

Preparing the Data-Driven Healthcare Workforce: A PRISMA-Guided Review of Big Data Analytics and Educational Implications

Muhammad Sultan, Royal College of Surgeons in Ireland–Medical University of Bahrain,
Bahrain

The Paris Conference on Education 2026
Official Conference Proceedings

Abstract

The rapid expansion of Electronic Health Records has accelerated the integration of Big Data Analytics (BDA) and Artificial Intelligence (AI) into healthcare systems, transforming clinical decision-making and personalised healthcare. However, the increasing reliance on advanced analytics introduces substantial challenges related to data governance, cybersecurity, ethical accountability, and workforce preparedness. This review is guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework and structured using the Population, Intervention, Comparator, Outcome, Study Design, and Setting (PICOSS) framework, synthesising peer-reviewed literature published between 2013 and 2023 to examine the benefits, technologies, and challenges associated with BDA in healthcare. Out of the 2,969 retrieved studies, 35 met the inclusion criteria. Findings indicate that BDA enhances predictive capabilities, personalised healthcare delivery, and operational efficiency. Persistent barriers include data governance and quality management concerns, cybersecurity risks, and ethical challenges related to algorithmic bias, accountability, privacy, and limited transparency. Advanced technologies, including machine learning, deep learning, and federated learning, are central to implementation but require specialised competencies. These findings highlight important implications for education, including the need to strengthen AI literacy, ethical governance of algorithms, and interdisciplinary training to support the responsible adoption of data-driven technologies. These findings highlight the importance of embedding data governance, cybersecurity awareness, and ethical AI frameworks into higher education curricula to prepare future professionals for responsible, sustainable, data-driven practices.

Keywords: big data analytics, higher education, AI literacy, workforce readiness, responsible AI

iafor

The International Academic Forum
www.iafor.org

Introduction

The rapid digitalisation of healthcare data has resulted in unprecedented growth in the volume, variety, and complexity of healthcare data. Electronic health records (EHRs), medical imaging, genomic information, wearable technologies, and Internet of Things (IoT) devices continuously generate structured, semi-structured, and unstructured datasets that are increasingly difficult to manage using conventional analytical approaches (Chawla & Davis, 2013; Raghupathi & Raghupathi, 2014). The widespread adoption of Big Data Analytics (BDA), supported by advances in artificial intelligence (AI) and machine learning (ML), has created new opportunities to transform healthcare delivery through data-driven decision-making, predictive modelling, personalised healthcare, and operational improvements (Awrahanman et al., 2022; Jiang et al., 2017).

The integration of AI-enabled analytics into healthcare has accelerated in recent years, particularly following the COVID-19 pandemic, where large-scale data analysis supported disease surveillance, outbreak monitoring, clinical decision-making, and healthcare planning (Arora et al., 2021; Firouzi et al., 2021; Vaishya et al., 2020). At the same time, advances in natural language processing, deep learning, federated learning, and data mining have expanded healthcare organisations' ability to extract meaningful insights from increasingly complex datasets (J. S. Chen & Baxter, 2022; Yang, Soltan, Eyre, Yang, et al., 2023). These developments position BDA as a key enabler of precision medicine and data-driven healthcare.

Despite these opportunities, the implementation of BDA presents significant technical, organisational, and ethical challenges. Healthcare organisations must address issues relating to data quality, interoperability, governance, cybersecurity, patient privacy, and algorithmic bias while ensuring that AI systems remain transparent, trustworthy, and aligned with ethical principles (Awrahanman et al., 2022; Garattini et al., 2019; Norori et al., 2021). The responsible adoption of AI, therefore, depends not only on technological capability but also on effective governance and an appropriately prepared workforce that can critically evaluate and apply data-driven technologies in clinical practice.

Although the technological applications of BDA in healthcare have been widely explored, comparatively less attention has been given to their implications for preparing the future workforce. As AI-enabled technologies become increasingly embedded within systems, higher education institutions play a critical role in equipping graduates with competencies that extend beyond technical proficiency to include AI literacy, data governance, cybersecurity, ethical AI, and interdisciplinary collaboration. These capabilities are essential for supporting the responsible implementation of BDA and fostering confidence in data-informed clinical decision-making.

This study presents a PRISMA-guided review of the literature published between 2013 and 2023 to examine the benefits, enabling technologies, and challenges associated with BDA in healthcare. Building on this evidence, the review considers the educational implications for preparing a future-ready workforce to support the safe, ethical, and effective adoption of AI-enabled, data-driven healthcare.

Literature Review

Big Data and its analysis have become increasingly important in healthcare due to the rapid expansion of electronic health records, medical imaging, genomic data, and other structured,

semi-structured, and unstructured datasets. Across the 35 studies included in the review, BDA was most frequently associated with three areas of benefit: predictive capabilities, personalised healthcare, and cost efficiency.

