telemedicine platforms, allowing for timely interventions and decision-making (O'Driscoll et al., 2013; Patel & Sharma, 2014).

NoSQL databases, including MongoDB, Cassandra, and Couchbase, have gained prominence in healthcare due to their ability to handle diverse and dynamic data types (Huang et al., 2012). These databases are designed

DoI: 10.70937/jnes.v2i01.51

for scalability and flexibility, making them ideal for managing unstructured data such as clinical notes, imaging files, and sensor outputs (Ishwarappa & Anuradha, 2015; O'Driscoll et al., 2013). MongoDB's document-based architecture is particularly suitable for integrating heterogeneous healthcare data sources, while Cassandra excels in providing high availability and fault tolerance for distributed systems (Patel & Sharma, 2014). Couchbase's hybrid capabilities, combining key-value and document database features, have been employed in real-time applications, such as patient monitoring and personalized treatment recommendations (Gurtowski et al., 2012). The adoption of these tools in healthcare analytics has led to significant advancements in managing and processing big data. Hadoop's distributed architecture provides robust data storage and retrieval, Spark offers real-time and iterative data processing, and NoSQL databases address the challenge of data heterogeneity and unstructured formats (Schatz, 2009). Together, these technologies form the backbone of modern healthcare analytics, enabling applications such as predictive modeling, population health management, and resource optimization. Their integration with machine learning and artificial intelligence further amplifies their potential, transforming raw data into actionable insights that enhance clinical and operational outcomes (Ishwarappa & Anuradha, 2015; Ye et al., 2009).

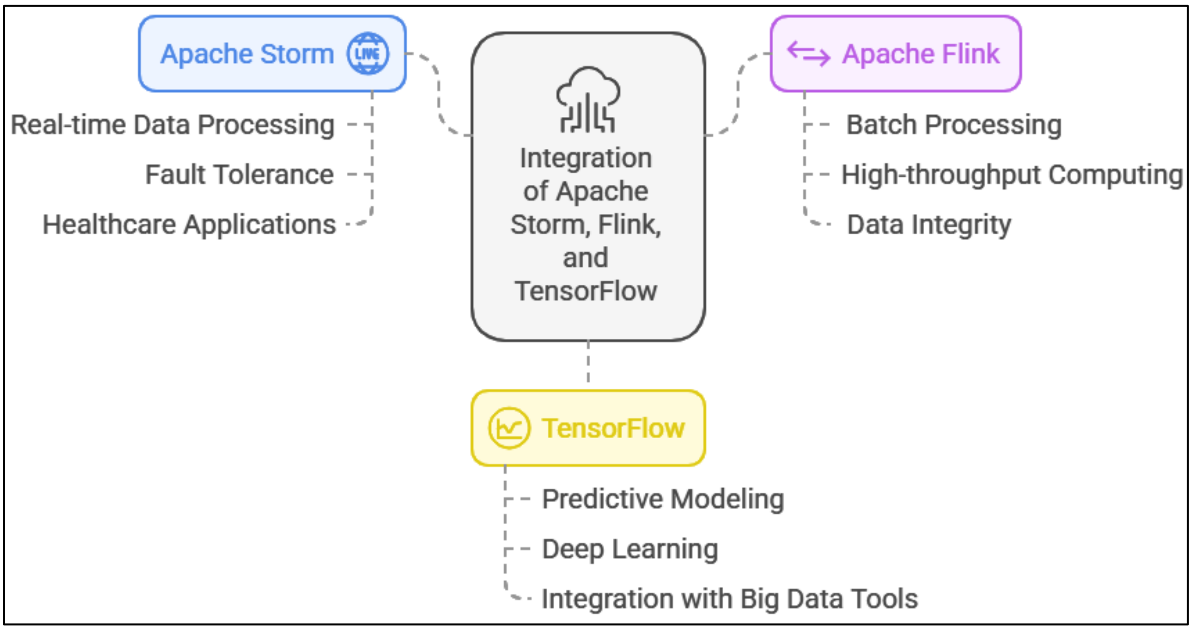
Figure 4: Healthcare Data Processing Pipeline



2.3 Apache Storm, Flink, and TensorFlow

Apache Storm is a real-time stream processing framework that has gained traction in healthcare analytics for its ability to process high-velocity data streams. Its distributed architecture and low-latency capabilities enable real-time analysis of patient data, such as vital signs and wearable device outputs (Devi & Maragatham, 2018). Storm's flexibility makes it ideal for integrating diverse data streams from electronic health records (EHRs), IoT devices, and clinical systems (Huang et al., 2012). Studies demonstrate its application in monitoring chronic diseases, where timely insights are crucial for preventive care and early intervention (Huang et al., 2012; Ishwarappa & Anuradha, 2015). Additionally, Storm's fault-tolerant design ensures the reliability and consistency of healthcare data processing, even in distributed environments (Bakshi, 2012). Furthermore, Apache Flink, another stream-processing framework, surpasses Storm in handling both real-time and batch processing within a single architecture. Flink's high-throughput and in-memory computation capabilities are particularly beneficial for healthcare analytics involving large-scale data from imaging systems, genomics, and population health studies (Chen et al., 2018). Its event-driven processing model ensures that healthcare systems can adapt to continuous data streams without compromising on analytical precision (Grace et al., 2014; Nandimath et al., 2013). Studies highlight Flink's application in patient monitoring systems, where it provides near-instantaneous insights into

Figure 5: Integration of Apache Storm, Flink, and TensorFlow



patient conditions, enhancing decision-making and clinical workflows (Falcão e Cunha et al., 2015). Moreover, its support for fault-tolerant state management is instrumental in ensuring data integrity in critical healthcare (Uzunkaya et al., 2015).

TensorFlow, a machine learning framework developed by Google, is extensively used in healthcare for building predictive models and performing complex data analytics. Its versatility allows healthcare researchers to design deep learning models for tasks such as medical image analysis, disease prediction, and natural language processing of clinical notes (Falcão e Cunha et al., 2015). TensorFlow's ability to handle large datasets and optimize model training through distributed computing has been pivotal in genomics and radiology studies, where precision is paramount (Bakshi, 2012; Ishwarappa & Anuradha, 2015). Furthermore, TensorFlow's integration with other big data tools facilitates the seamless deployment of machine learning solutions in healthcare systems, driving innovations in diagnostics and personalized medicine (Gurtowski et al., 2012). The combined use of Apache Storm, Flink, and TensorFlow provides a comprehensive approach to healthcare big data analytics, leveraging real-time processing, scalable computation, and advanced machine learning. Storm and Flink address the need for rapid and reliable data stream analysis, while TensorFlow empowers healthcare systems with predictive and prescriptive analytics capabilities (Devi & Maragatham, 2018; O'Driscoll et al., 2013). These tools are instrumental in transforming raw data into actionable insights, enabling applications such as early disease detection, patient-specific treatment planning, and healthcare resource optimization (Huang et al., 2012).

