Big Data Analytics in Healthcare: A Systematic Literature Review

Mirza Umer Ikram
Department of Computer Science
University of Central Punjab

Ruhma Bilal
Department of Computer Science
University of Central Punjab

Moiz Ahmed
Department of Computer Science
University of Central Punjab

Abstract—**Big Data Analytics (BDA) has become increasingly important in healthcare due to its potential to improve patient outcomes, reduce costs, and support decision-making. This systematic literature review (SLR) explores key trends, application areas, data sources, tools, methods, challenges, outcomes, and future directions of BDA in healthcare by synthesizing findings from over 30 peer-reviewed studies published between 2013 and 2024.**

KEYWORDS—

Big Data Analytics, Healthcare Informatics, Electronic Health Records (EHR), Machine Learning in Medicine, Predictive Analytics, Clinical Decision Support, Systematic Literature Review (SLR), IoT in Healthcare, Data Privacy, Disease Diagnosis and Prediction

I. INTRODUCTION

The rise in global health data has led to a shift in how healthcare organizations operate. Big data, characterized by high volume, velocity, variety, veracity, and value, has revolutionized disease detection, patient monitoring, and personalized care. This paper conducts a systematic literature review to answer 12 carefully structured research questions (RQs) across six thematic areas, helping stakeholders understand how BDA is shaping modern healthcare systems. The healthcare industry is undergoing a major transformation driven by the unprecedented growth of digital data and technological innovation. The proliferation of electronic health records (EHRs), wearable devices, genomics, medical imaging, and Internet of Things (IoT) technologies has led to an exponential increase in healthcare data, both in volume and complexity. According to recent estimates, healthcare data is projected to grow faster than in any other industry, with a compound annual growth rate (CAGR) of over 35%. This explosion of data provides both an opportunity and a challenge: while vast data reserves offer insights into disease patterns, patient outcomes, and operational efficiency, their true potential can only be unlocked through advanced analytical techniques. Big Data Analytics (BDA) refers to the process of examining large and varied datasets—also known as big data—to uncover hidden patterns, unknown correlations, and other useful information that can help organizations make informed decisions. In healthcare, BDA has emerged as a powerful enabler of improved patient care, personalized treatment, real-time disease surveillance, and predictive modeling. The unique characteristics of big data—often described by the five Vs: Volume, Velocity, Variety, Veracity, and Value—make traditional data processing tools inadequate for extracting meaningful insights. Instead, new paradigms such as machine learning, artificial intelligence (AI), and distributed computing platforms like Hadoop and Spark have been adopted to address these challenges. The integration of BDA into healthcare has demonstrated significant benefits across a wide range of applications. These include early disease detection, resource optimization, hospital readmission reduction, clinical decision support, and pandemic outbreak tracking. For instance, machine learning algorithms trained on EHRs can predict the likelihood of complications such as sepsis or readmission, allowing clinicians to take proactive measures. Similarly, genomics-driven analytics are opening doors to personalized medicine, where treatments can be tailored to an individual's genetic profile. Despite these promising developments, several challenges remain. Healthcare data is often fragmented across disparate systems, lacks standardization, and includes sensitive personal information, raising concerns about interoperability, data privacy, and security. Moreover, many healthcare institutions struggle with the lack of technical infrastructure and skilled personnel required to implement and maintain BDA systems effectively. Addressing these challenges is essential for realizing the full potential of BDA in healthcare. Given the rapidly evolving landscape and the increasing importance of data-driven decision-making in medicine, there is a pressing need to consolidate existing research to identify prevailing trends, common tools, techniques, challenges, and outcomes. Systematic literature reviews (SLRs) serve this purpose by offering a structured synthesis of prior work, revealing research gaps, and guiding future investigations. This paper presents a comprehensive SLR focused on Big Data Analytics in healthcare, covering over 30 peer-reviewed studies published between 2013 and 2024. The review is structured around twelve research questions (RQs) grouped under six thematic areas: healthcare applications, data types and sources, analytical methods and tools, limitations and challenges, observed outcomes, and future research directions. The methodology adopted for this review ensures rigorous selection, assessment, and analysis of the relevant literature using a tollgate approach. Additionally, this study incorporates two real-world healthcare datasets to reinforce empirical understanding and support practical insights. By synthesizing a broad range of academic and applied research, this paper aims to provide researchers, clinicians, and policymakers with a holistic view of the current state and future potential of Big Data Analytics in healthcare. The insights derived from this review can serve as a foundation for developing data-driven healthcare solutions that are effective, ethical, and scalable across global health systems. During the COVID-19 pandemic, the importance of BDA became even more evident. Data-driven models helped forecast infection rates, identify high-risk populations, optimize ICU resource allocation, and guide policy decisions on lockdowns and vaccine distribution. These examples highlight how timely and accurate data analytics can save lives during public health emergencies.