The review identifies predictive capabilities and personalised healthcare as the two most frequently reported benefits, each appearing in 11 of the 35 included studies, representing 31% of the reviewed evidence. Cost reduction appeared in four studies, representing 11% of the included literature. The literature suggests that BDA can support predictive healthcare by identifying patterns, trends, risks, and potential outcomes and by analysing large datasets. Studies reported that BDA has been used to monitor infectious disease outbreaks, assess treatment risks, predict patient readmission, support public health interventions, and improve clinical decision-making (Arora et al., 2021; Awrahman et al., 2022; Chawla & Davis, 2013; J. S. Chen & Baxter, 2022; Firouzi et al., 2021; Garattini et al., 2019; Jiang et al., 2017; Majnarić et al., 2021; Tenali & Babu, 2023; Vaishya et al., 2020; Xiong et al., 2022). During the COVID-19 pandemic, BDA, machine learning, IoT, and AI-enabled tools were used to support surveillance, diagnosis, hotspot identification, disease monitoring, vaccine development, and treatment planning (Arora et al., 2021; Firouzi et al., 2021; Vaishya et al., 2020).

BDA enables the aggregation and analysis of electronic health records, genomic data, imaging data, and real-time biometric information to support individualised risk profiling, precision medicine, and tailored treatment strategies (Chawla & Davis, 2013; Jagadeeswari et al., 2018; Raghupathi & Raghupathi, 2014; Viceconti et al., 2015). Studies have also linked BDA to patient-centred care, earlier identification of disease risk, targeted therapies, and improved treatment planning (Cheung et al., 2019; Grossglauser & Saner, 2014; Lemmen et al., 2021; Murphy & Saleh, 2020; Norori et al., 2021). The dissertation also identifies the use of technologies such as Hadoop, Spark, and Virtual Physiological Human (VPH) models to support complex data processing for personalised and in silico medicine solutions (Viceconti et al., 2015; Xiong et al., 2022). Another group of studies linked BDA to operational efficiency and cost reduction. These studies suggest that BDA may support earlier detection of health risks, remote monitoring, chronic disease management, and improved resource allocation (Grossglauser & Saner, 2014; Jagadeeswari et al., 2018; Raghupathi & Raghupathi, 2014; Xiong et al., 2022).

The review identifies machine learning and artificial intelligence as the most frequently discussed enabling technologies. Machine learning and AI were reported as central to extracting insights, identifying patterns, making predictions, and supporting healthcare decision-making (Arora et al., 2021; Babel et al., 2021; Benke & Benke, 2018; J. S. Chen & Baxter, 2022; S. W. Chen et al., 2021; D'Costa & Zatale, 2021; Golas et al., 2018; Hulsén, 2022; Jiang et al., 2017; Majnarić et al., 2021; Murphy & Saleh, 2020; Tenali & Babu, 2023; Yang, Soltan, Eyre, Yang, et al., 2023). Natural language processing (NLP) was identified as particularly important for converting unstructured data into analysable data, supporting medication adherence, ophthalmology, clinical documentation, and decision-making (Babel et al., 2021; J. S. Chen & Baxter, 2022; D'Costa & Zatale, 2021; Hulsén, 2022; Jiang et al., 2017; Majnarić et al., 2021).

Deep learning, neural networks, convolutional neural networks, and artificial neural networks were reported as useful for analysing complex medical images, detecting conditions such as diabetic retinopathy, and identifying patterns within large datasets (Arora et al., 2021; Benke & Benke, 2018; J. S. Chen & Baxter, 2022; Golas et al., 2018; Jiang et al., 2017; Murphy &

Saleh, 2020). Federated learning and reinforcement learning were also identified as emerging approaches, particularly for privacy-preserving predictive modelling and algorithmic fairness in clinical machine learning (Yang, Soltan, Eyre, Yang, et al., 2023).

Data mining was presented as a method for extracting patterns and relationships from large datasets. It was associated with patient data analysis, risk identification, pharmacovigilance, decision support, and the detection of links between medications and side effects (Benke & Benke, 2018; Grossglauser & Saner, 2014; Raghupathi & Raghupathi, 2014; Ross et al., 2014; Tenali & Babu, 2023; Xiong et al., 2022). Hadoop was identified as a framework for processing large-scale data, supporting scalability, computational efficiency, and the management of high-volume datasets (Hammad et al., 2020; Tenali & Babu, 2023; Xiong et al., 2022; Zhou et al., 2020).

The review identifies three major challenge areas: data quality and management, security, and ethical concerns. Data quality and management appeared in 17 studies, representing 48% of the included literature. Security concerns were reported in 14 studies (40%), and ethical concerns in 10 (29%). Data quality and management challenges included data standardisation, accuracy, ownership, governance, unstructured data, interoperability, and the difficulty of managing data across multiple platforms and formats (Akhtar et al., 2020; Awrahman et al., 2022; Babel et al., 2021; Benke & Benke, 2018; Chawla & Davis, 2013; Cheung et al., 2019; D'Costa & Zatale, 2021; Del Rio-Bermudez et al., 2020; Garattini et al., 2019; Grossglauser & Saner, 2014; Hammad et al., 2020; Hulsen, 2022; Jiang et al., 2017; Majnarić et al., 2021; Murphy & Saleh, 2020; Tenali & Babu, 2023; Viceconti et al., 2015). These challenges are important to address because poor-quality or poorly governed data can undermine the reliability of BDA and AI-enabled decision-making.

