

Leveraging Big Data and Predictive Analytics to Forecast Postpartum Depression Risk

Reem Osama^{1*}, Ghada Elkhayat², Abeer A. Amer³

^{1,2}Department of Information Systems and Computers, Faculty of Business, Alexandria University, Alexandria, Alexandria Governorate, Egypt.

³Department of Computer Science and Information Systems, Faculty of Management Science, Sadat Academy for Management and Sciences, Cairo, Cairo Governorate, Egypt.

reem.osama@alexu.edu.eg¹, ghada.elkhayat@alexu.edu.eg², abamer.2000@gmail.com³.

*Corresponding author

Abstract: This research paper highlights the transformative role of Big Data Analytics (BDA) in enhancing women's health, with a particular focus on Postpartum Depression (PPD). Focusing on the field of FemTech—an abbreviation for Female Technology—this study emphasises technology's potential to address women's specific healthcare needs. The global FemTech market, anticipated to reach its peak in 2024, encompasses applications, devices, wearables, and diagnostics tailored to women's health. This research emphasises the importance of leveraging big data derived from sources such as FemTech and the Internet of Things (IoT) to predict and manage health conditions, such as PPD. It introduces a novel framework integrating data from FemTech apps, smartwatches, and Electronic Medical Records (EMRs) to identify women at risk of PPD with unprecedented accuracy and efficiency. The paper argues that this predictive capacity could dramatically alter healthcare delivery for women, ensuring more personalised and proactive care. By harnessing advanced machine learning models, the research aims to revolutionise the early detection of PPD, ultimately enhancing the wellbeing of mothers and families worldwide. It advocates interdisciplinary collaboration to explore further the intersections of technology, healthcare, and big data analytics.

Keywords: Big Data Analytics; Women's Health; Femtech and Healthcare; Postpartum Depression; Internet of Things; Electronic Medical Records; Bank for International Settlements; Intersections of Technology.

Cite as: R. Osama, G. Elkhayat, and A. A. Amer, "Leveraging Big Data and Predictive Analytics to Forecast Postpartum Depression Risk," *AVE Trends in Intelligent Health Letters*, vol. 2, no. 3, pp. 178–186, 2025.

Journal Homepage: <https://avepubs.com/user/journals/details/ATIHL>

Received on: 09/11/2024, **Revised on:** 26/02/2025, **Accepted on:** 19/05/2025, **Published on:** 07/09/2025

DOI: <https://doi.org/10.64091/ATIHL.2025.000175>

1. Introduction

Despite women constituting at least half of the global population, the number of technology vendors catering to their healthcare needs represents only a tiny segment of the worldwide technology market. The FemTech industry is just beginning to develop, and experts suggest it will reach its peak by 2024. The term "FemTech" is an abbreviation of "Female Technology" and was coined by Ida Tin, the Danish founder of the Clue ovulation-tracking app, which launched in Germany in 2013 [1]. Femtech is a new and emerging industry that develops innovative technology solutions such as apps, devices, wearables, diagnostic software, services, or any combination thereof, specialising in the health needs of women. A research report by the Bank for

Copyright © 2025 R. Osama *et al.*, licensed to AVE Trends Publishing Company. This is an open access article distributed under [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows unlimited use, distribution, and reproduction in any medium with proper attribution.

International Settlements (BIS) estimated the value of big data in healthcare at \$14.27 billion in 2017, with projections that this would increase to \$68.75 billion by the end of 2025.

In 2017, North America led as the largest market for healthcare big data. It introduced the PreManage Emergency Department system, which enables practitioners to share patient records and reports, thereby reducing time and saving money. Sharing patient records is pivotal for efficient disease management. Accessing information from shared patient databases enables the prediction of various types of cancer and the identification of suitable treatment options. Consequently, focusing on data generated by FemTech and applying advanced analytics will yield valuable insights and models relevant to women's health. In this study, researchers concentrate on women's health by exploring a critical and sensitive area: postpartum depression (PPD). PPD is a prevalent and severe mood disorder impacting women after childbirth, carrying significant implications for the mother, infant, and family [2]. Early detection and intervention can enhance outcomes; however, PPD frequently remains underdiagnosed and undertreated due to obstacles such as stigma, a lack of awareness, and constrained resources [3].