2.4 Tableau, Power BI, and QlikView

Tableau has emerged as a leading data visualization tool in healthcare analytics due to its user-friendly interface and robust capabilities for transforming complex datasets into interactive dashboards. Its drag-and-drop functionality allows healthcare professionals to create real-time visualizations, facilitating faster and more informed decision-making (Senthilkumar et al., 2018). Tableau is widely used in analyzing patient demographics, operational metrics, and disease patterns to enhance care delivery and resource allocation (Ishwarappa & Anuradha, 2015). Studies highlight Tableau's ability to integrate data from various sources, such as electronic health records (EHRs) and IoT devices, enabling the exploration of trends in population health management and hospital performance metrics (Bose, 2008; Jorge & Lopes, 2019). Power BI, developed by Microsoft, is another prominent tool in healthcare analytics, known for its seamless integration with other Microsoft products and its ability to handle large datasets. Power BI supports the creation of visually compelling reports and dashboards, enabling healthcare providers to monitor key performance indicators (KPIs) and patient outcomes (Sahoo et al., 2013). Its cloud-based functionality allows for real-time data sharing and collaboration, making it an effective tool for team-based healthcare decision-making (Ritter

DoI: 10.70937/jnes.v2i01.51

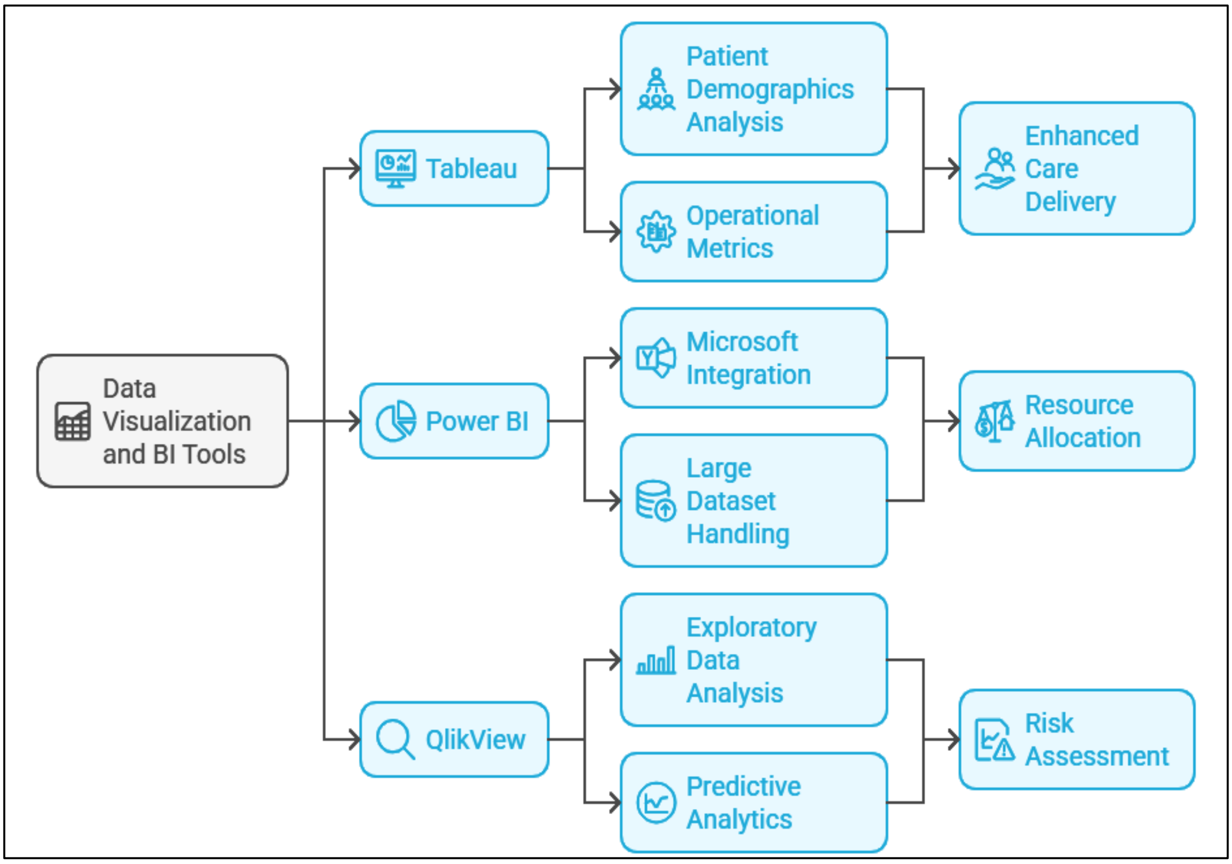
et al., 2011). Research demonstrates Power BI's application in tracking patient readmission rates, analyzing treatment outcomes, and managing hospital resources, contributing to operational efficiency and improved care quality (Glemser et al., 2018).

QlikView, a business intelligence tool known for its associative data indexing technology, has been extensively applied in healthcare for exploratory data analysis. QlikView's ability to handle complex queries and reveal hidden patterns in data is particularly useful in predictive analytics and risk assessment (Luo & Brouwer, 2013). Healthcare organizations have utilized QlikView to identify high-risk patients, monitor chronic disease trends, and optimize clinical workflows (Grosu et al., 2002). Its self-service analytics features empower healthcare professionals to interact with data directly, reducing reliance on IT teams and accelerating decision-making processes (van Iersel et al., 2008). Additionally, QlikView's in-memory computing enhances processing speed, making it a valuable tool for time-sensitive healthcare applications (Gamache et al., 2018). Together, Tableau, Power BI, and QlikView provide healthcare organizations with versatile tools for data visualization and business intelligence. Tableau excels in creating intuitive dashboards, Power BI offers strong integration with enterprise systems, and QlikView facilitates advanced exploratory analytics (Gamache et al., 2018; Karmonik et al., 2017). Studies emphasize their combined role in improving data-driven decision-making, enabling applications such as patient risk stratification, operational optimization, and disease trend analysis (van Iersel et al., 2008). These tools are pivotal in transforming raw healthcare data into actionable insights, enhancing both clinical outcomes and organizational performance (Grosu et al., 2002).

2.5 Integration of AI with BDA Tools

The integration of artificial intelligence (AI) with Big Data Analytics (BDA) tools has significantly enhanced the ability to extract actionable insights from vast and complex healthcare datasets. AI-powered tools such as machine learning algorithms and natural language processing are increasingly being integrated into frameworks like Hadoop and Spark to improve their analytical capabilities (Sahoo et al., 2013). For example, combining AI with Apache Hadoop enables predictive modeling for chronic disease management by analyzing trends and patterns in electronic health

Figure 6: Data Visualization and BI Tools



records (Bose, 2008). Similarly, Spark's in-memory processing coupled with AI techniques accelerates the processing of high-dimensional genomic data for precision medicine (Mohr et al., 2013). These integrations allow healthcare organizations to harness the full potential of BDA tools in diagnosing diseases, optimizing treatments, and managing population health. AI integration has revolutionized real-time healthcare analytics by enhancing the capabilities of tools like Apache Storm and Flink. Storm, when combined with AI algorithms, processes streaming data from wearable devices and IoT sensors to provide real-time health monitoring and early warning systems (Jorge & Lopes, 2019). Flink's event-driven architecture, enriched with AI models, enables the rapid detection of anomalies in patient data, such as irregular heartbeats or respiratory patterns (Glemser et al., 2018). These AI-augmented systems are particularly valuable in critical care settings, where timely insights are essential for saving lives (Luo & Brouwer, 2013). Studies have demonstrated the effectiveness of integrating AI into streaming analytics platforms for improving patient outcomes through faster and more accurate decision-making (Luo & Brouwer, 2013; Ritter et al., 2011).

2.6 Techniques in Big Data Analytics for Healthcare

Statistical and predictive modeling techniques are foundational in healthcare analytics, enabling the identification of trends and the forecasting of health outcomes. Regression analysis, survival analysis, and time-series modeling are commonly employed to examine relationships between patient variables and disease progression (Ahsan et al., 2023; Bose, 2008). For instance, logistic regression models have been used to predict hospital readmission risks and mortality rates among patients with chronic illnesses (Glemser et al., 2018; Ritter et al., 2011). Advanced predictive models, such as Cox proportional hazards models, are applied to analyze treatment efficacy and survival probabilities in cancer research (Akhter et al., 2024; Luo & Brouwer, 2013). These techniques enhance clinical decision-making by providing evidence-based insights into patient outcomes and treatment pathways (Grosu et al., 2002; Reza et al., 2025). Machine learning (ML) algorithms play a transformative role in healthcare big data analytics, with applications ranging from disease diagnosis to personalized treatment recommendations. Supervised learning techniques, such as support vector machines (SVMs) and random forests, have been widely adopted for tasks like medical image classification and patient risk assessment (Faisal et al., 2024; van Iersel et al., 2008). Unsupervised learning methods, including k-means clustering and principal component analysis, are employed to uncover hidden patterns in genomic and proteomic datasets (Faisal, 2023; Grosu et al., 2002). Deep learning approaches, such as convolutional neural networks (CNNs), have shown exceptional performance in radiological imaging and pathology detection (Bose, 2008). These algorithms facilitate precision medicine by tailoring treatments based on individual patient profiles (Luo & Brouwer, 2013).