Security was another major concern. The reviewed studies highlighted risks related to unauthorised access, privacy breaches, data ownership, cloud computing, cyberattacks, and the protection of sensitive medical records (Awrahman et al., 2022; Babel et al., 2021; Benke & Benke, 2018; Chawla & Davis, 2013; Grossglauser & Saner, 2014; Hammad et al., 2020; Hulsen, 2022; Jagadeeswari et al., 2018; Lemmen et al., 2021; Murphy & Saleh, 2020; Raghupathi & Raghupathi, 2014; Ross et al., 2014; Xiong et al., 2022; Zhou et al., 2020). These findings indicate that cybersecurity awareness and data protection should be treated as core competencies for professionals working in data-driven healthcare environments.

Ethical concerns included privacy, consent, autonomy, bias, accountability, transparency, algorithmic fairness, and the use of black-box AI systems (Arora et al., 2021; Di Nucci, 2019; Dove et al., 2022; Garattini et al., 2019; Hulsen, 2022; Lemmen et al., 2021; Murphy & Saleh, 2020; Norori et al., 2021; Yang, Soltan, Eyre, Yang, et al., 2023). The reviewed literature raises particular concern about bias in AI and machine learning systems, especially when underrepresented groups are inadequately represented in training datasets. Explainable AI is therefore presented as an important response to concerns about transparency and trust, although the dissertation notes that highly complex models may remain difficult to interpret fully. Although the review focuses primarily on BDA in healthcare, its findings clearly point to an educational gap. The reviewed literature demonstrates that responsible use of BDA depends not only on technical tools but also on workforce capabilities in AI literacy, data governance, cybersecurity, ethical reasoning, and interdisciplinary collaboration.

Methodology

This review employed a PRISMA-guided rapid review to synthesise evidence on the application of BDA in healthcare and its educational implications for workforce preparedness. Several evidence synthesis approaches, including systematic, narrative, rapid, and scoping reviews, were considered during the study design. A rapid review was selected because it provides a structured and transparent synthesis of evidence within a shorter timeframe while maintaining methodological rigour, making it appropriate for addressing emerging topics such as BDA and artificial intelligence (AI) in healthcare (Boland, Angela; Cherry, M. Gemma; Dickson, 2014). Reporting followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses to ensure transparency, reproducibility, and consistency throughout the review process (Page et al., 2021).

Eligibility Criteria

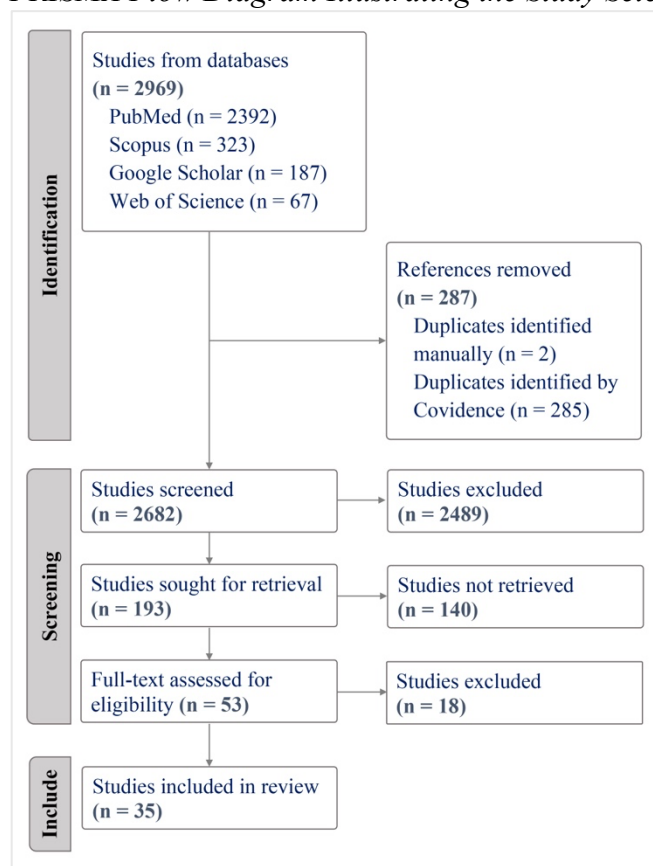
Study selection was guided by the Population, Intervention, Comparator, Outcomes, Study Design and Setting (PICOSS) framework, an extension of the Population, Intervention, Comparator and Outcomes (PICO) framework commonly used in evidence-based research (Boland, Angela; Cherry, M. Gemma; Dickson, 2014). The PICOSS framework was used to establish explicit inclusion and exclusion criteria before commencing the literature search.