A promising method to overcome these barriers involves using big data analytics to identify the women most at risk of PPD, thereby directing preventive measures and support interventions to them. Big data analytics (BDA) is pivotal in the healthcare sector, enabling significant advancements in patient care, operational efficiency, and medical research. It also aids in early detection of disease spread and in preemptive prediction of illnesses. Additionally, it assists in monitoring clinical performance, underscoring its essential role in improving patient care and outcomes. In the following sections, researchers will delve into the background of big data in the healthcare industry, including its analysis, and into postpartum depression and the application of the Internet of Things (IoT) in healthcare, as detailed in Section II. Section III will outline the proposed framework for predicting women most at risk of developing PPD. To wrap up, researchers will provide a summary and discuss potential future directions.

2. Background

2.1. Big Data

2.1.1. Big Data in General

Big data refers to large, complex datasets that pose challenges for traditional database management systems in processing and handling. ZDNet.com describes big data as encompassing the tools and techniques for generating, utilising, and storing vast amounts of data using storage devices. Every individual continuously generates substantial amounts of data. All computer systems and social networking sites produce big data. Digital systems, sensors, cameras, wearable devices, smartphones, and their applications distribute it [4]. Big data arrives with unparalleled velocity, encompassing a vast array of data from diverse sources. Deriving meaningful insights from such extensive datasets requires significant computing power, analytical power, and specialised expertise. This explosion of data is prompting a reevaluation of all aspects related to big data. Notably, transaction data, database information, sensor data, and electronic medical records are expanding rapidly, reflecting the evolving landscape of data generation and utilisation [5]. This data can be categorised into several types, including web data, sensor-based data, demographic data, transactional data, and machine-generated data, as detailed below:

- **Web Data:** Gathered from social networking platforms like Facebook, Twitter, and blogs.
- **Machine-Generated Data:** Collected by sensors installed on various devices and gadgets.
- **Transactional Data:** Derived from biometrics, vital signs, radiology, and other medical imaging sources.
- **Human-Generated Data:** Includes emails, doctors' prescriptions, and scanned copies of medical reports.

2.1.2. Characteristics of Big Data

The following attributes often summarise the fundamental characteristics of big data:

- **Volume:** Represents the immense size of data, typically measured in terabytes (TB, 10¹² bytes), petabytes (PB, 10¹⁵ bytes), zettabytes (ZB, 10²¹ bytes), etc.
- **Velocity:** Indicates the rapid rate at which data is generated and needs to be processed.
- **Variety:** big data encompasses three different types of data, including structured, semi-structured, and unstructured formats.
- **Veracity:** Concerns about the reliability and accuracy of the data.
- **Value:** Reflects the valuable insights and significance one can derive from the data.

Beyond these foundational characteristics, the definition and features of big data have evolved to encompass additional dimensions as technology, data storage, processing speeds, and other capabilities have advanced [6]; [7]. Various frameworks

and scholars have expanded the concept beyond the initial 3Vs or 4Vs to include 5Vs, 10Vs, and even more [8]. As of 2017, 42 distinct dimensions of big data were identified, a number expected to increase as the field continues to expand [9]. These dimensions underscore the multifaceted nature of big data and its growing importance across sectors, underscoring the need for sophisticated tools and methodologies to harness its full potential [10] (Figure 1).

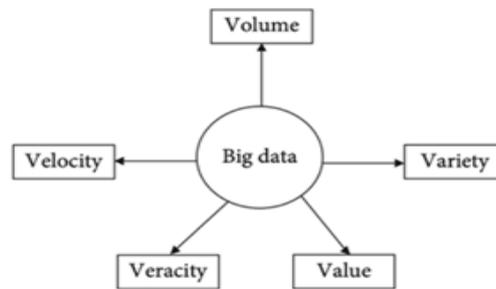


Figure 1: Generic concept of big data [11]

2.1.3. Healthcare Big Data

The healthcare sector is undergoing a significant transformation, driven by vast amounts of health data generated by technological advancements and the digitisation of medical records. Recent developments in Health Information Technologies (HIT) have enabled the electronic generation, storage, and global transmission of data in mere seconds, offering the potential to significantly enhance the efficiency and quality of health services. This innovation permits each healthcare provider to maintain a digital database of patient records. Consequently, the healthcare sector has amassed substantial data from medical records, regulatory approvals, requirements, and patient care [12]. This collective data constitutes a vast health dataset, specifically referred to as Big Health Data. Big Health Data is characterised by Electronic Medical Records (EMRs) that encompass a patient's medical history, doctors' notes, clinical reports, biometric information, and other health-related data, along with contributions from social media such as blog posts, tweets, Facebook news alerts, and medical journal entries [12].