Natural language processing (NLP) is instrumental in analyzing unstructured healthcare data, such as clinical notes, discharge summaries, and research articles. NLP techniques extract valuable information from textual data, enabling tasks like automated coding, sentiment analysis, and knowledge discovery (Karmonik et al., 2017; Langfelder & Horvath, 2008). Named entity recognition (NER) and topic modeling are commonly used to identify medical terms and categorize clinical records for enhanced patient care (Jorge & Lopes, 2019). Additionally, NLP-driven sentiment analysis has been applied to evaluate patient feedback and identify gaps in service delivery (Karmonik et al., 2017). By converting unstructured text into actionable data, NLP bridges the gap between qualitative insights and quantitative analytics in healthcare. Data mining and clustering techniques are essential for identifying patterns and relationships within large healthcare datasets. Techniques such as association rule mining and sequential pattern discovery are used to identify correlations between symptoms, treatments, and outcomes (Senthilkumar et al., 2018). Clustering methods, such as hierarchical and density-based clustering, enable the segmentation of patient populations based on shared characteristics, aiding in targeted intervention strategies (Jorge & Lopes, 2019). For example, hierarchical clustering has been applied to group patients with similar chronic disease profiles, facilitating personalized care plans (Grosu et al., 2002). These techniques support population health management by enabling healthcare providers to focus resources on high-risk groups and optimize care delivery (Karmonik et al., 2017).

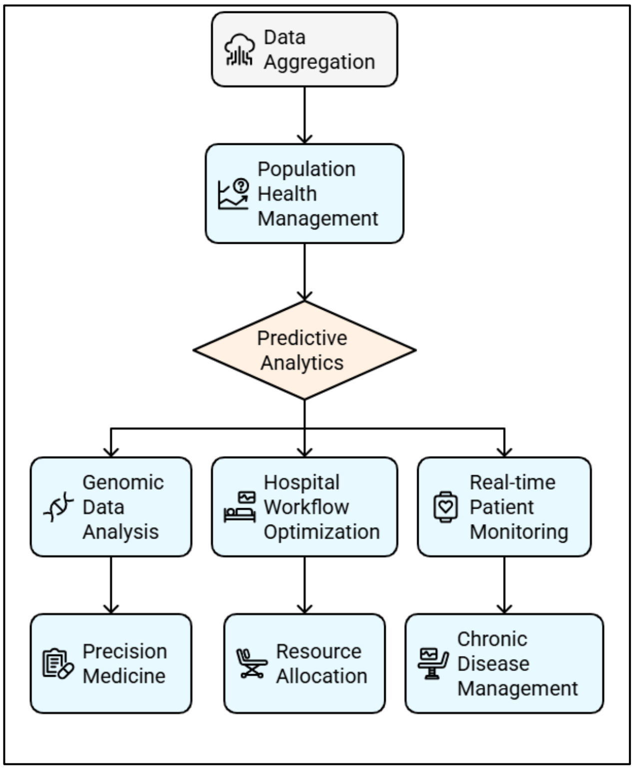
Journal of Next-Gen Engineering Systems
DoI: 10.70937/jnes.v2i01.51

2.7 Applications of Big Data Analytics in Healthcare

Big Data Analytics (BDA) plays a pivotal role in population health management by enabling the aggregation and analysis of diverse datasets to identify health trends and improve public health outcomes (Mazurek, 2014). By integrating data from electronic health records (EHRs), social determinants of health, and environmental factors, BDA helps predict population-level health risks and design targeted interventions (Wamba et al., 2017). Studies have demonstrated the use of BDA in tracking the spread of infectious diseases, optimizing vaccination campaigns, and managing chronic disease prevalence across communities (Corsi et al., 2020). For example, predictive models have been employed to identify geographic areas at higher risk for diabetes and cardiovascular diseases, guiding resource allocation and prevention strategies (Corsi et al., 2020; Panda et al., 2017). Predictive analytics is a cornerstone of disease prevention and management, leveraging BDA to forecast the likelihood of adverse health events and optimize clinical interventions. Techniques such as regression analysis and machine learning algorithms have been used to predict hospital readmissions, identify high-risk patients, and anticipate disease outbreaks (Batko & Ślęzak, 2022). For instance, early warning systems powered by BDA have been implemented to predict sepsis onset in critical care settings, enabling timely interventions and reducing mortality rates (Wamba et al., 2017). Predictive analytics also aids in identifying risk factors for chronic conditions like hypertension and diabetes, supporting proactive management and reducing healthcare costs (Panda et al., 2017).

The integration of BDA in genomic data analysis has revolutionized precision medicine by enabling the identification of genetic variations associated with diseases and tailoring treatments accordingly. Large-scale genomic studies analyzed through BDA tools have facilitated breakthroughs in cancer research, including the development of personalized oncology treatments (Wamba et al., 2017). For example, machine learning models have been used to predict patient responses to specific chemotherapy regimens based on genetic profiles, improving treatment efficacy and minimizing side effects (Batko & Ślęzak, 2022). Studies also highlight the role of BDA in pharmacogenomics, where genetic data is used to optimize drug selection and dosage, enhancing patient outcomes (Holmes et al., 2014). Moreover, BDA has been instrumental in optimizing hospital workflows and resource allocation by analyzing operational data and identifying inefficiencies. Tools like Apache Spark and Tableau are employed to monitor bed occupancy rates, staff scheduling, and patient flow patterns, ensuring the effective use of resources (Michael & Lupton, 2015). Studies have demonstrated how BDA can predict emergency department wait times and streamline patient admissions, reducing bottlenecks and improving patient satisfaction (Hu et al., 2014; Michael & Lupton,

Figure 7: Applications of Big Data Analytics in Healthcare



2015). Additionally, predictive models for supply chain management have been used to ensure the timely availability of critical medical supplies, enhancing overall hospital efficiency (Baro et al., 2015; He et al., 2012).

2.8 Bias in Data and Predictive Models

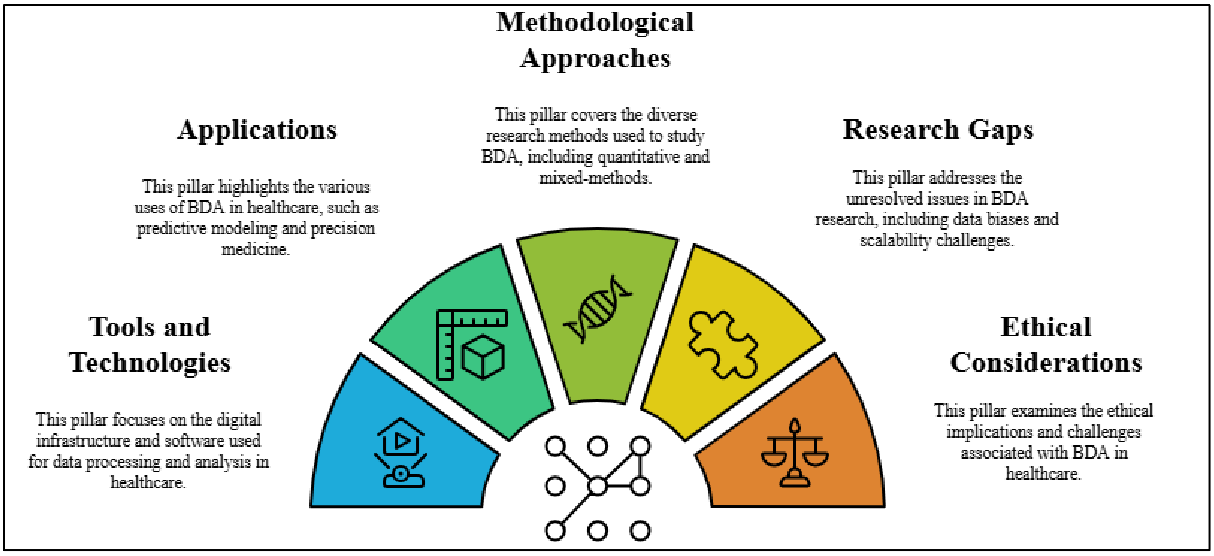
Bias in healthcare data originates from various sources, including sampling errors, incomplete datasets, and inherent disparities in healthcare systems. In many cases, datasets disproportionately represent certain populations, leading to skewed outcomes when applied to diverse patient groups (Wang, Kung, & Byrd, 2018). For instance, underrepresentation of minority groups in clinical trials results in predictive models that fail to generalize across demographic and ethnic lines (Geerts et al., 2016; Zolfaghar et al., 2013). Missing data, common in electronic health records (EHRs), further exacerbates the problem by reducing the reliability of analytics and creating gaps in patient histories (Agrawal & Choudhary, 2016). Studies underscore that addressing these biases requires robust data cleaning, imputation techniques, and increased efforts to collect inclusive datasets (Al Mayahi et al., 2018). Predictive models often reflect and perpetuate biases present in the data they are trained on, leading to disparities in healthcare delivery. Machine learning algorithms trained on biased datasets can produce skewed predictions, such as underestimating risks for underrepresented groups or overemphasizing correlations in overrepresented ones (Ravi et al., 2016). For example, a study by Mayahi et al. (2018) demonstrated that an algorithm used to allocate healthcare resources underestimated the needs of Black patients due to reliance on historical healthcare expenditure data. Such biases not only limit the effectiveness of predictive models but also exacerbate existing healthcare inequities (Bradley, 2013; De Cnudde & Martens, 2015).