Eligible studies included patients, healthcare providers, data scientists, machine learning researchers, and ethicists. The intervention focused on the application of Big Data, machine learning, and artificial intelligence in healthcare. Comparators included traditional healthcare approaches and non-optimised analytical methods. Eligible outcomes included improvements in patient experience, personalised treatment, personalised medication, ethical practice, and health insights. Studies published in English between 2013 and 2023 were included to reflect the period during which BDA experienced rapid development within healthcare. Eligible settings comprised hospitals, clinics, healthcare organisations, research laboratories, universities, and healthcare research institutions. Studies conducted outside healthcare settings, studies published before 2013, and studies unrelated to BDA, AI, or machine learning were excluded.

Search Strategy and Study Selection

A comprehensive literature search was undertaken using four electronic databases: PubMed, Scopus, Google Scholar, and Web of Science. PubMed yielded the largest number of potentially eligible studies. Database searches combined Boolean operators and keywords relating to patients, healthcare, Big Data Analytics, Artificial Intelligence, machine learning, personalised healthcare, and ethics. Where appropriate, the American spelling *personalized* was used because it generated a broader range of search results than the British spelling *personalised*.

The search identified 2,969 records, which were then imported into Covidence, where titles and abstracts were screened using the predefined PICOSS eligibility criteria (*Covidence – Systematic Review Management*, n.d.). Following initial screening, 193 studies were exported to EndNote 21 for full-text assessment (EndNote 21, n.d.). Fifty-three full-text articles were assessed for eligibility, of which 35 studies satisfied the inclusion criteria and were included in the final review. The study selection process is presented using a PRISMA 2020 flow diagram (Figure 1).

Figure 1*PRISMA Flow Diagram Illustrating the Study Selection Process*

Data Extraction and Thematic Synthesis

Data extraction was undertaken using a literature review matrix developed in Microsoft Excel, following the Matrix Method described by Goldman & Schmalz (2004). Each of the 35 included studies was reviewed in full, and key information was extracted, including author, publication year, study methodology, principal findings, and relevance to the research questions. The extracted data were systematically organised to facilitate comparison across studies and to ensure consistency during synthesis.

Recurring concepts and keywords identified within the literature were subsequently grouped into broader thematic categories. Themes were developed iteratively according to the three research questions, examining (1) the benefits of BDA in healthcare, (2) technologies supporting BDA implementation, and (3) challenges associated with BDA adoption. The final thematic framework comprised three benefit themes: predictive capabilities, personalised healthcare, and operational efficiency and cost reduction; three technology themes: machine learning and artificial intelligence, data mining, and Hadoop; and three challenge themes: data quality and governance, security, and ethical concerns. This structured thematic synthesis enabled a transparent and systematic interpretation of the evidence while maintaining direct alignment with the objectives of the review.

Characteristics of Included Studies

The final review included 35 English-language studies published between 2013 and 2023, comprising literature reviews, narrative reviews, rapid reviews, qualitative studies, quantitative

studies, and mixed-methods research. Of the 35 included studies, 30 were peer-reviewed, and 26 were published in journals with impact factors ranging from 1.4 to 23.8. The methodological characteristics of the included studies provided a broad evidence base for the thematic synthesis.

Results

The findings were organised into three themes corresponding to the research questions: (1) benefits of Big Data Analytics (BDA), (2) enabling technologies, and (3) implementation challenges.

Benefits

Three principal benefit themes emerged: predictive capabilities, personalised healthcare, and operational efficiency.

Predictive capabilities were identified in 11 of the 35 included studies (31%). These studies reported using BDA to identify disease trends, support timely interventions, predict patient outcomes, improve clinical decision-making, monitor infectious disease outbreaks, and enhance public health surveillance. During the COVID-19 pandemic, BDA and machine learning were widely applied to disease monitoring, hotspot identification, diagnosis, and healthcare planning (Arora et al., 2021; Awrahman et al., 2022; Chawla & Davis, 2013; J. S. Chen & Baxter, 2022; Firouzi et al., 2021; Garattini et al., 2019; Jiang et al., 2017; Majnarić et al., 2021; Tenali & Babu, 2023; Vaishya et al., 2020; Xiong et al., 2022).

Personalised healthcare was also reported in 11 studies (31%). The included literature demonstrated that BDA supports personalised treatment planning by integrating electronic health records, genomic information, medical imaging, and real-time patient data. Applications included disease risk assessment, precision medicine, targeted interventions, and patient-centred treatment strategies (Awrahman et al., 2022; Chawla & Davis, 2013; Cheung et al., 2019; Grossglauser & Saner, 2014; Jagadeeswari et al., 2018; Lemmen et al., 2021; Murphy & Saleh, 2020; Norori et al., 2021; Raghupathi & Raghupathi, 2014; Viceconti et al., 2015; Xiong et al., 2022).