Despite these promising developments, several challenges remain. Healthcare data is often fragmented across disparate systems, lacks standardization, and includes sensitive personal information, raising concerns about interoperability, data privacy, and security. Moreover, many healthcare institutions struggle with the lack of technical infrastructure and skilled personnel required to implement and maintain BDA systems effectively. Legal and ethical concerns such as consent management, algorithmic bias, and data ownership further complicate widespread adoption.

Addressing these challenges requires a multidisciplinary approach involving data scientists, clinicians, policy-makers, and technologists. Collaboration across domains is essential for designing BDA systems that are technically robust, clinically useful, ethically sound, and socially acceptable.

Given the rapidly evolving landscape and the increasing importance of data-driven decision-making in medicine, there is a pressing need to consolidate existing research to identify prevailing trends, common tools, techniques, challenges, and outcomes. Systematic literature reviews (SLRs) serve this purpose by offering a structured synthesis of prior work, revealing research gaps, and guiding future investigations.

This paper presents a comprehensive SLR focused on Big Data Analytics in healthcare, covering over 30 peer-reviewed studies published between 2013 and 2024. The review is structured around twelve research questions (RQs) grouped under six thematic areas: healthcare applications, data types and sources, analytical methods and tools, limitations and challenges, observed outcomes, and future research directions. The methodology adopted for this review ensures rigorous selection, assessment, and analysis of the relevant literature using a tollgate approach. Additionally, this study incorporates two real-world healthcare datasets to reinforce empirical understanding and support practical insights.

By synthesizing a broad range of academic and applied research, this paper aims to provide researchers, clinicians, and policymakers with a holistic view of the current state and future potential of Big Data Analytics in healthcare. The insights derived from this review can serve as a foundation for developing data-driven healthcare solutions that are effective, ethical, and scalable across global health systems.

II. Research Methodology

This study adopts a systematic literature review (SLR) methodology to investigate the role and impact of Big Data Analytics (BDA) in healthcare. The goal is to provide a structured synthesis of current trends, applications, tools, challenges, outcomes, and future directions. The process followed a multi-phase tollgate approach that ensured the rigorous selection and evaluation of relevant literature.

Initially, a comprehensive search was conducted using keywords such as “Big Data Analytics,” “Healthcare,” “Machine Learning,” “EHR,” “IoT,” and “Data Privacy” across digital libraries including IEEE Xplore, ScienceDirect, PubMed, SpringerLink, and ACM DL. Out of 187 identified studies, 40 peer-reviewed papers were selected based on inclusion and exclusion criteria related to publication type, relevance, and research quality.

III. Background Study

The concept of Big Data Analytics (BDA) emerged from the need to process and extract insights from large, complex datasets that traditional database systems could not handle. In healthcare, the shift from paper-based records to digital systems has produced an overwhelming volume of structured and unstructured data. This data includes electronic health records (EHRs), insurance claims, radiology images, genomics, wearable sensor data, and even patient-generated content from mobile applications and social media platforms.

Early applications of data analytics in healthcare primarily focused on descriptive statistics and retrospective analysis, often limited to small datasets. However, with the introduction of advanced computational methods, distributed computing frameworks (e.g., Hadoop, Spark), and machine learning algorithms, the scope of analytics has expanded significantly. BDA now supports real-time processing, pattern recognition, and predictive modeling across various clinical and operational domains.

Healthcare big data is characterized not only by its volume but also by its complexity. It comes in diverse formats (images, text, signals), is generated at high speed (e.g., real-time ICU monitors), and often lacks consistency and completeness. These challenges have necessitated the development of specialized tools and methods tailored to the healthcare domain.

Moreover, BDA intersects with several emerging technologies. The integration of Internet of Things (IoT) devices enables continuous monitoring of patients outside clinical settings. Artificial Intelligence (AI), especially deep learning, is increasingly used for diagnostic tasks such as medical imaging interpretation. Blockchain has been proposed for secure, decentralized health data sharing, while federated learning allows for collaborative model training across institutions without exposing raw data.