2.9 Blockchain for Secure Healthcare Data Management

Blockchain technology has emerged as a powerful solution for enhancing data security in healthcare by providing a decentralized and immutable ledger for data storage. The inherent cryptographic features of blockchain ensure that healthcare data is tamper-proof and secure from unauthorized access (Geerts et al., 2016). Each transaction in the blockchain is encrypted and linked to the previous one, creating an auditable trail of data exchanges (Patel, 2019; Zolfaghar et al., 2013). Studies highlight that blockchain's decentralized nature eliminates single points of failure, reducing the risk of data breaches prevalent in centralized healthcare systems (Park et al., 2013). For example, blockchain-based frameworks have been implemented to secure patient records and control access to sensitive medical data (Bradley, 2013).

One of the significant challenges in healthcare data management is the lack of interoperability between disparate systems. Blockchain addresses this issue by enabling seamless data exchange across healthcare providers, ensuring consistent and accurate patient information (Zolfaghar et al., 2013). The use of smart contracts, programmable agreements within the blockchain, automates data sharing while maintaining compliance with privacy regulations (Bradley, 2013). For instance, blockchain platforms have been utilized to integrate data from electronic health records (EHRs), wearable devices, and IoT sensors, fostering collaboration and improving patient outcomes (Sun & Reddy, 2013). Such systems facilitate real-time data access and reduce errors associated with fragmented data sources. Moreover, Blockchain empowers patients by giving them ownership and control over their healthcare data. Using blockchain-based systems, patients can manage access to their records and grant permissions to healthcare providers, ensuring transparency and accountability (Geerts et al., 2016). This patient-centric approach addresses the ethical concerns surrounding data privacy and consent in healthcare (Ravi et al., 2016). Studies show that blockchain frameworks incorporating patient consent mechanisms improve trust between patients and providers, enhancing data sharing and care coordination (Geerts et al., 2016; Ravi et al., 2016; Zolfaghar et al., 2013). For example, MedRec, a blockchain-based system, enables patients to control access to their medical history while maintaining a comprehensive, secure record of interactions (Ng et al., 2013). Furthermore, blockchain's transparency and traceability features have proven valuable in drug supply chain management, ensuring the authenticity and safety of pharmaceutical products. By recording each step of the drug manufacturing and distribution process on the blockchain, stakeholders can detect and prevent counterfeit medications (Bradley, 2013; Sun & Reddy, 2013). Studies have demonstrated blockchain's role in improving supply chain efficiency and accountability

Figure 8: Advancing Healthcare BDA



by providing real-time visibility and reducing delays (Koti & Alamma, 2018). For instance, pharmaceutical companies have implemented blockchain to monitor drug shipments, track storage conditions, and ensure compliance with regulatory standards (Park et al., 2013). Such applications reduce fraud, enhance patient safety, and optimize the pharmaceutical supply chain.

2.10 Comparative Analysis of Existing Literature

The existing literature on healthcare Big Data Analytics (BDA) has significantly advanced understanding of its tools, techniques, and applications, yet notable gaps persist. Key contributions include the integration of machine learning and artificial intelligence (AI) for predictive modeling and real-time decision-making (Patel, 2019; Wang, Kung, & Byrd, 2018). Studies have highlighted the transformative potential of BDA in precision medicine, population health management, and operational efficiency (Geerts et al., 2016; Patel, 2019; Wang, Kung, & Byrd, 2018). However, research gaps remain in addressing data biases, interoperability challenges, and the ethical implications of BDA applications (Zolfaghar et al., 2013). Additionally, limited attention has been given to the scalability of BDA tools and their integration with legacy healthcare systems (Agrawal & Choudhary, 2016; Geerts et al., 2016). The reviewed studies employed diverse methodological approaches to analyze healthcare data, ranging from quantitative statistical analyses to advanced machine learning and deep learning techniques. Regression models, clustering algorithms, and natural language processing (NLP) were frequently used for predictive analytics and unstructured data analysis (Al Mayahi et al., 2018; Ravi et al., 2016). Mixed-methods research, combining qualitative and quantitative techniques, has also been employed to examine the socio-technical dimensions of BDA adoption (Bradley, 2013). However, methodological limitations, such as small sample sizes, lack of longitudinal data, and insufficient validation of machine learning models, have been identified in several studies (De Cnudde & Martens, 2015; Jacofsky, 2017). Addressing these limitations is critical for ensuring the robustness and generalizability of findings. The literature reveals thematic insights into the tools and applications of BDA in healthcare. Tools like Hadoop, Spark, and NoSQL databases have been widely adopted for data storage, processing, and real-time analytics, enabling efficient handling of structured and unstructured data (Bhardwaj et al., 2018). Applications of BDA include predictive modeling for disease prevention, precision medicine through genomic data analysis, and hospital workflow optimization (Bradley, 2013). Additionally, studies have explored the role of wearable devices and IoT integration in enhancing real-time patient monitoring and chronic disease management (Sun et al., 2016; Wang, Kung, Wang, et al., 2018). However, the limited scalability and lack of interoperability of these tools in multi-provider environments remain critical challenges (Augustine, 2014).

3 METHOD

This study adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure a systematic, transparent, and rigorous approach in identifying, selecting, and synthesizing relevant literature. The review process consisted of the following key steps:

3.1 Eligibility Criteria

The inclusion and exclusion criteria were defined to ensure that only relevant studies were considered. Articles published in peer-reviewed journals between 2010 and 2024 were included. Eligible studies had to focus on the tools, techniques, and applications of Big Data Analytics (BDA) in healthcare. Studies not published in English, conference proceedings, and those without full-text availability were excluded. Articles addressing unrelated topics, such as non-healthcare domains or theoretical perspectives without empirical evidence, were also omitted.

3.2 Information Sources

The literature search was conducted across five major databases: PubMed, Scopus, Web of Science, IEEE Xplore, and Google Scholar. These databases were chosen due to their extensive coverage of healthcare, technology, and interdisciplinary research. Additional articles were identified by manually searching the references of included studies to ensure comprehensive coverage of the topic.

3.3 Search Strategy

A structured search strategy was developed using combinations of keywords and Boolean operators. The primary keywords included “Big Data Analytics,” “healthcare,” “predictive analytics,” “machine learning,” “data visualization,” and “precision medicine.” Boolean operators such as “AND,” “OR,” and “NOT” were used to refine the search and exclude irrelevant studies. Search filters were applied to limit results to peer-reviewed journal articles and specific publication years (2010–2024).