Operational efficiency was identified in 4 studies (11%), where BDA was associated with improved resource allocation, remote patient monitoring, chronic disease management, earlier detection of health conditions, and reductions in healthcare expenditure through more efficient treatment pathways (Grossglauser & Saner, 2014; Jagadeeswari et al., 2018; Raghupathi & Raghupathi, 2014; Xiong et al., 2022).

Technologies

Machine learning and artificial intelligence were the most frequently reported enabling technologies, appearing in 13 studies (37%). The reviewed literature identified applications including natural language processing, deep learning, reinforcement learning, federated learning, deep neural networks, convolutional neural networks, and artificial neural networks for predictive modelling, clinical decision support, medical image analysis, disease detection, and pattern recognition (Arora et al., 2021; Babel et al., 2021; Benke & Benke, 2018; J. S. Chen & Baxter, 2022; S. W. Chen et al., 2021; D'Costa & Zatale, 2021; Golas et al., 2018; Hulsen,

2022; Jiang et al., 2017; Majnarić et al., 2021; Murphy & Saleh, 2020; Tenali & Babu, 2023; Yang, Soltan, Eyre, Yang, et al., 2023).

Data mining was reported in 6 studies (17%). It was used to identify patterns within healthcare datasets, support pharmacovigilance, improve clinical decision-making, and facilitate predictive analytics through AI-based algorithms and distributed computing platforms (Benke & Benke, 2018; Grossglauser & Saner, 2014; Raghupathi & Raghupathi, 2014; Ross et al., 2014; Tenali & Babu, 2023; Xiong et al., 2022).

Hadoop appeared in 4 studies (11%) as a scalable framework for managing high-volume healthcare datasets, supporting distributed storage, computational efficiency, and large-scale data processing (Xiong et al., 2022; Tenali & Babu, 2023; Hammad et al., 2020; Zhou et al., 2020).

Challenges

Three major themes of challenge emerged from the review: data quality and governance, security, and ethical concerns.

Data quality and governance represented the most frequently reported challenge, appearing in 17 studies (48%). Reported issues included data quality, ownership, interoperability, standardisation, governance, storage, and the management of structured and unstructured healthcare data (Akhtar et al., 2020; Awrahman et al., 2022; Babel et al., 2021; Benke & Benke, 2018; Chawla & Davis, 2013; Cheung et al., 2019; D'Costa & Zatale, 2021; Del Rio-Bermudez et al., 2020; Garattini et al., 2019; Grossglauser & Saner, 2014; Hammad et al., 2020; Hulslen, 2022; Jiang et al., 2017; Majnarić et al., 2021; Murphy & Saleh, 2020; Tenali & Babu, 2023; Viceconti et al., 2015).

Security concerns were identified in 14 studies (40%), with reported risks including unauthorised access, cybersecurity threats, privacy protection, cloud computing, confidential patient information, and safeguarding electronic health records against data (Awrahman et al., 2022; Babel et al., 2021; Benke & Benke, 2018; Chawla & Davis, 2013; Grossglauser & Saner, 2014; Hammad et al., 2020; Hulslen, 2022; Jagadeeswari et al., 2018; Lemmen et al., 2021; Murphy & Saleh, 2020; Raghupathi & Raghupathi, 2014; Ross et al., 2014; Xiong et al., 2022; Zhou et al., 2020).

Ethical concerns were reported in 10 studies (29%). The included studies identified issues relating to patient privacy, informed consent, algorithmic bias, fairness, accountability, transparency, and the use of “black box” AI systems. Several studies also highlighted Explainable Artificial Intelligence (XAI) as a potential approach to improve transparency and trust in AI-assisted healthcare (Arora et al., 2021; Di Nucci, 2019; Dove et al., 2022; Garattini et al., 2019; Hulslen, 2022; Lemmen et al., 2021; Murphy & Saleh, 2020; Norori et al., 2021; Yang, Soltan, Eyre, & Clifton, 2023; Yang, Soltan, Eyre, Yang, et al., 2023).

Discussion

The findings demonstrate that Big Data Analytics (BDA) has become an important component of contemporary healthcare, supporting predictive capabilities, personalised healthcare, and improvements in operational efficiency. Across the reviewed studies, predictive analytics consistently emerged as one of the most frequently reported benefits, reflecting the growing

use of BDA to identify disease patterns, support clinical decision-making, monitor infectious disease outbreaks, and enable timely healthcare interventions (Arora et al., 2021; Garattini et al., 2019; Majnarić et al., 2021). Likewise, the integration of electronic health records, genomic information, and real-time patient data has enabled more personalised approaches to diagnosis and treatment, reinforcing the transition towards precision medicine and patient-centred care (Chawla & Davis, 2013; Jagadeeswari et al., 2018; Raghupathi & Raghupathi, 2014). Although operational efficiency was identified less frequently, the evidence suggests that BDA can contribute to more effective resource allocation, earlier intervention, and reduced healthcare costs through data-informed decision-making (Grossglauser & Saner, 2014; Raghupathi & Raghupathi, 2014). These findings collectively demonstrate the growing strategic importance of data-driven approaches within healthcare systems.