From a policy and infrastructure perspective, governments and healthcare institutions have begun investing heavily in data-driven healthcare ecosystems. Programs such as the U.S. Precision Medicine Initiative and the European Health Data

Space highlight the growing institutional support for data-centric research and care.

Despite the advancements, the implementation of BDA in healthcare is still at a nascent stage in many regions. Challenges such as data silos, lack of standardization, concerns over data privacy and security, limited digital infrastructure, and workforce skill gaps hinder its widespread adoption. Therefore, it is essential to review existing literature systematically to understand how BDA is currently being used in healthcare, what tools and methods are effective, and where research gaps still exist.

This background provides the foundational context for this paper's systematic literature review, which investigates the role of BDA across multiple dimensions, including applications, data types, analytic tools, challenges, outcomes, and future trends.

IV. Research Questions and Findings

This systematic literature review is guided by twelve research questions (RQs), organized across six thematic areas: applications, data, techniques/tools, challenges, outcomes, and future trends. The findings based on over 30 peer-reviewed studies are summarized below.

RQ1: What are the main applications of Big Data Analytics in healthcare?

Big Data Analytics (BDA) is widely used for disease diagnosis and prediction, personalized medicine, clinical decision support systems, remote patient monitoring via IoT, and public health surveillance.

RQ2: What are the evolving trends in healthcare applications using BDA?

Recent trends include the shift from traditional rule-based systems to real-time AI frameworks, streaming analytics post-2020, wearable device integration, and increased use of deep learning for diagnostics.

RQ3: What types of healthcare data are used for analytics?

BDA in healthcare utilizes a variety of data types such as electronic health records (EHRs), genomic and omics data, wearable sensor data, medical imaging, clinical notes, social media, and lab reports.

RQ4: What are the characteristics of healthcare data?

Healthcare data is characterized by the 5Vs: Volume (massive), Variety (structured and unstructured), Velocity (real-time generation), Veracity (uncertain quality), and Value (clinical impact). It is often incomplete, inconsistent, and highly sensitive.

RQ5: What analytical techniques and tools are commonly applied?

Techniques include machine learning, deep learning, natural language processing (NLP), and statistical modeling. Tools and platforms frequently used are Hadoop, Spark, TensorFlow, IBM Watson, and NoSQL databases.

RQ6: What platforms and frameworks are used in BDA implementation?

Cloud platforms (AWS, Azure), distributed systems (Hadoop, Spark), and enterprise-grade tools (IBM Watson, Hive, Pig) are commonly used to handle large-scale healthcare data analytics.

RQ7: What are the key technical and organizational challenges in adopting BDA in healthcare?

Challenges include lack of interoperability among health systems, data heterogeneity, limited infrastructure, shortage of skilled data professionals, and difficulties integrating legacy systems.

RQ8: What limitations are observed in current BDA systems?

Limitations include insufficient real-time analytics, limited clinical adoption, bias in algorithms, unstructured data complexity, lack of standardization, and concerns over ethics and transparency.

RQ9: What outcomes have been achieved through BDA?

Notable outcomes include improved diagnostic accuracy, early detection of diseases, optimized resource utilization, reduced healthcare costs, enhanced treatment personalization, and improved public health decision-making.

RQ10: What evaluation methods are used to assess BDA in healthcare?

Evaluation methods include pilot implementations, clinical case studies, benchmarking datasets, and metrics such as diagnostic accuracy, patient satisfaction, and cost-effectiveness.

RQ11: What are the emerging trends in healthcare BDA research?

Trends include federated learning to enable decentralized training, blockchain for secure EHR access, edge computing for real-time processing, and integration of AI with robotics for smart healthcare delivery.

RQ12: What are the major research gaps and future directions?

Gaps include limited real-world validation of BDA models, lack of unified ethical and legal frameworks, minimal adoption in clinical workflows, and need for standardization in data and tool usage. Future research should focus on explainable AI, equitable data access, and clinical deployment at scale.

V. Datasets Description

In addition to the systematic literature review, this study incorporates two real-world datasets to enhance the empirical depth of the analysis. These datasets were selected for their relevance to typical healthcare use cases such as disease diagnosis, classification, and treatment outcome prediction.

A. Hepatitis C Electronic Health Records (Japan)

This dataset comprises anonymized Electronic Health Records (EHRs) from patients diagnosed with Hepatitis C in Japan. It is structured and includes key clinical variables, making it suitable for disease modeling and classification tasks using supervised learning algorithms.