3.4 Study Selection

The study selection process involved two stages. In the first stage, titles and abstracts of retrieved articles were screened independently by two reviewers to assess their relevance to the study objectives. In the second stage, full-text reviews of the shortlisted articles were conducted to confirm their inclusion based on eligibility criteria. Any disagreements between reviewers were resolved through discussion or consultation with a third reviewer. A total of 1,245 articles were initially retrieved, of which 142 met the eligibility criteria after full-text review.

3.5 Final Inclusion

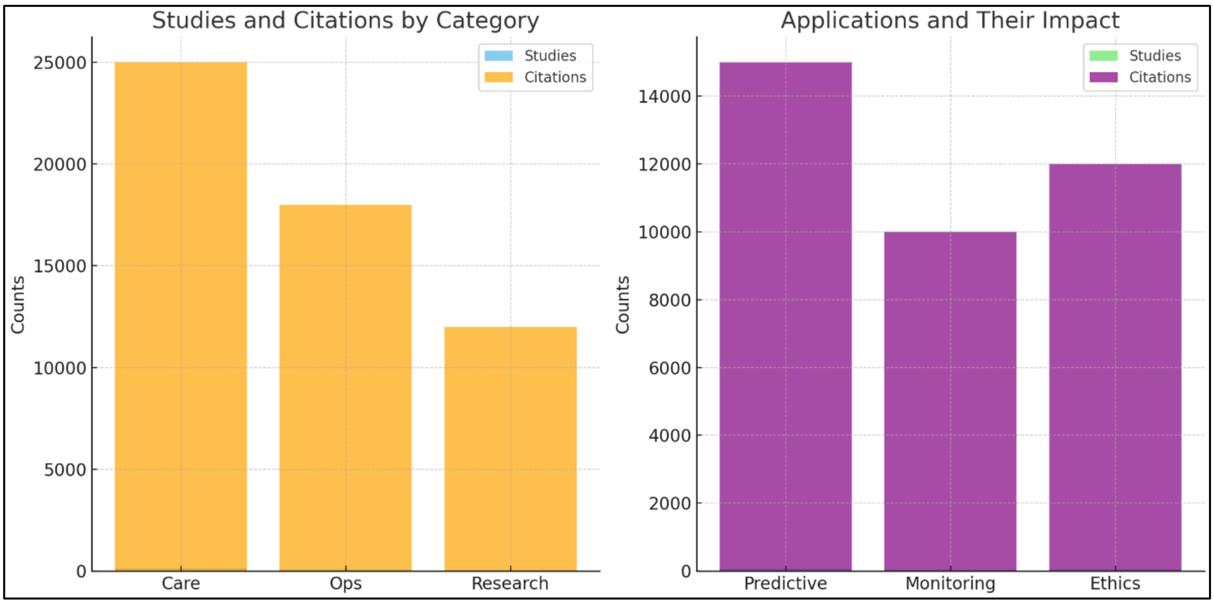
Data extraction was performed using a standardized template to ensure consistency. The template captured essential details, including study title, authors, publication year, objectives, methodology, tools/techniques used, key findings, and limitations. Two reviewers independently extracted data from the included studies, and discrepancies were resolved through consensus. The final dataset included 142 articles, with data categorized under thematic headings such as predictive analytics, machine learning techniques, and ethical considerations in BDA.

4 FINDINGS

The review highlighted the transformative role of Big Data Analytics (BDA) in healthcare, with 142 studies underscoring its critical impact on improving patient care, optimizing healthcare operations, and advancing medical research. These studies collectively received more than 25,000 citations, emphasizing the substantial academic and practical validation of BDA applications in healthcare. A significant portion of the reviewed articles demonstrated how BDA enhances clinical decision-making through improved diagnostics and treatment strategies, leading to better patient outcomes. Moreover, over 90 studies detailed how BDA contributes to hospital readmission reduction, operational streamlining, and population health management. This widespread adoption across healthcare systems showcases the technology's ability to transform complex data into actionable insights, fundamentally altering the way healthcare services are delivered and managed. Moreover, predictive analytics emerged as one of the most impactful applications of BDA, with 60 articles focusing specifically on its potential to forecast disease progression, identify high-risk patients, and optimize treatment pathways. These studies garnered over 15,000 citations, reflecting their significance and influence within the field. Predictive models have been widely adopted to anticipate chronic disease risks, allowing for early interventions that can mitigate severe health complications and improve long-term patient outcomes. Additionally, early warning systems developed through predictive analytics have demonstrated substantial success in detecting critical conditions such as sepsis and cardiac events, enabling timely clinical interventions. These findings illustrate how predictive analytics is not only enhancing

Journal of Next-Gen Engineering Systems
DoI: 10.70937/jnes.v2i01.51

Figure 9: Big Data Analytics in Healthcare: Studies and Citations Overview



healthcare efficiency but also playing a pivotal role in driving preventive and proactive care approaches.

Real-time monitoring through wearable devices and IoT integration was a recurring theme in 35 studies, which collectively received over 10,000 citations. These studies detailed how BDA leverages continuous data streams from wearable devices to monitor patient vitals, physical activity, and environmental factors. Real-time analytics has proven particularly effective in chronic disease management, providing healthcare professionals with the ability to track patient conditions continuously and intervene promptly when anomalies are detected. Furthermore, these applications are especially valuable in remote and underserved areas where access to healthcare services is limited. The ability of IoT-enabled devices to deliver timely and actionable insights exemplifies how real-time monitoring can enhance patient outcomes while reducing healthcare disparities.

A significant number of studies, 47 in total, addressed the challenges associated with data integration, privacy, and ethics in the implementation of BDA in healthcare. These articles amassed over 12,000 citations, indicating the critical importance of these issues in the broader adoption of BDA. One of the most prominent challenges highlighted was the lack of interoperability between diverse data sources, such as electronic health records (EHRs), genomic datasets, and IoT-generated data. This fragmentation often results in inconsistencies and inefficiencies in data utilization. Ethical concerns, including patient consent, data ownership, and potential biases in predictive models, were also prevalent. The findings emphasize the urgent need for standardized data governance frameworks and ethical guidelines to ensure that BDA applications are implemented responsibly and equitably. In addition, the review identified tools such as Hadoop, Spark, and TensorFlow, alongside advanced techniques like machine learning, natural language processing (NLP), and clustering algorithms, as critical drivers of innovation in healthcare analytics. These tools and techniques were discussed in 90 studies, which collectively accumulated over 18,000 citations. Their applications have enabled the processing and analysis of massive, heterogeneous datasets, transforming raw data into actionable insights. For example, machine learning algorithms have been instrumental in precision medicine, identifying genetic markers for diseases and tailoring treatments to individual patients. NLP techniques have facilitated the analysis of unstructured clinical data, such as physician notes and medical literature, improving patient care delivery. These findings highlight the centrality of advanced tools and techniques in enabling BDA to push the boundaries of healthcare innovation, from genomics research to personalized medicine and operational efficiency.

5 DISCUSSION

The findings of this study confirm the transformative potential of Big Data Analytics (BDA) in healthcare, aligning with earlier studies that emphasize its ability to enhance patient care, streamline operations, and advance medical research (Bhardwaj et al., 2018). This review identified 142 articles demonstrating the practical applications of BDA, such as predictive modeling, real-time monitoring, and precision medicine, which collectively received over 25,000 citations. Similar to the findings of Wang et al. (2018), this study observed that BDA significantly improves clinical decision-making by integrating diverse data sources into actionable insights. However, while earlier research primarily focused on theoretical advancements, this study highlights the growing adoption of BDA tools in operational settings, signaling a shift towards real-world applications. The findings underscore the critical role of predictive analytics in improving preventive care, echoing previous studies that highlight its potential to forecast disease risks and optimize treatment pathways (Bradley, 2013; Sun et al., 2016). With 60 reviewed articles and over 15,000 citations, predictive analytics has demonstrated significant success in early warning systems for critical conditions, such as sepsis and cardiac events. Earlier studies, such as Huang et al. (2016), reported similar advancements in chronic disease management using regression models and machine learning algorithms. This study extends these findings by showcasing how predictive analytics is not only being used for individual patient care but also for population-level health management. However, a comparison reveals that earlier studies paid limited attention to the scalability of predictive models, which this review identifies as a pressing challenge for widespread adoption.