This change in healthcare includes not only more data, but also a faster rate of data production and a wider range of sorts of medical data being created. Sensor systems, cameras, and cellphones are just a few examples of technology that are moving quickly and are very important for collecting healthcare data. As new data sources come out every day, it gets harder for traditional database management solutions to analyse and analyse healthcare big data. By properly recording, storing, and analysing large amounts of healthcare data, you may make better decisions that lead to better service results and lower healthcare expenditures. To do this, though, you need advanced analytical tools and methods, as well as strong computer systems.

Healthcare Big Data Analytics (BDA) has emerged as a critical tool for addressing challenges across multiple healthcare domains. Additionally, a data analyst's responsibilities include mining big data to uncover relationships and understand trends and patterns in healthcare data. This process not only improves individual health and quality of life but also enables the provision of early-stage treatments at reduced costs. The rapidly growing volume of data in the healthcare sector has significantly heightened interest in big data's capabilities. With an extensive pool of data ready for analysis, a key goal of leveraging big data is to improve healthcare services. Nations worldwide are motivated to improve healthcare facilities and reduce medical expenses. However, the huge amount of data in healthcare is still a big problem that makes it hard to reach these goals. In 2012, it was thought that there were 500 petabytes of electronic health data around the world.

By 2020, that number had grown to 25 petabytes [13]. So, healthcare is a broad term that refers to the many services that medical professionals offer to individuals, families, or communities to help them stay healthy, improve their health, or get back to good health. The quality of the healthcare system is important because it affects how well a hospital can grow over time and how well patients can stay healthy. In some cases, the quality of healthcare services is very good, which makes them more expensive for patients. So, it's important to focus on the most important healthcare procedures and the quality measures that go along with them. This will help produce the best results for patients while keeping expenses down.

2.1.4. Sources of Healthcare Big Data

Healthcare big data has different forms, as shown in Figure 2. Structured, semi-structured, and unstructured data are forms of big data. Structured healthcare big data can be represented as clinical data, such as EMR data. However, semi-structured data

are claims data and laboratory data. Moreover, the majority of healthcare data is unstructured, represented in imaging data and other text, such as physician notes.

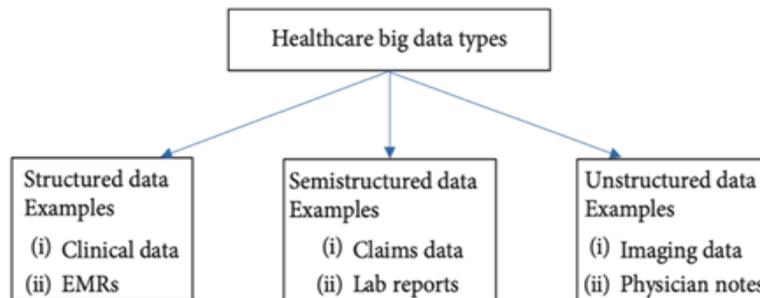


Figure 2: Types of healthcare big data [11]

2.1.5. The 5Vs of Healthcare Big Data Characteristics

In this section, the five key characteristics that have been found in most literature to define healthcare big data are discussed as follows [14]; [15]:

- **Volume:** Based on the broader discussion of big data, healthcare data is a prime example. The volume of large datasets can reach 44 zettabytes [16]. Compared with many other sectors, the healthcare industry generates a large volume of data, including electronic medical records (EMRs), biometric data, clinical information, X-ray images, genomic data, and more, collectively known as Big Health Data [17]-[19].
- **Velocity:** Velocity refers to the rate at which data is generated and acquired from different healthcare systems [20].
- **Variety:** In the context of big data, it refers to the diversity of data types and sources, highlighting the heterogeneity of the data collected and analysed. The healthcare sector is generating and collecting data rapidly from various sources, such as social media, sensors, cameras, and smartphones. However, this health data can be in any structured, unstructured, or semi-structured form. An example of structured data is clinical data. In contrast, data such as doctors' notes, images, social media data, mobile data, and X-ray videos are unstructured or semi-structured, as shown in Figure 2.
- **Veracity:** Veracity characteristics of health data refer to data reliability, which in this context corresponds to data quality assurance. This gives health knowledge a certain degree of authenticity.
- **Value:** The most important and distinctive characteristic of Big Data in 5V healthcare is its ability to transform healthcare data into valuable information. This concept fits perfectly with the concept of health data.