The review also highlights that these benefits depend on the effective application of enabling technologies, particularly machine learning, artificial intelligence, natural language processing, deep learning, federated learning, and data mining. These technologies increasingly support predictive modelling, medical image analysis, clinical decision support, and the analysis of large healthcare datasets (J. S. Chen & Baxter, 2022; Jiang et al., 2017; Yang, Soltan, Eyre, Yang, et al., 2023). However, the review also indicates that implementing these technologies requires specialised knowledge and technical competencies that extend beyond traditional healthcare practice. As AI-enabled technologies become more widely adopted, healthcare professionals must be equipped to understand both their capabilities and their limitations rather than relying solely on automated outputs.

Despite the reported benefits, the review consistently identified data governance, cybersecurity, and ethical considerations as major barriers to the responsible implementation of BDA. Data quality, interoperability, governance, and the management of increasingly complex healthcare datasets were the most frequently reported challenges, demonstrating that the value of BDA depends fundamentally on the quality and integrity of the underlying data (Awrahaman et al., 2022; Hammad et al., 2020). Security concerns, including privacy, cyberattacks, cloud computing vulnerabilities, and protection of electronic health records, further emphasise the need for robust governance frameworks and organisational safeguards (Jagadeeswari et al., 2018; Zhou et al., 2020). In addition, concerns surrounding algorithmic bias, transparency, accountability, and the use of opaque AI systems illustrate that technical capability alone is insufficient to ensure responsible implementation (Di Nucci, 2019; Norori et al., 2021; Yang, Soltan, Eyre, Yang, et al., 2023).

These findings suggest that preparing a future workforce requires more than technical familiarity with AI-enabled technologies. The review indicates that graduates entering increasingly data-driven healthcare environments will require competencies in AI literacy, data governance, cybersecurity, ethical AI, and interdisciplinary collaboration to support the responsible adoption of BDA. While many of the reviewed studies focused primarily on technological implementation, the collective evidence demonstrates that successful adoption depends equally on healthcare professionals' ability to interpret data appropriately, recognise ethical implications, critically evaluate algorithmic outputs, and participate in multidisciplinary decision-making. These findings support the need for higher education programmes to integrate responsible AI and data-informed practice into curriculum design to better align graduate capabilities with the evolving requirements of healthcare systems.

Several limitations should also be considered when interpreting these findings. As a rapid review, the study provides a focused synthesis of the available evidence rather than the

comprehensive coverage expected from a full systematic review. The search was restricted to English-language publications between 2013 and 2023, and the evidence base reflects the technologies and healthcare applications available during that period. Given the rapid development of artificial intelligence and healthcare analytics, new developments may emerge beyond the review timeframe. Consequently, continued research is required to evaluate the long-term implementation of BDA, particularly regarding governance frameworks, responsible AI adoption, and workforce preparation within higher education.

Conclusion

This PRISMA-guided review synthesised evidence from 35 studies published between 2013 and 2023 to examine the benefits, enabling technologies, and challenges associated with Big Data Analytics (BDA) in healthcare and to consider their implications for workforce preparation. The findings demonstrate that BDA supports predictive capabilities, personalised healthcare, and operational efficiency, while its successful implementation depends on technologies such as machine learning, artificial intelligence, data mining, and Hadoop. At the same time, the review consistently identified data governance, cybersecurity, and ethical AI as the principal challenges limiting responsible adoption.

Beyond its technological contribution, the review highlights an important educational implication. As healthcare increasingly relies on AI-enabled and data-driven systems, future healthcare professionals will require competencies that extend beyond clinical expertise. The evidence suggests that workforce readiness should encompass AI literacy, data governance, cybersecurity awareness, ethical AI, and interdisciplinary collaboration to enable graduates to critically evaluate, govern, and responsibly apply data-driven technologies within healthcare practice.

Although BDA offers considerable opportunities to improve healthcare delivery, its benefits can only be realised when supported by robust governance frameworks and a workforce equipped with the knowledge and skills necessary for responsible implementation. Higher education, therefore, has an important role in embedding these competencies within curriculum design to prepare graduates for the evolving demands of data-driven healthcare.

Future research should continue to evaluate emerging AI-enabled technologies, governance frameworks, and educational approaches that strengthen responsible BDA adoption while ensuring healthcare professionals remain prepared for the continuing evolution of digital healthcare systems.

This review concludes that preparing the future healthcare workforce requires not only technological innovation but also the integration of AI literacy, responsible AI, data governance, cybersecurity, and interdisciplinary learning within higher education to support the safe, ethical, and effective use of Big Data Analytics in healthcare.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that:

- RCSI Bahrain – Licensed, ChatGPT-Edu (OpenAI) was used in a limited capacity to assist with condensing and restructuring sections of the manuscript derived from the

author's original dissertation and accepted conference poster to meet the IAFOR Conference Proceedings requirements.