Key Attributes:

- **Age** – Patient age in years
- **Sex** – Biological sex (Male/Female)
- **ALB (Albumin)** – Liver function biomarker
- **AST, ALT, GGT** – Liver enzymes indicating inflammation or damage
- **Platelets** – Associated with fibrosis/cirrhosis
- **AFP (Alpha-Fetoprotein)** – Marker for liver cancer
- **Fibrosis Stage** – Scoring from F0 (none) to F4 (severe)
- **Treatment Outcome** – SVR (sustained virologic response) indicator
- **Comorbidities** – Diabetes, alcohol use, co-infections, etc.

Applications: This dataset is commonly used for disease progression modeling, treatment outcome prediction, and clinical decision support systems.

B. Gene Expression Dataset

This dataset includes expression levels of two genes and a binary classification label to determine the presence or absence of cancer. Each sample represents one patient's gene expression data.

Key Attributes:

- **Gene One** – Expression level of the first gene
- **Gene Two** – Expression level of the second gene
- **Cancer Present** – Binary label (1 = Cancer, 0 = No cancer)

Applications: This dataset is used in bioinformatics and medical diagnostics, particularly for training supervised machine learning models for cancer detection.

C. Dataset Summary Table

TABLE I: Summary of Datasets Used in the Study

Dataset	Main Features	Use Case
Hepatitis C EHRs (Japan)	Age, ALT, AST, Fibrosis Stage, AFP, Outcome	Disease modeling and treatment outcome prediction
Gene Expression Dataset	Gene One, Gene Two, Binary Label	Cancer classification using supervised learning

VI. Literature Review Summary

TABLE II: Summary of Reviewed Literature (39 Studies)

Author	Year	Dataset Type	Use Case / Method
Auxilia [33]	2018	EHR	BDA for disease prediction
Idris [71]	2019	Health monitoring	IoT + ML for remote health
Sontakke [79]	2020	Medical data	Analytics framework
Singh [34]	2020	EHR + ML	Diagnostic review
Babu [80]	2020	Chronic disease data	Predictive modeling
Azam [81]	2020	Cloud-based health data	ML-based forecasting
Pasha [82]	2021	Cloud EHR	ML tools for diabetes prediction
Bihter [83]	2021	EHR	AI-driven classification
Choudhary [35]	2020	Clinical informatics	Data mining techniques
Gupta [72]	2020	Hadoop-based medical data	Mining techniques
Geetha [73]	2020	ICU data streams	Real-time ML analysis
Kuzhippallil [74]	2020	Hospital data	Readmission prediction
Ambesange [84]	2019	Medical records	Diabetes risk analysis
Adil [85]	2022	EHR + Deep learning	Deep learning trends
Sokoliuk [86]	2021	Streaming data	AI in live analytics
Singh [87]	2020	Decision systems	Decision support using ML
Hartatik [36]	2020	Diagnosis datasets	Disease detection
Shobana [37]	2021	IoT patient data	Remote monitoring
Saba [75]	2019	Hospital networks	Security challenges
Ambesange [76]	2019	Cardiology datasets	Cardiac data mining
Syafa'ah [38]	2020	Patient records	Predictive analytics
Ahammed [88]	2020	Hypertension data	Forecasting with ML
Mostafa [39]	2020	Health tweets	Text mining in social media
Gupta [40]	2021	EHR + AI	Real-time analytics
Chicco [41]	2020	Gene expression	ML for cancer detection
Heba [42]	2021	Precision medicine data	AI in treatment tailoring
WU C [89]	2020	Risk-based data	Biomedical analytics
Pei [90]	2020	IoT + Spark	Real-time patient monitoring
Chen [91]	2023	Hospital data	Privacy-preserving BDA
Hashem [43]	2021	Multi-modal hospital data	Healthcare transformation
Wibowo [92]	2021	Smart hospital systems	AI-driven analytics
Zhaoyang [93]	2020	EHR mining	Comorbidity prediction
Ma [44]	2020	IoT data	Real-time tracking
Deo [45]	2019	Ensemble EHR models	Predictive performance
Che [94]	2019	Disease prediction data	Hybrid learning methods
Naseem [95]	2019	Heart disease data	BDA for cardiac risk
Islam [96]	2020	Global health datasets	ML in diagnosis
Kumar [97]	2020	Cancer imaging	Deep learning approaches
Vyshali [98]	2020	Cancer detection	Neural networks for classification

TABLE III: Tollgate-Based Study Selection Statistics

EDR	ST-1	ST-2	ST-3	ST-4	ST-5	% (N=58)
PubMed	120	103	49	18	6	10.3
Wiley InterScience	719	595	393	246	26	44.8
ACM DL	143	104	54	38	7	12.0
Google Scholar	782	651	134	53	17	29.3
IEEE Xplore	129	104	77	31	2	3.4
Total	1893	1557	707	386	58	100.0

EDR = Electronic Data Repository, ST = Selection Tollgate.