2.1.6. Types of Healthcare Big Data Analytics (BDA)

BDA does three main types of analysis: descriptive, predictive, and regulatory. Descriptive analytics facilitate the examination of information and assist clinicians in comprehending the dynamics of a certain situation [21]; [22]. In the realm of health data, descriptive analytics examines gathered data to interpret, comprehend, condense, and represent pertinent health information visually. On the other hand, predictive analytics helps doctors figure out what services to offer and how to answer patient questions. It also lets doctors make decisions about patients based on predictions from the system [21]; [22]. Predictive analytics includes a number of statistical methods for looking at large datasets and getting useful information from them. [6]. One of the most common ways to make predictive models of healthcare systems is to use Hadoop/MapReduce. Prescriptive analytics is a newer sort of analytics that blends descriptive and predictive analytics [23]. Predictive analytics tells us what will happen in the future, whereas prescriptive analytics tells healthcare practitioners what they should do [21]; [22]. Prescriptive Analytics integrates clinical and genetic data to continually forecast healthcare outcomes and enhance prediction accuracy, hence enabling healthcare providers to offer more suitable diagnoses and treatments [24]; [25].

2.1.7. Healthcare Big Data Analytics Tools

Several Big Data Analytics (BDA) techniques and tools have been developed to support distributed processing within the Hadoop ecosystem [28]; [29]. Hadoop Common provides a set of core libraries and utilities that support other Hadoop modules and is considered a fundamental component of the Apache Hadoop platform alongside HDFS, YARN, and MapReduce. It is often referred to as Hadoop Core and offers essential services required for the proper functioning of the Hadoop framework [26]. Apache HDFS, or Hadoop Distributed File System, is designed to store and process large volumes of unstructured data

on standard hardware. It divides files into fixed-size blocks and distributes them across multiple nodes in a cluster. HDFS follows a master-slave architecture consisting of a NameNode that manages metadata and multiple DataNodes that store actual data, ensuring scalability, fault tolerance, and high availability [30]; [31].

Hadoop MapReduce is a distributed programming model that enables parallel processing of massive datasets across clusters. It consists of two primary phases: the map phase, which processes input data and generates intermediate key-value pairs, and the reduce phase, which aggregates and summarizes the results after the mapping phase is completed [27]; [32]; [33]. Apache Hive is a data warehouse framework built on Hadoop that facilitates querying and analyzing large datasets stored in HDFS. It provides an SQL-like language known as Hive Query Language (HiveQL) and supports Extract, Transform, and Load (ETL) operations, allowing efficient data analysis for users familiar with SQL [34]. Apache Pig is a high-level parallel computing platform that operates on Hadoop and uses the Pig Latin scripting language. Pig Latin is similar to SQL but is more flexible and capable of handling semi-structured and unstructured data, making it suitable for complex data processing tasks [30]; [35].

Apache HBase is an open-source, distributed, multi-dimensional NoSQL database that runs on HDFS. It is designed to store very large datasets ranging from terabytes to petabytes and provides fast read and write access, although it does not support traditional SQL queries [30]. Apache Mahout is an open-source distributed framework for big data analytics and machine learning built on Hadoop. It supports scalable algorithms for tasks such as classification, clustering, collaborative filtering, association analysis, and statistical modeling using MapReduce-based processing [30]; [36].

2.2. Postpartum Depression

Postpartum depression (PPD) is regarded as the most frequent complication associated with childbirth, affecting one in seven women [37]. It is estimated that over 20% of women worldwide suffer from PPD [38]. PPD usually occurs 6-8 weeks after birth [39]. Mothers often face discomfort related to physical changes, poor sleep quality, and various insecurities related to the newborn in the postpartum period [39]. A study developed predictive models for PPD using various risk factors, including medical and psychiatric history, as well as factors related to childbirth and pregnancy. These models were validated using logistic regression and assessed for their reproducibility and transportability. The models showed good performance in predicting PPD within six months after birth [40]. Another study explored machine learning methods to predict postpartum depressive symptoms. Various algorithms, including Ridge Regression, LASSO Regression, Distributed Random Forests, and Extremely Randomised Trees, were evaluated. The Extremely Randomised Trees model showed the highest accuracy and a good balance in sensitivity and specificity. The study also looked at the stratification of results based on previous depression, showing varied performance based on women's past depression history [41].