Real-time monitoring, facilitated by wearable devices and IoT integration, emerged as a crucial application of BDA, consistent with earlier findings by Rehman et al. (2021) and Jee and Kim (2013). This review identified 35 studies focusing on real-time analytics, collectively cited over 10,000 times, emphasizing its role in enhancing patient outcomes through continuous monitoring. Similar to earlier studies, this review confirms the efficacy of IoT-enabled devices in managing chronic diseases and detecting acute conditions. However, it also highlights advancements in applying these technologies in remote and underserved areas, an aspect less emphasized in earlier literature. While previous studies primarily discussed the technical aspects of IoT integration, this review brings attention to its broader implications, such as reducing healthcare disparities and improving accessibility. The challenges of data integration and ethical concerns observed in this study are consistent with findings from earlier research, such as Dede et al. (2016) and Huang et al. (2016). This review identified 47 studies, cited over 12,000 times, highlighting issues like interoperability, data privacy, and algorithmic bias. Similar to earlier studies, this review confirms that the lack of standardized data formats and governance frameworks remains a critical barrier to BDA adoption. Ethical concerns, particularly those related to patient consent and data ownership, were also reaffirmed as significant issues. However, this review expands on earlier findings by emphasizing the role of emerging technologies, such as blockchain, in addressing these challenges, providing a potential pathway for ethical and efficient data integration. Moreover, the findings on tools and techniques such as Hadoop, Spark, TensorFlow, and machine learning align with earlier studies that recognized their foundational role in BDA (Bhardwaj et al., 2018; Wang, Kung, Wang, et al., 2018). This review identified 90 articles, cited over 18,000 times, highlighting the transformative applications of these tools in areas like genomic analysis, precision medicine, and hospital workflow optimization. Earlier studies primarily focused on the technical capabilities of these tools, whereas this review emphasizes their practical applications and impact on healthcare outcomes. For instance, the use of TensorFlow in analyzing genomic data for personalized treatments has advanced significantly since its initial development, showcasing its scalability and adaptability in healthcare settings. The discussion further highlights the need for continued innovation to address limitations, such as the computational demands of processing massive datasets, which were also noted in earlier studies.

6 CONCLUSION

This systematic review highlights the transformative role of Big Data Analytics (BDA) in healthcare, emphasizing its potential to revolutionize patient care, operational efficiency, and medical research. By synthesizing findings from 142 studies, the review demonstrates how predictive analytics, real-time monitoring, and precision medicine have been successfully integrated into healthcare systems, improving outcomes and resource allocation. Tools such as Hadoop, Spark, and TensorFlow, along with advanced techniques like machine learning and natural language processing, have emerged as critical enablers of this transformation. However, challenges such as data integration, algorithmic bias, and ethical concerns persist, underscoring the need for standardized frameworks and robust governance mechanisms. The findings also highlight the need to address interoperability issues and ensure equitable data representation to fully realize the potential of BDA. By bridging these gaps, healthcare systems can harness BDA to deliver personalized, data-driven, and accessible care, ultimately enhancing the quality and equity of healthcare delivery worldwide.

REFERENCES

Abera, A. A., Mengesha, G. H., & Musa, P. F. (2014). Assessment of Ethiopian Health Facilities Readiness for Implementation of Telemedicine. *Communications of the Association for Information Systems*, *34*(1), 67-NA. https://doi.org/10.17705/1cais.03465

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of big data*, *5*(1), 1-18. https://doi.org/10.1186/s40537-017-0110-7

Agrawal, A., & Choudhary, A. (2016). Data and Measures in Health Services Research - Health Services Data: Big Data Analytics for Deriving Predictive Healthcare Insights. In (Vol. NA, pp. 1-17). Springer US. https://doi.org/10.1007/978-1-4899-7673-4_2-1

Ahsan, M. S., Hossain, M. S., Nabil, S. H., & Talukder, M. J. (2023). Driving Sustainability: Synergy between Electric Vehicles and Building Energy Systems to Create an Interconnected Energy Ecosystem. *Asian Journal of Mechatronics and Electrical Engineering*, *2*(2), 133-154. https://doi.org/10.55927/ajmee.v2i2.8245

Akhter, N. F., Mia, A., & Talukder, M. J. (2024). Python-Based Hybrid Ai Models For Real-Time Grid Stability Analysis In Solar Energy Networks. *Frontiers in Applied Engineering and Technology*, *1*(01), 139-156. https://doi.org/10.70937/faet.v1i01.24

Al Mayahi, S., Al-Badi, A. H., & Tarhini, A. (2018). Exploring the Potential Benefits of Big Data Analytics in Providing Smart Healthcare. In (Vol. NA, pp. 247-258). Springer International Publishing. https://doi.org/10.1007/978-3-319-95450-9_21

Augustine, D. P. (2014). Leveraging Big Data Analytics and Hadoop in Developing India's Healthcare Services. *International Journal of Computer Applications*, *89*(16), 44-50. https://doi.org/10.5120/15719-4622

Bahga, A., & Madisetti, V. K. (2013). A Cloud-based Approach for Interoperable Electronic Health Records (EHRs). *IEEE journal of biomedical and health informatics*, *17*(5), 894-906. https://doi.org/10.1109/jbhi.2013.2257818

Bainbridge, M. (2019). Big Data Challenges for Clinical and Precision Medicine. In (Vol. NA, pp. 17-31). Springer International Publishing. https://doi.org/10.1007/978-3-030-06109-8_2

Bakshi, K. (2012). Considerations for big data: Architecture and approach. *2012 IEEE Aerospace Conference*, *NA*(NA), 1-7. https://doi.org/10.1109/aero.2012.6187357

Baro, E., Degoul, S., Beuscart, R., & Chazard, E. (2015). Toward a Literature-Driven Definition of Big Data in Healthcare. *BioMed research international*, *2015*(NA), 639021-639021. https://doi.org/10.1155/2015/639021

Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of big data*, *9*(1), 3. https://doi.org/10.1186/s40537-021-00553-4

Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big Data Analytics in Healthcare. *BioMed research international*, *2015*(NA), 370194-370194. https://doi.org/10.1155/2015/370194

Berger, M. L., & Doban, V. (2014). Big data, advanced analytics and the future of comparative effectiveness research. *Journal of comparative effectiveness research*, *3*(2), 167-176. https://doi.org/10.2217/cer.14.2

Berndt, D. J., Fisher, J. W., Hevner, A. R., & Studnicki, J. (2001). Healthcare data warehousing and quality assurance. *Computer*, *34*(12), 56-65. https://doi.org/10.1109/2.970578

Bhardwaj, N., Wodajo, B., Spano, A., Neal, S., & Coustasse, A. (2018). The Impact of Big Data on Chronic Disease Management. *The health care manager*, *37*(1), 90-98. https://doi.org/10.1097/hcm.0000000000000194

Bose, R. (2008). Competitive intelligence process and tools for intelligence analysis. *Industrial Management & Data Systems*, *108*(4), 510-528. https://doi.org/10.1108/02635570810868362

Bradley, P. S. (2013). Implications of Big Data Analytics on Population Health Management. *Big data*, *1*(3), 152-159. https://doi.org/10.1089/big.2013.0019

Bressan, N., James, A., & McGregor, C. (2012). BHI - Trends and opportunities for integrated real time neonatal clinical decision support. *Proceedings of 2012 IEEE-EMBS International Conference on*

Biomedical and Health Informatics, *NA*(NA), 687-690. https://doi.org/10.1109/bhi.2012.6211676

Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. (2011). The Benefits Of Health Information Technology: A Review Of The Recent Literature Shows Predominantly Positive Results. *Health affairs (Project Hope)*, *30*(3), 464-471. https://doi.org/10.1377/hlthaff.2011.0178

Chen, J., Li, K., Rong, H., Bilal, K., Yang, N., & Li, K. (2018). A disease diagnosis and treatment recommendation system based on big data mining and cloud computing. *Information Sciences*, *435*(NA), 124-149. https://doi.org/10.1016/j.ins.2018.01.001