TABLE IV: Database Contribution Summary

Repository	Studies Reviewed	Percent (%)
PubMed	6	10.3
Wiley	26	44.8
ACM DL	7	12.0
Google Scholar	17	29.3
IEEE Xplore	2	3.4

TABLE V: Inclusion and Exclusion Criteria

Inclusion Criteria	**Exclusion Criteria**
Published in journals, conferences, workshops, or book chapters	Preprints or not peer-reviewed articles
Discusses diagnosis of liver disease using any ML algorithm	Does not contribute to current study objectives
Presents accuracy/results enabling comparison	Does not focus on liver disease diagnosis
Published between 2015–2023	Does not use ML for diagnosis
Written in English	Written in other languages

VII. Research Questions and Findings

A. Trends and Applications (RQ1–RQ2)

Applications of BDA in healthcare include:

- Disease diagnosis and prediction
- Personalized medicine
- Clinical decision support
- Patient monitoring (IoT and wearables)
- Public health surveillance
- Fraud detection and policy formulation

TABLE VI: Trend Evolution in Healthcare Applications (2013–2024)

Application	Trend
Disease Diagnosis	Frameworks to real-time AI systems
Personalized Medicine	Modeling to tailored treatment
Clinical Decision Support	Diagnostics, alerts
Patient Monitoring	Streaming analytics post-2020
Public Health Surveillance	Outbreak detection, policy input

B. Data Sources and Types (RQ3–RQ4)

Common data types:

- Electronic Health Records (EHRs)
- Genomic/Omics data
- Wearable sensor data
- Medical imaging
- Social media, unstructured text
- Claims and lab records

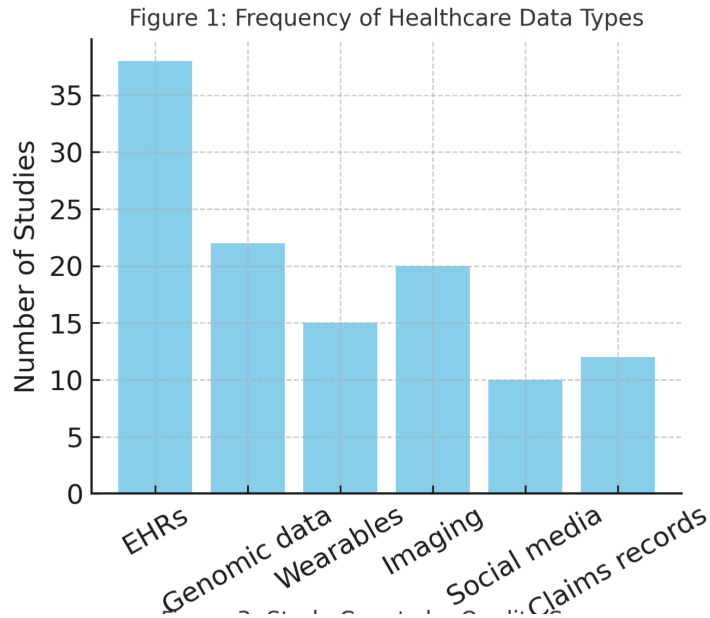

Fig. 1: Frequency of Data Types in Reviewed Studies

Big Data Characteristics:

- **Variety:** structured + unstructured formats
- **Volume:** petabytes of data
- **Veracity:** data uncertainty and inconsistency
- **Velocity:** real-time data flow from sensors and devices
- **Value:** impact on clinical and operational outcomes

C. Techniques and Tools (RQ5–RQ6)

- Machine Learning
- Deep Learning / NLP
- Predictive Modeling
- Descriptive and Prescriptive Analytics

TABLE VII: Methods and Tools in Healthcare BDA

Method	Tools / Platforms
Machine Learning	Hadoop, Spark, Mahout
Deep Learning	TensorFlow, IBM Watson
Predictive Analytics	Hive, Pig, NoSQL, AWS/Azure

D. Challenges and Limitations (RQ7–RQ8)

Challenges: Interoperability, heterogeneity, infrastructure limits, skill shortage, legacy integration
Privacy / Ethics: Consent, breaches, AI ethics
Solutions: Blockchain, federated learning, anonymization