Besides, a study by Saqib et al. [42] synthesised the literature on machine learning (ML) and big data analytics for maternal mental health, particularly for predicting PPD. The authors identified 14 studies that used supervised learning techniques to predict PPD. The most commonly used algorithms were Support Vector Machine, Random Forest, Naive Bayes, regression, artificial neural networks, decision trees, and XGBoost (Extreme Gradient Boosting). The study concluded that ML algorithms can significantly improve the early detection of PPD. Furthermore, a study conducted in Lebanon used the Edinburgh Postnatal Depression Scale (EPDS) to predict PPD. They conducted a statistical analysis to determine the likelihood of developing PPD based on EPDS scores and identified potential risk factors for PPD. Their findings suggest the effectiveness of EPDS in early detection and the importance of considering various factors, including personal and family history, in predicting PPD [43].

2.3. Internet of Things

The Internet of Things (IoT) refers to “the interconnected network of physical objects, or 'things,'” that are equipped with sensors, software, and other technologies, aiming to connect and share data with other devices and systems across the internet. This network includes a wide array of devices, from simple household items to advanced industrial machinery [44]. These devices span from everyday household items to complex industrial equipment. Accordingly, IoT in healthcare manifests primarily through asset monitoring, offering numerous advantages to the industry. Medical staff, including doctors, nurses, and orderlies, frequently require precise information about the location of patient-assistance tools, such as wheelchairs.

By outfitting hospital wheelchairs with IoT sensors, their locations can be monitored via an IoT asset-monitoring application, enabling anyone seeking a wheelchair to easily locate the nearest available one. This method of tracking can be applied to various hospital assets, facilitating efficient use and financial accounting of physical resources across different departments [45]. Besides monitoring health and environmental conditions, IoT wearables empower individuals to gain a deeper understanding of their health and enable doctors to monitor patients remotely. This technology also allows companies to oversee the health and safety of their workforce, proving particularly beneficial for employees working under dangerous conditions. Accordingly, in this research, we leverage big data generated by IoT devices to achieve our goal.

3. Methodology

This research proposes a new framework that leverages the strengths of each data source to provide a holistic, predictive view of potential PPD cases. The framework's primary objective is to integrate diverse data sources, including FemTech applications, smartwatches (wearable devices), and electronic medical records (EMRs), to predict and identify women at risk of developing Postpartum Depression (PPD) with high accuracy and efficiency.

3.1. Framework Components

3.1.1. Data Collection Module

- **FemTech Data Collection:** Leverage APIs to gather data from FemTech apps focusing on menstrual cycle tracking, fertility, and pregnancy monitoring.
- **Smartwatch (wearable devices) Data Integration:** Collect health metrics from smartwatches and wearable devices, including sleep patterns, heart rate variability, physical activity levels, and stress markers.
- **Medical Data Aggregation:** Securely access EMRs for clinical history, pregnancy, childbirth details, and postnatal check-ups.

3.1.2. Data Processing and Integration Engine

- **Preprocessing and Normalisation:** Standardise data formats, handle missing values, and normalise data scales across different sources.
- **Feature Engineering:** Extract and select features that strongly correlate with PPD risk, including hormonal changes (from FemTech data), sleep disturbances (from smartwatches), and a history of mental health issues (from EMRs).
- **Data Integration:** Merge data from all sources into a unified data model, ensuring privacy and compliance with data protection regulations.

3.1.3. Analytics and Prediction Module

- **Machine Learning Models:** Develop and train machine learning models on the integrated dataset. Models such as Random Forests, Gradient Boosting, and Neural Networks could be evaluated for their predictive accuracy and sensitivity.
- **Predictive Analytics:** Implement algorithms that can analyse the data in real-time or near-real-time to predict the likelihood of developing PPD.
- **Model Evaluation and Optimisation:** Continuously assess predictive model performance using metrics such as accuracy, precision, recall, and F1 score. Adjust and refine models based on feedback loops and updated data (Figure 3).