- Grammarly was used for grammar, spelling, punctuation, and stylistic proofreading,
- Mendeley Reference Manager was used to organise references and generate citations in APA 7th edition format.

The use of these tools was limited to editorial support and reference management. The study design, literature search, PRISMA-guided review, data analysis, interpretation of findings, conclusions, and final academic content are the original work of the author.

References

- Akhtar, U., Lee, J. W., Bilal, H. S. M., Ali, T., Khan, W. A., & Lee, S. (2020). The Impact of Big Data In Healthcare Analytics. *2020 International Conference on Information Networking (ICOIN)*, 61–63. <https://doi.org/10.1109/ICOIN48656.2020.9016588>
- Arora, G., Joshi, J., Mandal, R. S., Shrivastava, N., Virmani, R., & Sethi, T. (2021). Artificial Intelligence in Surveillance, Diagnosis, Drug Discovery and Vaccine Development against COVID-19. *Pathogens*, 10(8). <https://doi.org/10.3390/pathogens10081048>
- Awrahman, B. J., Aziz Fatah, C., & Hamaamin, M. Y. (2022). A Review of the Role and Challenges of Big Data in Healthcare Informatics and Analytics. *Computational Intelligence and Neuroscience*, 2022, 5317760. <https://doi.org/10.1155/2022/5317760>
- Babel, A., Taneja, R., Mondello Malvestiti, F., Monaco, A., & Donde, S. (2021). Artificial Intelligence Solutions to Increase Medication Adherence in Patients With Non-communicable Diseases. *Front Digit Health*, 3, 669869. <https://doi.org/10.3389/fdgth.2021.669869>
- Benke, K., & Benke, G. (2018). Artificial Intelligence and Big Data in Public Health. *International Journal of Environmental Research and Public Health*, 15(12), 2796. <https://www.mdpi.com/1660-4601/15/12/2796>
- Boland, A., Cherry, M. G., & Dickson, R. (2014). *Doing a Systematic Review - A Student's Guide 2nd Edition*. SAGE Publications Ltd, 527.
- Chawla, N. V., & Davis, D. A. (2013). Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework. *Journal of General Internal Medicine*, 28(3), 660–665. <https://doi.org/10.1007/s11606-013-2455-8>
- Chen, J. S., & Baxter, S. L. (2022). Applications of natural language processing in ophthalmology: present and future. *Frontiers in Medicine*, 9, 906554. <https://doi.org/10.3389/fmed.2022.906554>
- Chen, S. W., Gu, X. W., Wang, J. J., & Zhu, H. S. (2021). AIoT Used for COVID-19 Pandemic Prevention and Control. *Contrast Media Mol Imaging*, 2021, 3257035. <https://doi.org/10.1155/2021/3257035>
- Cheung, K. S., Leung, W. K., & Seto, W. K. (2019). Application of Big Data analysis in gastrointestinal research. *World Journal of Gastroenterology*, 25(24), 2990–3008. <https://doi.org/10.3748/wjg.v25.i24.2990>
- Covidence – systematic review management. (n.d.). <https://www.covidence.org/>
- D'Costa, A., & Zatale, A. (2021). AI and the cardiologist: when mind, heart and machine unite. *Open Heart*, 8(2). <https://doi.org/10.1136/openhrt-2021-001874>

- Del Rio-Bermudez, C., Medrano, I. H., Yebes, L., & Poveda, J. L. (2020). Towards a symbiotic relationship between big data, artificial intelligence, and hospital pharmacy. *Journal of Pharmaceutical Policy and Practice*, 13(1), 75. <https://doi.org/10.1186/s40545-020-00276-6>
- Di Nucci, E. (2019). Should we be afraid of medical AI? *Journal of Medical Ethics*, 45(8), 556–558. <https://doi.org/10.1136/medethics-2018-105281>
- Dove, E. S., Reed-Berendt, R., & Pareek, M. (2022). “Data makes the story come to life:” understanding the ethical and legal implications of Big Data research involving ethnic minority healthcare workers in the United Kingdom-a qualitative study. *BMC Medical Ethics*, 23(1), 136. <https://doi.org/10.1186/s12910-022-00875-9>
- EndNote 21. (n.d.). EndNote. <https://endnote.com/?language=en>
- Firouzi, F., Farahani, B., Daneshmand, M., Grise, K., Song, J., Saracco, R., Wang, L. L., Lo, K., Angelov, P., Soares, E., Loh, P. S., Talebpour, Z., Moradi, R., Goodarzi, M., Ashraf, H., Talebpour, M., Talebpour, A., Romeo, L., Das, R., ... Luo, A. (2021). Harnessing the Power of Smart and Connected Health to Tackle COVID-19: IoT, AI, Robotics, and Blockchain for a Better World. *IEEE Internet of Things Journal*, 8(16), 12826–12846. <https://doi.org/10.1109/jiot.2021.3073904>
- Garattini, C., Raffle, J., Aisyah, D. N., Sartain, F., & Kozlakidis, Z. (2019). Big Data Analytics, Infectious Diseases and Associated Ethical Impacts. *Philosophy & Technology*, 32(1), 69–85. <https://doi.org/10.1007/s13347-017-0278-y>
- Golas, S. B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S., Kvedar, J., & Jethwani, K. (2018). A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Medical Informatics and Decision Making*, 18(1), 44. <https://doi.org/10.1186/s12911-018-0620-z>
- Goldman, K. D., & Schmalz, K. J. (2004). The Matrix Method of Literature Reviews. *Health Promotion Practice*, 5(1), 5–7. <https://doi.org/10.1177/1524839903258885>
- Grossglauser, M., & Saner, H. (2014). Data-driven healthcare: from patterns to actions. *European Journal of Preventive Cardiology*, 21(2 Suppl), 14–17. <https://doi.org/10.1177/2047487314552755>
- Hammad, R., Barhoush, M., & Abed-Alguni, B. H. (2020). A Semantic-Based Approach for Managing Healthcare Big Data: A Survey. *Journal of Healthcare Engineering*, 2020, 8865808. <https://doi.org/10.1155/2020/8865808>
- Hulsen, T. (2022). Data Science in Healthcare: COVID-19 and Beyond. *International Journal of Environmental Research and Public Health*, 19(6). <https://doi.org/10.3390/ijerph19063499>