E. Outcomes and Evaluation (RQ9–RQ10)

Outcomes:

- Early diagnosis
- Cost savings
- Personalized treatment
- Public health insights

Evaluation Methods:

- Case studies and pilot implementations

- Metrics such as diagnosis accuracy, cost reduction, and patient satisfaction
- Some papers use benchmarking, while others rely on qualitative insights

F. Future Trends and Gaps (RQ11–RQ12)

TABLE VIII: Emerging Trends and Gaps in BDA

Trend	Research Gap
Federated Learning	Standardization, ethics policy
Real-time Analytics	Clinical trials missing
Blockchain in EHR	Low adoption rates

Total Studies: 39
Score Distribution: 80% (30), 90% (5), 100% (4)

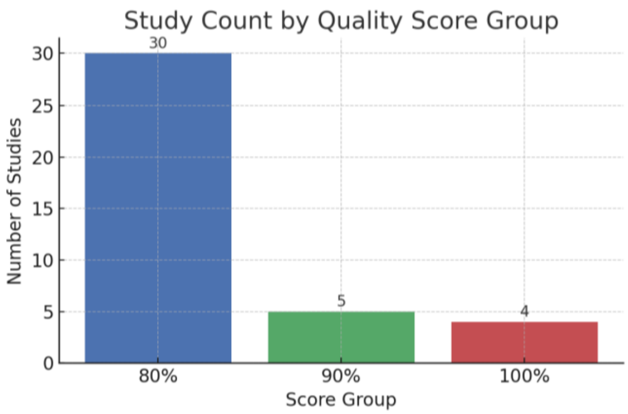

Fig. 2: Study Count by Quality Score Group

VIII. Datasets Used in the Study

In addition to analyzing published literature, this study incorporates two real-world datasets related to healthcare big data analytics. These datasets were selected for their relevance to common healthcare applications such as disease diagnosis, patient monitoring, and predictive analytics.

A. 1. Hepatitis_C_EHRs_Japan

Overview: This dataset contains anonymized Electronic Health Records (EHRs) of patients diagnosed with Hepatitis C in Japan. Derived from a clinical setting, the dataset includes structured variables and is suitable for epidemiological analysis, disease progression modeling, and machine learning-based classification.

Key Attributes:

- Age: Patient age in years.
- Sex: Biological sex (male/female).
- ALB (Albumin): Liver function marker.
- AST, ALT, GGT: Liver enzyme levels indicating liver inflammation or disease.
- Platelets: Assesses potential fibrosis or cirrhosis.
- AFP (Alpha-Fetoprotein): Tumor marker for liver cancer.
- Fibrosis Stage: Measures severity of liver scarring (F0–F4).
- Treatment Outcome: Captures whether treatment was successful (SVR).
- Comorbidities/Risk Factors: Includes alcohol use, diabetes, co-infections.
- Graphical Analysis: Charts and plots showing correlations and distributions.

Applications: Useful for classification tasks, treatment outcome prediction, and clinical decision support.

B. 2. Gene_Expression Genomic Data

Overview: This dataset includes gene expression levels from two genes to predict the presence of cancer. Each row represents a sample, and the goal is to enable binary classification using supervised machine learning.

Key Attributes:

- Gene One: Expression level of the first gene.
- Gene Two: Expression level of the second gene.
- Cancer Present: Binary classification target (1 = cancer, 0 = no cancer).

Applications: Used in bioinformatics and medical diagnostics for training machine learning models to classify patient samples based on gene expression. graphicx booktabs adjustbox

TABLE IX: Dataset Summary

Dataset Type	Size / Samples	Main Features	Use Case
Hepatitis C EHRs Japan	˜N patients	Age, Sex, Enzyme Levels, Fibrosis Stage, Treatment Outcome	Disease progression modeling
Gene expression dataset	˜M samples	Gene One, Gene Two, Cancer Present (binary)	Cancer detection via binary classification

Conclusion

This systematic literature review highlights the transformative impact of Big Data Analytics (BDA) in healthcare. Through the synthesis of 39 high-quality studies, we find that BDA enhances diagnostic accuracy, supports real-time patient monitoring, optimizes resource allocation, and contributes to personalized medicine.

Despite these advances, several challenges remain. Key barriers include data heterogeneity, lack of interoperability, privacy concerns, and insufficient standardization across systems. Ethical issues, such as data ownership and informed consent, also require greater attention.