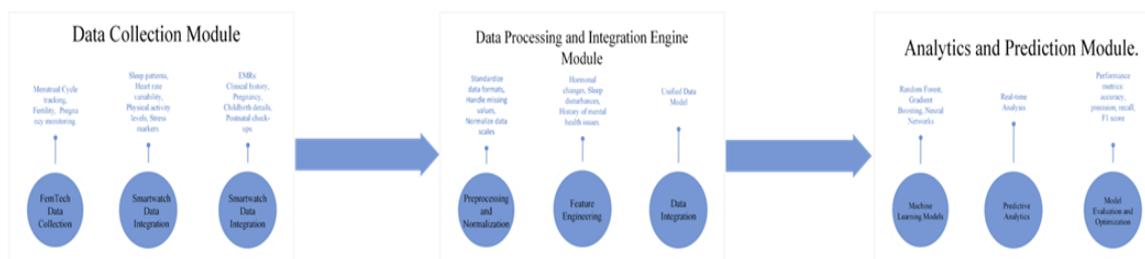


Figure 3: System architecture of the health data analysis pipeline

The development of this comprehensive framework is crucial, particularly in fields where data is scarce and must be meticulously gathered from a variety of sources. This approach is not commonly addressed in the existing literature, highlighting the innovative nature of this framework. By consolidating disparate data into a single, coherent structure, it becomes possible to uncover patterns and insights that would otherwise remain hidden. Furthermore, the efficacy of predictive analytics is heavily reliant on the application of machine learning techniques and the availability of big data. Machine learning algorithms require vast amounts of data to learn from, to make accurate predictions and uncover meaningful trends. The proposed framework facilitates the aggregation and processing of large datasets, enabling these algorithms to function optimally. In essence, the framework serves as the foundation for predictive analytics, leveraging big data to drive innovation and decision-making in the healthcare sector.

4. Conclusion

In conclusion, this research highlights the important role of big data analytics and emerging technologies, such as FemTech and IoT devices, in advancing women's health, specifically in the early prediction of Postpartum Depression (PPD). The integration of diverse data sources, including data from FemTech applications, wearable devices, and electronic medical records (EMRs), offers a novel approach to understanding and predicting PPD. This condition significantly impacts the well-being of mothers and their families globally. Our proposed framework leverages the strengths of big data analytics to process and analyse the vast, multifaceted datasets derived from these technologies. By employing advanced machine learning models, we aim to predict the likelihood of PPD with greater accuracy and efficiency than ever before. This predictive capability could revolutionise the way healthcare providers approach the diagnosis and treatment of PPD, offering a more proactive and personalised healthcare experience for women.

The integration of data from FemTech and wearable devices with traditional medical records presents a promising avenue for uncovering hidden patterns and indicators of PPD. It is a testament to the transformative power of big data analytics in healthcare, particularly in areas where early detection can lead to improved outcomes. Moreover, this approach emphasises the importance of interdisciplinary collaboration between technologists, healthcare professionals, and researchers to address complex health issues. As researchers look to the future, the continued evolution of big data technologies and their applications in healthcare holds immense potential to enhance patient care and health outcomes. The ongoing development and refinement of frameworks like the one proposed in this research will be crucial in harnessing the full potential of big data analytics for the betterment of women's health and beyond.

4.1. Future work (Recommendations)

Researchers are planning to implement the proposed framework in future research, as well as some recommendations listed below:

- Study the effect of demographic variations on women's health using femTech analytics.
- Study women's psychological health using data from the femtech application.
- Discuss the variations in demographics and education level for women who develop PPD.

The exploration of demographic variations, psychological health, and educational impacts on PPD through FemTech analytics as part of our future work will further contribute to the body of knowledge in this field. Our commitment to implementing and testing the proposed framework in real-world settings will not only validate its effectiveness but also pave the way for its adoption in healthcare practices worldwide, marking a significant step forward in the fight against Postpartum Depression.

Acknowledgement: The authors express their sincere appreciation to Alexandria University and Sadat Academy for Management and Sciences for their guidance, resources, and continuous academic support throughout this work.

Data Availability Statement: The data supporting the findings of this study are available from the authors upon reasonable request, ensuring full transparency and facilitating future scholarly use.

Funding Statement: The authors confirm that this research was conducted without any external funding, and all activities were carried out using internal and self-supported resources.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest that may have affected the conduct, analysis, or conclusions of this study.

Ethics and Consent Statement: The authors collectively provide their consent to share this work for academic, educational, and research purposes, ensuring open access for all interested readers.

References

1. Cambridge University Press, "Femtech," *Cambridge Dictionary*, 2016. Available : <https://dictionary.cambridge.org/dictionary/english/femtech> [Accessed by 10/09/2024].
2. S. Mgonja and A. Schoening, "Postpartum depression screening at well-child appointments: a quality improvement project," *J. Pediatr. Health Care*, vol. 31, no. 2, pp. 178–183, 2017.