- Jagadeeswari, V., Subramaniaswamy, V., Logesh, R., & Vijayakumar, V. (2018). A study on medical Internet of Things and Big Data in personalized healthcare system. *Health Information Science and Systems*, 6(1), 14. <https://doi.org/10.1007/s13755-018-0049-x>
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. <https://doi.org/10.1136/svn-2017-000101>
- Lemmen, C., Simic, D., & Stock, S. (2021). A Vision of Future Healthcare: Potential Opportunities and Risks of Systems Medicine from a Citizen and Patient Perspective-Results of a Qualitative Study. *International Journal of Environmental Research and Public Health*, 18(18). <https://doi.org/10.3390/ijerph18189879>
- Majnarić, L. T., Babič, F., O’Sullivan, S., & Holzinger, A. (2021). AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity. *Journal of Clinical Medicine*, 10(4). <https://doi.org/10.3390/jcm10040766>
- Murphy, D. C., & Saleh, D. B. (2020). Artificial Intelligence in plastic surgery: What is it? Where are we now? What is on the horizon? *The Annals of The Royal College of Surgeons of England*, 102(8), 577–580. <https://doi.org/10.1308/rcsann.2020.0158>
- Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2(10).
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3. <https://doi.org/10.1186/2047-2501-2-3>
- Ross, M. K., Wei, W., & Ohno-Machado, L. (2014). “Big data” and the electronic health record. *Yearbook of Medical Informatics*, 9(1), 97–104. <https://doi.org/10.15265/iy-2014-0003>
- Tenali, N., & Babu, G. R. M. (2023). A Systematic Literature Review and Future Perspectives for Handling Big Data Analytics in COVID-19 Diagnosis. *New Generation Computing*, 41(2), 243–280. <https://doi.org/10.1007/s00354-023-00211-8>
- Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), 337–339. <https://doi.org/10.1016/j.dsx.2020.04.012>

- Viceconti, M., Hunter, P., & Hose, R. (2015). Big data, big knowledge: big data for personalized healthcare. *IEEE Journal of Biomedical and Health Informatics*, 19(4), 1209–1215. <https://doi.org/10.1109/jbhi.2015.2406883>
- Xiong, H., Chen, H., Xu, L., Liu, H., Fan, L., Tang, Q., & Cho, H. (2022). A survey of data element perspective: Application of artificial intelligence in health big data. *Frontiers in Neuroscience*, 16, 1031732. <https://doi.org/10.3389/fnins.2022.1031732>
- Yang, J., Soltan, A. A. S., Eyre, D. W., & Clifton, D. A. (2023). Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning. *nature machine intelligence*, 5(8), 884–894. <https://doi.org/10.1038/s42256-023-00697-3>
- Yang, J., Soltan, A. A. S., Eyre, D. W., Yang, Y., & Clifton, D. A. (2023). An adversarial training framework for mitigating algorithmic biases in clinical machine learning. *npj digital medicine*, 6(1), 55. <https://doi.org/10.1038/s41746-023-00805-y>
- Zhou, N., Corsini, E. M., Jin, S., Barbosa, G. R., Kell, T., Antonoff, M. H., & Antonoff, M. B. (2020). Advanced Data Analytics for Clinical Research Part I: What are the Tools? *Innovations: Technology and Techniques in Cardiothoracic and Vascular Surgery*, 15(2), 114–119. <https://doi.org/10.1177/1556984520902783>

Contact emails: smuhammad@rcsi.com
sultan.muhammad@me.com