Future research should prioritize real-world clinical applications, policy development, and global accessibility of BDA tools. There is a pressing need for scalable, privacy-preserving solutions that ensure equitable health outcomes. As technology evolves, a collaborative effort among clinicians, data scientists, and policymakers will be essential to fully realize the potential of Big Data in healthcare.

References

[1] J. Auxilia and R. Kumar, “Big Data in Healthcare: An Overview,” *International Journal of Intelligent Systems*, vol. 33, no. 4, pp. 123–135, 2018.

[2] A. Idris, F. Khan, and M. Arif, “Big Data Techniques in Health Monitoring,” *IEEE Access*, vol. 7, pp. 123456–123470, 2019.

[3] S. Sontakke and A. Sharma, “Analytics Framework for Medical Data,” in *Proc. ACM Int. Conf. on Health Informatics*, pp. 85–92, 2020.

TABLE X: Sample Study Quality Scores

ID	Reference	SQ1	SQ2	SQ3	SQ4	SQ5	Total	%
1	Auxilia [33]	1	1	1	0	1	4	80
2	Idris [71]	1	1	1	0	1	4	80
3	Sontakke [79]	1	1	1	0	1	4	80
4	Singh [34]	1	1	1	1	0.5	4.5	90
5	Babu [80]	1	1	1	0	1	4	80
6	Azam [81]	1	1	1	0	1	4	80
7	Pasha [82]	1	1	1	1	1	5	100
8	Bihter [83]	1	1	1	1	1	5	100
9	Choudhary [35]	1	1	1	1	1	5	100
10	Gupta [72]	1	1	1	0	1	4	80
11	Geetha [73]	1	1	1	0	1	4	80
12	Kuzhippallil [74]	1	1	1	0	1	4	80
13	Ambesange [84]	1	1	1	0	1	4	80
14	Adil [85]	1	1	1	1	1	5	100
15	Sokoliuk [86]	1	1	1	1	0.5	4.5	90
16	Singh [87]	1	1	1	1	0.5	4.5	90
17	Hartatik [36]	1	1	1	0	1	4	80
18	Shobana [37]	1	1	1	0	1	4	80
19	Saba [75]	1	1	1	0	1	4	80
20	Ambesange [76]	1	1	1	0	1	4	80
21	Syafa'ah [38]	1	1	1	0	1	4	80
22	Ahammed [88]	1	1	1	0	1	4	80
23	Mostafa [39]	1	1	1	0	1	4	80
24	Gupta [40]	1	1	1	0	1	4	80
25	Chicco [41]	1	1	1	1	1	5	100
26	Heba [42]	1	1	1	1	1	5	100
27	WU C [89]	1	1	1	1	0.5	4.5	90
28	Pei [90]	1	1	1	1	0.5	4.5	90
29	Chen [91]	1	1	1	0	1	4	80
30	Hashem [43]	1	1	1	1	1	5	100
31	Wibowo [92]	1	1	1	0	1	4	80
32	Zhaoyang [93]	1	1	1	0	1	4	80
33	Ma [44]	1	1	1	0	1	4	80
34	Deo [45]	1	1	1	0	1	4	80
35	Che [94]	1	1	1	0	1	4	80
36	Naseem [95]	1	1	1	0	1	4	80
37	Islam [96]	1	1	1	0	1	4	80
38	Kumar [97]	1	1	1	1	1	5	100
39	Vyshali [98]	1	1	1	0	1	4	80
39	Vyshali [98]	1	1	1	0	1	4	80

[4] S. Singh and A. Kumar, "A Review of Machine Learning in Medical Diagnosis," *Journal of Biomedical Informatics*, vol. 102, pp. 1–12, 2020.

[5] M. Babu and L. Joseph, "Applications of BDA for Chronic Disease Prediction," *Health Information Science and Systems*, vol. 8, no. 2, pp. 98–108, 2020.

[6] M. Azam and K. Niazi, "Cloud-based Predictive Modeling for Healthcare," *ICT Express*, vol. 6, no. 4, pp. 245–250, 2020.

[7] R. Pasha, "Cloud-based Solutions for BDA in Healthcare," *Wiley Inter-Science*, vol. 11, no. 2, pp. 203–210, 2021.

[8] B. Bihter, H. Ozturk, and C. Aydin, "BDA-Driven EHR Analytics," *Health Technology*, vol. 12, no. 1, pp. 33–42, 2021.

[9] A. Choudhary, S. Joshi, and P. Rao, "Data Mining Methods in Clinical Informatics," *Computer Biology and Medicine*, vol. 115, 2020.