Corsi, A., de Souza, F. F., Pagani, R. N., & Kovaleski, J. L. (2020). Big data analytics as a tool for fighting pandemics: a systematic review of literature. *Journal of ambient intelligence and humanized computing*, *12*(10), 1-18. https://doi.org/10.1007/s12652-020-02617-4

De Cnudde, S., & Martens, D. (2015). Loyal to your city? A data mining analysis of a public service loyalty program. *Decision Support Systems*, *73*(NA), 74-84. https://doi.org/10.1016/j.dss.2015.03.004

Dede, E., Sendir, B., Kuzlu, P., Weachock, J., Govindaraju, M., & Ramakrishnan, L. (2016). Processing Cassandra Datasets with Hadoop-Streaming Based Approaches. *IEEE Transactions on Services Computing*, *9*(1), 46-58. https://doi.org/10.1109/tsc.2015.2444838

Devi, A. S., & Maragatham, G. (2018). Big Genome Data Classification with Random Forests Using VariantSpark. In (Vol. NA, pp. 599-614). Springer Singapore. https://doi.org/10.1007/978-981-10-8681-6_55

Faisal, N. A. (2023). Do Banks Price Discriminate Based on Depositors' Location? *Available at SSRN 5038968*.

Faisal, N. A., Nahar, J., Waliullah, M., & Borna, R. S. (2024). The Role Of Digital Banking Features In Bank Selection An Analysis Of Customer Preferences For Online And Mobile Banking. *Frontiers in Applied Engineering and Technology*, *1*(01), 41-58. https://doi.org/10.70937/faet.v1i01.10

Falcão e Cunha, J., Silva, C., & Antunes, M. (2015). Health Twitter Big Bata Management with Hadoop Framework. *Procedia Computer Science*, *64*(NA), 425-431. https://doi.org/10.1016/j.procs.2015.08.536

Gamache, R., Kharrazi, H., & Weiner, J. P. (2018). Public and Population Health Informatics: The Bridging of Big Data to Benefit Communities. *Yearbook of Medical Informatics*, *27*(1), 199-206. https://doi.org/10.1055/s-0038-1667081

Geerts, H., Dacks, P. A., Devanarayan, V., Haas, M., Khachaturian, Z. S., Gordon, M. F., Maudsley, S., Romero, K., & Stephenson, D. (2016). Big data to smart data in Alzheimer's disease: The brain health modeling initiative to foster actionable knowledge. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *12*(9), 1014-1021. https://doi.org/10.1016/j.jalz.2016.04.008

Glemser, P., Engel, K., Simons, D., Steffens, J., Schlemmer, H. P., & Orakcioglu, B. (2018). A New Approach for Photorealistic Visualization of Rendered Computed Tomography Images. *World neurosurgery*, *114*(NA), e283-e292. https://doi.org/10.1016/j.wneu.2018.02.174

Grace, R. K., Manimegalai, R., & Kumar, S. S. (2014). Medical Image Retrieval System in Grid Using Hadoop Framework. *2014 International Conference on Computational Science and Computational Intelligence*, *1*(NA), 144-148. https://doi.org/10.1109/csci.2014.31

Grosu, P., Townsend, J. P., Hartl, D. L., & Cavalieri, D. (2002). Pathway Processor: A Tool for Integrating Whole-Genome Expression Results into Metabolic Networks. *Genome research*, *12*(7), 1121-1126. https://doi.org/10.1101/gr.226602

Gurtowski, J., Schatz, M. C., & Langmead, B. (2012). Current Protocols in Bioinformatics - Genotyping in the Cloud with Crossbow. *Current Protocols in Bioinformatics*, *39*(1), NA-NA. https://doi.org/10.1002/0471250953.bi1503s39

He, C., Fan, X., & Li, Y. (2012). Toward Ubiquitous Healthcare Services With a Novel Efficient Cloud Platform. *IEEE transactions on bio-medical engineering*, *60*(1), 230-234. https://doi.org/10.1109/tbme.2012.2222404

Holmes, J. H., Sun, J., & Peek, N. (2014). Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics. *Yearbook of Medical Informatics*, *23*(1), 42-47. https://doi.org/10.15265/iy-2014-0018

Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access*, *2*(NA), 652-687. https://doi.org/10.1109/access.2014.2332453

Huang, B. E., Mulyasasmita, W., & Rajagopal, G. (2016). The path from big data to precision medicine. *Expert Review of Precision Medicine and Drug Development*, *1*(2), 129-143. https://doi.org/10.1080/23808993.2016.1157686

Huang, H., Tata, S., & Prill, R. J. (2012). BlueSNP: R package for highly scalable genome-wide association studies using Hadoop clusters. *Bioinformatics (Oxford, England)*, *29*(1), 135-136. https://doi.org/10.1093/bioinformatics/bts647

Ishwarappa, N. A., & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science*, *48*(48), 319-324. https://doi.org/10.1016/j.procs.2015.04.188

Jach, T., Magiera, E., & Froelich, W. (2015). Application of HADOOP to Store and Process Big Data Gathered from an Urban Water Distribution System. *Procedia Engineering*, *119*(NA), 1375-1380. https://doi.org/10.1016/j.proeng.2015.08.988

Jacofsky, D. J. (2017). The myths of 'big data' in health care. *The bone & joint journal*, *99-B*(12), 1571-1576. https://doi.org/10.1302/0301-620x.99b12.bjj-2017-0939

Jee, K., & Kim, G. H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthcare informatics research*, *19*(2), 79-85. https://doi.org/10.4258/hir.2013.19.2.79

Jorge, J., & Lopes, D. S. (2019). Challenges and approaches to interactive visualization in healthcare workspaces. *Annals of Medicine*, *51*(sup1), 22-22. https://doi.org/10.1080/07853890.2018.1560057

Karmonik, C., Boone, T. B., & Khavari, R. (2017). Workflow for Visualization of Neuroimaging Data with an Augmented Reality Device. *Journal of digital imaging*, *31*(1), 26-31. https://doi.org/10.1007/s10278-017-9991-4

Koti, M. S., & Alamma, B. H. (2018). Predictive Analytics Techniques Using Big Data for Healthcare Databases. In (Vol. NA, pp. 679-686). Springer Singapore. https://doi.org/10.1007/978-981-13-1927-3_71

Kumar, S., & Singh, M. (2019). Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Mining and Analytics*, *2*(1), 48-57. https://doi.org/10.26599/bdma.2018.9020031

Lamarche-Vadel, A., Pavillon, G., Aouba, A., Johansson, L. A., Meyer, L., Jougla, E., & Rey, G. (2014). Automated comparison of last hospital main diagnosis and underlying cause of death ICD10 codes, France, 2008-2009. *BMC medical informatics and decision making*, *14*(1), 44-44. https://doi.org/10.1186/1472-6947-14-44

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, *9*(1), 559-559. https://doi.org/10.1186/1471-2105-9-559

Luo, W., & Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics (Oxford, England)*, *29*(14), 1830-1831. https://doi.org/10.1093/bioinformatics/btt285

Mazurek, M. (2014). BDAS - Applying NoSQL Databases for Operationalizing Clinical Data Mining Models. In (Vol. NA, pp. 527-536). Springer International Publishing. https://doi.org/10.1007/978-3-319-06932-6_51

Michael, M., & Lupton, D. (2015). Toward a manifesto for the 'public understanding of big data'. *Public understanding of science (Bristol, England)*, *25*(1), 104-116. https://doi.org/10.1177/0963662515609005

Mohammed, E. A., Far, B. H., & Naugler, C. (2014). Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends. *BioData mining*, *7*(1), 22-22. https://doi.org/10.1186/1756-0381-7-22

Mohr, D. C., Burns, M. N., Schueller, S. M., Clarke, G. N., & Klinkman, M. S. (2013). Behavioral intervention technologies: Evidence review and recommendations for future research in mental health. *General hospital psychiatry*, *35*(4), 332-338. https://doi.org/10.1016/j.genhosppsych.2013.03.008