3. M. W. L. Moreira, J. J. P. C. Rodrigues, N. Kumar, K. Saleem, and I. V. Illin, "Postpartum depression prediction through pregnancy data analysis for emotion-aware smart systems," *Information Fusion*, vol. 47, no. 5, pp. 23–31, 2019.
4. S. Sagioglu and D. Sinanc, "Big data: A review," *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, California, United States of America, 2013.
5. I. Lee, "Big data: dimensions, evolution, impacts, and challenges," *Business Horizons*, vol. 60, no. 3, pp. 293–303, 2017.
6. A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.
7. C. H. Lee and Y. Hyung-Jin, "Medical big data: promise and challenges," *Kidney Research and Clinical Practice*, vol. 36, no. 1, pp. 3–11, 2017.
8. R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives," *Computers and Industrial Engineering*, vol. 101, no. 11, pp. 572–591, 2016.
9. K. F. Tiampo, S. McGinnis, Y. Kropivnitskaya, J. Qin, and M. A. Bauer, "Big data challenges and hazards modeling," in *Risk Modeling for Hazards and Disasters*, Elsevier, Amsterdam, Netherlands, 2018.
10. L. Wang and C. A. Alexander, "Big data in medical applications and health care," *American Medical Journal*, vol. 6, no. 1, pp. 1–8, 2015.
11. R. Raja, I. Mukherjee, and B. K. Sarkar, "A Systematic Review of Healthcare Big Data," *Scientific Programming*, vol. 2020, no. 1, p. 5471849, 2020.
12. K. Priyanka and N. Kulennavar, "A survey on big data analytics in health care," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5865–5868, 2014.
13. J. Sun and K. C. Reddy, "Big data analytics for healthcare," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining-KDD '13*, New York, United States of America, 2013.
14. S. Nazir, M. Nawaz, A. Adnan, S. Shahzad, and S. Asadi, "Big data features, applications, and analytics in cardiology—a systematic literature review," *IEEE Access*, vol. 7, no. 9, pp. 143742–143771, 2019.
15. B. K. Sarkar, "Big data for secure healthcare system: a conceptual design," *Complex and Intelligent Systems*, vol. 3, no. 2, pp. 133–151, 2017.
16. M. S. Hajirahimova and A. S. Aliyeva, "About big data measurement methodologies and indicators," *International Journal of Modern Education and Computer Science*, vol. 9, no. 10, pp. 1-9, 2017.
17. A. Widmer, R. Schaer, D. Markonis, and H. Müller, "Gesture Interaction for Content-based Medical Image Retrieval," in *Proceedings of International Conference on Multimedia Retrieval (ICMR '14)*, New York, United States of America, 2014.
18. J. A. Seibert, "Modalities and data acquisition," in *Practical Imaging Informatics*, Springer, New York, United States of America, 2009.
19. A. S. Panayides, M. S. Pattichis, S. Leandrou, C. Pitris, A. Constantinidou, and C. S. Pattichis, "Radiogenomics for precision medicine with a big data analytics perspective," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2063–2079, 2018.
20. B. Cyganek, M. Graña, B. Krawczyk, A. Kasprzak, P. Porwik, K. Walkowiak, and M. Woźniak, "A survey of big data issues in electronic health record analysis," *Applied Artificial Intelligence*, vol. 30, no. 6, pp. 497–520, 2016.
21. G. Phillips-Wren, L. S. Iyer, U. Kulkarni, and T. Ariyachandra, "Business analytics in the context of big data: a roadmap for research," *Communications of the Association for Information Systems*, vol. 37, no. 1, p. 23, 2015.
22. H. J. Watson, "Tutorial: big data analytics: concepts, technologies, and applications," *Communications of the Association for Information Systems*, vol. 34, no. 1, p. 65, 2014.
23. D. Delen, "Real-World Data Mining: Applied Business Analytics and Decision Making," *F.T. Press*, Upper Saddle River, New Jersey, United States of America, 2014.
24. M. Riabacke, M. Danielson, and L. Ekenberg, "State-of-the-art prescriptive criteria weight elicitation," *Advances in Decision Sciences*, vol. 2012, no. 11, pp. 1-24, 2012.
25. Z. Pang, H. Yuan, Y. T. Zhang, and M. Packirisamy, "Guest editorial health engineering driven by the industry 4.0 for aging society," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1709-1710, 2018.
26. C. Ngufor and J. Wojtusiak, "Learning from large-scale distributed health data: an approximate logistic regression approach," in *Proc. ICML 13: Role of Machine Learning in Transforming Healthcare*, Atlanta, Georgia, United States of America, 2013.
27. R. Zhang, G. Simon, and F. Yu, "Advancing Alzheimer's research: a review of big data promises," *International Journal of Medical Informatics*, vol. 106, no. 10, pp. 48–56, 2017.
28. P. Zikopoulos, D. deRoos, K. Parasuraman, T. Deutsch, J. Giles, and D. Corrigan, "Harness the Power of Big Data the IBM Big Data Platform," *McGraw Hill Professional*, New York, United States of America, 2012.
29. P. Zikopoulos and C. Eaton, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data," *McGraw Hill Professional*, New York, United States of America, 2011.