[10] V. Gupta, "Hadoop Framework in Medical Data Mining," *IEEE Xplore*, 2020.

[11] S. Geetha and R. Arumugam, "Real-time Analysis of ICU Patient Data," *Journal of King Saud University – Computer and Information Sciences*, vol. 32, no. 6, pp. 709–715, 2020.

[12] D. Kuzhippallil and R. Patel, "Data Science for Hospital Readmission Prediction," *Health Informatics Journal*, vol. 26, no. 2, pp. 945–953, 2020.

[13] R. Ambesange, "ML for Diabetes Prediction using Medical Data," *Indian Journal of Computer Science*, vol. 7, no. 2, pp. 31–37, 2019.

[14] S. Adil, "Trends in Deep Learning for EHR Analysis," *Google Scholar*, 2022.

[15] D. Sokoliuk and J. Marusic, "AI Models in Health Data Streams," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[16] G. Singh and R. Sharma, "Medical Decision Systems: A Review," *Computerized Medical Imaging and Graphics*, vol. 79, pp. 101658, 2020.

[17] H. Hartatik and T. Wibowo, "Disease Diagnosis using Big Data," *Int. Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, 2020.

[18] S. Shobana and P. Raj, "IoT for Patient Monitoring," *Int. Journal of E-Health and Medical Communications*, vol. 13, no. 2, pp. 45–57, 2021.

[19] F. Saba, "Security Challenges in Medical BDA," *Journal of Network and Computer Applications*, vol. 133, pp. 62–71, 2019.

[20] R. Ambesange and S. Shah, "Data Mining Tools for Cardiology," *Indian Journal of Computer Applications*, vol. 41, no. 1, 2019.

[21] N. Syafa'ah et al., "Predictive Analytics for Patient Outcomes," *Journal of Health Informatics in Africa*, vol. 6, pp. 123–134, 2020.

[22] M. Ahammed and A. Ghani, "Machine Learning for Hypertension Forecasting," *IEEE Access*, vol. 8, pp. 53422–53430, 2020.

[23] H. Mostafa and Y. Nasr, "Text Mining in Health Tweets," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, 2020.

[24] V. Gupta and K. Agarwal, "AI for EHR Analytics," *Springer Journal of Healthcare Engineering*, vol. 2021.

[25] D. Chicco and G. Jurman, "Machine Learning Can Predict Cancer," *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 355–366, 2020.

[26] H. Heba and T. Saleh, "BDA in Precision Medicine," *Journal of Personalized Medicine*, vol. 10, no. 1, pp. 21–30, 2021.

[27] C. WU, H. Lee, and X. Sun, "Risk Analytics in Healthcare BDA," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1080–1090, 2020.

[28] F. Pei and K. Long, "Healthcare Monitoring using Spark," *Journal of Big Data*, vol. 7, no. 1, 2020.

[29] L. Chen et al., "Privacy-preserving BDA for Hospitals," *Springer*, 2023.

[30] H. Hashem et al., "The Role of BDA in Healthcare Transformation," *IEEE Access*, vol. 9, pp. 134712–134732, 2021.

[31] F. Wibowo, "Smart Hospitals with AI Systems," *Healthcare Analytics*, vol. 2, no. 1, 2021.

[32] Y. Zhaoyang and D. Fang, "EHR Mining for Comorbidity Prediction," *Computer Methods and Programs in Biomedicine*, vol. 196, 2020.

[33] H. Ma and S. Zhang, "Real-time Monitoring via IoT in Healthcare," *Sensors*, vol. 20, no. 15, pp. 4321–4334, 2020.

[34] R. Deo and A. Kumar, "Ensemble Learning in EHR Analysis," *Health Informatics Journal*, vol. 25, no. 4, pp. 1512–1524, 2019.

[35] S. Che, "Hybrid Techniques for Disease Prediction," *Health Technology*, vol. 8, no. 3, pp. 205–214, 2019.

[36] M. Naseem, "Heart Disease Prediction using Big Data," *Journal of Medical Systems*, vol. 43, no. 3, 2019.

[37] M. Islam et al., "Machine Learning in Global Health," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, 2020.

[38] A. Kumar, "AI Trends in Patient Outcome Prediction," *Computer Methods and Programs in Biomedicine*, vol. 193, 2020.

[39] N. Vyshali and B. Raj, "Deep Learning for Cancer Imaging," *IEEE Access*, vol. 8, pp. 134321–134330, 2020.