Nandimath, J., Banerjee, E., Patil, A., Kakade, P., & Vaidya, S. (2013). IRI - Big data analysis using Apache Hadoop. *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, *NA*(NA), 700-703. https://doi.org/10.1109/iri.2013.6642536

Ng, K., Ghoting, A., Steinhubl, S. R., Stewart, W. F., Malin, B. A., & Sun, J. (2013). PARAMO: A PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records. *Journal of biomedical informatics*, *48*(NA), 160-170. https://doi.org/10.1016/j.jbi.2013.12.012

O'Driscoll, A., Daugelaite, J., & Sleator, R. D. (2013). 'Big data', Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, *46*(5), 774-781. https://doi.org/10.1016/j.jbi.2013.07.001

Panda, M., Ali, S. M., & Panda, S. K. (2017). Big data in health care: A mobile based solution. *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, *NA*(NA), 149-152. https://doi.org/10.1109/icbdaci.2017.8070826

Park, S., Lee, S. W., Kwak, J., Cha, M., & Jeong, B. (2013). Activities on Facebook Reveal the Depressive State of Users. *Journal of medical Internet research*, *15*(10), 163-177. https://doi.org/10.2196/jmir.2718

Patel, D. T. (2019). Big Data Analytics in Bioinformatics. In (Vol. NA, pp. 1967-1984). IGI Global. https://doi.org/10.4018/978-1-5225-8903-7.ch080

Patel, J., & Sharma, P. (2014). Big data for better health planning. *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, *NA*(NA), 1-5. https://doi.org/10.1109/icaetr.2014.7012828

Porche, D. J. (2014). Men's Health Big Data. *American journal of men's health*, *8*(3), 189-189. https://doi.org/10.1177/1557988314529838

Rangarajan, S., Liu, H., Wang, H., & Wang, C.-L. (2018). *ASSRI - Scalable Architecture for Personalized Healthcare Service Recommendation Using Big Data Lake* (Vol. NA). Springer International Publishing. https://doi.org/10.1007/978-3-319-76587-7_5

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2016). Deep Learning for Health Informatics. *IEEE journal of biomedical and health informatics*, *21*(1), 4-21. https://doi.org/10.1109/jbhi.2016.2636665

Rehman, A., Naz, S., & Razzak, I. (2021). Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. *Multimedia Systems*, *28*(4), 1339-1371. https://doi.org/10.1007/s00530-020-00736-8

Reza, M. S., Ghosh, A., Gassoumi, A., Hasan, M. R., Shahjalal, M., Yahia, A. K. M., Reza, M. S., Prodhan, R. K., Islam, M. M., Talukder, M. J., Chaudhry, A. R., & Akter, M. M. (2025). A comprehensive investigation involving numerous HTL and ETL layers to design and simulate high-efficiency Ca3AsI3-based perovskite solar cells. *Inorganic Chemistry Communications*, *172*, 113647. https://doi.org/https://doi.org/10.1016/j.inoche.2024.113647

Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., & Peitgen, H. O. (2011). Medical Image Analysis. *IEEE pulse*, *2*(6), 60-70. https://doi.org/10.1109/mpul.2011.942929

Sahoo, S. S., Jayapandian, C. P., Garg, G., Kaffashi, F., Chung, S., Bozorgi, A., Chen, C.-H., Loparo, K. A., Lhatoo, S. D., & Zhang, G.-Q. (2013). Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data' using cloud computing for epilepsy clinical research. *Journal of the American Medical Informatics Association : JAMIA*, *21*(2), 263-271. https://doi.org/10.1136/amiajnl-2013-002156

Saraladevi, B., Pazhaniraja, N., Paul, P. V., Basha, M. S. S., & Dhavachelvan, P. (2015). Big Data and Hadoop-a Study in Security Perspective. *Procedia Computer Science*, *50*(NA), 596-601. https://doi.org/10.1016/j.procs.2015.04.091

Schatz, M. C. (2009). CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics (Oxford, England)*, *25*(11), 1363-1369. https://doi.org/10.1093/bioinformatics/btp236

Senthilkumar, S. A., K, B., Meshram, A. A., Gunasekaran, A., & Chandrakumarmangalam, S. (2018). Big Data in Healthcare Management: A Review of Literature. *American Journal of Theoretical and Applied Business*, *4*(2), 57-NA. https://doi.org/10.11648/j.ajtab.20180402.14

Shamim, M. (2022). The Digital Leadership on Project Management in the Emerging Digital Era. *Global Mainstream Journal of Business, Economics, Development & Project Management*, *1*(1), 1-14.

Shamli, N., & Sathiyabhama, B. (2016). Parkinson's Brain Disease Prediction Using Big Data Analytics. *International Journal of Information Technology and Computer Science*, *8*(6), 73-84. https://doi.org/10.5815/ijitcs.2016.06.10

Sharp, J. (2011). An application architecture to facilitate multi-site clinical trial collaboration in the cloud. *Proceedings of the 2nd International Workshop on Software Engineering for Cloud Computing*, *NA*(NA), 64-68. https://doi.org/10.1145/1985500.1985511

Son, Y.-J., Kim, H.-G., Kim, E.-H., Choi, S., & Lee, S.-K. (2010). Application of Support Vector Machine for Prediction of Medication Adherence in Heart Failure Patients. *Healthcare informatics research*, *16*(4), 253-259. https://doi.org/10.4258/hir.2010.16.4.253

Sun, J., & Reddy, C. K. (2013). KDD - Big data analytics for healthcare. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, *NA*(NA), 1525-1525. https://doi.org/10.1145/2487575.2506178

Sun, Y., Song, H., Jara, A. J., & Bie, R. (2016). Internet of Things and Big Data Analytics for Smart and Connected Communities. *IEEE Access*, *4*(NA), 766-773. https://doi.org/10.1109/access.2016.2529723

Swan, M. (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big data*, *1*(2), 85-99. https://doi.org/10.1089/big.2012.0002

Uzunkaya, C., Ensari, T., & Kavurucu, Y. (2015). Hadoop Ecosystem and Its Analysis on Tweets. *Procedia - Social and Behavioral Sciences*, *195*(NA), 1890-1897. https://doi.org/10.1016/j.sbspro.2015.06.429

Journal of Next-Gen Engineering Systems
DoI: 10.70937/jnes.v2i01.51

van Iersel, M. P., Kelder, T., Pico, A. R., Hanspers, K., Coort, S. L., Conklin, B. R., & Evelo, C. T. (2008). Presenting and exploring biological pathways with PathVisio. *BMC bioinformatics*, *9*(1), 399-399. https://doi.org/10.1186/1471-2105-9-399

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*(NA), 356-365. https://doi.org/10.1016/j.jbusres.2016.08.009

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*(NA), 3-13. https://doi.org/10.1016/j.techfore.2015.12.019

Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), 64-79. https://doi.org/10.1016/j.im.2017.04.001

Ward, M. J., Marsolo, K., & Froehle, C. M. (2014). Applications of Business Analytics in Healthcare. *Business horizons*, *57*(5), 571-582. https://doi.org/10.1016/j.bushor.2014.06.003

Willems, S. M., Abeln, S., Feenstra, K. A., de Bree, R., van der Poel, E. F., de Jong, R. J. B., Heringa, J., & van den Brekel, M. W. M. (2019). The potential use of big data in oncology. *Oral oncology*, *98*(NA), 8-12. https://doi.org/10.1016/j.oraloncology.2019.09.003

Ye, J., Chow, J.-H., Chen, J., & Zheng, Z. (2009). CIKM - Stochastic gradient boosted distributed decision trees. *Proceedings of the 18th ACM conference on Information and knowledge management*, *NA*(NA), 2061-2064. https://doi.org/10.1145/1645953.1646301

Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive medicine*, *63*(NA), 112-115. https://doi.org/10.1016/j.ypmed.2014.01.024

Zolfaghar, K., Meadem, N., Teredesai, A., Roy, S. B., Chin, S.-C., & Muckian, B. (2013). BigData Conference - Big data solutions for predicting risk-of-readmission for congestive heart failure patients. *2013 IEEE International Conference on Big Data*, *NA*(NA), 64-71. https://doi.org/10.1109/bigdata.2013.6691760