30. A. Oussous, B. Fatima-Zahra, A. A. Lahcen, and S. Belfkih, "Big data technologies: a survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 431–448, 2018.
31. V. Rajaraman, "Big data analytics," *Resonance*, vol. 21, no. 8, pp. 695–716, 2016.
32. M. Idris, S. Hussain, M. Ali, A. Abdulali, M. H. Siddiqi, B. H. Kang, and S. Lee, "Context-aware scheduling in MapReduce: a compact review," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 17, pp. 5332–5349, 2015.
33. H. Senger, V. Gil-Costa, L. Arantes, C. A. C. Marcondes, M. Marin, L. M. Sato, and F. A. B. Da silva, "BSP cost and scalability analysis for MapReduce operations," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 8, pp. 2503–2527, 2016.
34. GeeksforGeeks, "Hadoop Ecosystem," *GeeksforGeeks*, 2025. Available: <https://www.geeksforgeeks.org/dbms/hadoop-ecosystem/> [Accessed by 17/09/2024].
35. A. K. Bhadani and D. Jothimani, "Big data: challenges, opportunities, and realities," in *Effective Big Data Management and Opportunities for Implementation*, IGI Global, Pennsylvania, United States of America, 2016.
36. N. Khan, I. Yaqoob, I. A. T. Hashem, Z. Inayat, W. K. M. Ali, M. Alam, M. Shiraz, and A. Gani, "Big data: survey, technologies, opportunities, and challenges," *The Scientific World Journal*, vol. 2014, no. 1, p. 712826, 2014.
37. S. Mgonja and A. Schoening, "Postpartum depression screening at well-child appointments: a quality improvement project," *J. Pediatr. Health Care*, vol. 31, no. 2, pp. 178–183, 2017.
38. N. Ukatu, C. A. Clare, and M. Brulja, "Postpartum depression screening tools: a review," *Psychosomatics*, vol. 59, no. 3, pp. 211–219, 2018.
39. J. J. Rogathi, R. Manongi, D. Mushi, V. Rasch, G. N. Sigalla, T. Gammeltoft, and D. W. Meyrowitsch, "Postpartum depression among women who have experienced intimate partner violence: a prospective cohort study at Moshi, Tanzania," *J. Affect. Disord.*, vol. 218, no. 8, pp. 238–245, 2017.
40. T. Munk-Olsen, X. Liu, K. B. Madsen, Z. K. Mette-Marie, L. V. Petersen, V. Bergink, A. Skalkidou, S. N. Vigod, V. G. Frokjaer, C. B. Pedersen, and M. L. Maegbaek, "Postpartum depression: a developed and validated model predicting individual risk in new mothers," *Transl. Psychiatry*, vol. 12, no. 1, p. 419, 2022.
41. S. Andersson, D. R. Bathula, S. I. Iliadis, M. Walter, and A. Skalkidou, "Predicting women with depressive symptoms postpartum with machine learning methods," *Sci. Rep.*, vol. 11, no. 1, p. 7877, 2021.
42. K. Saqib, A. F. Khan, and Z. A. Butt, "Machine Learning Methods for Predicting Postpartum Depression: Scoping Review," *JMIR Mental Health*, vol. 8, no. 11, pp. 1-14, 2021.
43. C. El-Hachem, J. Rohayem, R. B. Khalil, S. Richa, A. Kesrouani, R. Gemayel, N. Aouad, N. Hatab, E. Zaccak, S. Salameh, and E. Attieh, "Early identification of women at risk of postpartum depression using the Edinburgh Postnatal Depression Scale (EPDS) in a sample of Lebanese women," *BMC Psychiatry*, vol. 14, no. 1, p. 242, 2014.
44. A. R. Sfar, Z. C. Challal, and Y. Challal, "A systematic and cognitive vision for IoT security: a case study of military live simulation and security challenges," in *Proc. 2017 International Conference on Smart, Monitored and Controlled Cities (SM2C)*, Sfax, Tunisia, 2017.
45. Z. N. Aghdam, A. M. Rahmani, and M. Hosseinzadeh, "The Role of the Internet of Things in Healthcare: Future Trends and Challenges," *Computer Methods and Programs in Biomedicine*, vol. 199, no. 2, p. 105903, 2021.