UNIVERSIDAD
NACIONAL
DE COLOMBIA

Advancing Healthcare Analytics: A Thematic Review of Machine Learning, Health Informatics, and Real-world Data Applications

María Isabel Arias Ríos

Universidad Nacional de Colombia
Facultad de Minas, Área Curricular de Sistemas e Informática
Medellín, Colombia
2025

Advancing Healthcare Analytics: A Thematic Review of Machine Learning, Health Informatics, and Real-world Data Applications

María Isabel Arias Ríos

Trabajo final presentado como requisito parcial para optar al título de:
Magister en ingeniería con énfasis en analítica – profundización

Director:
Juan David Velásquez Henao, MSc, PhD

Universidad Nacional de Colombia
Facultad de Minas, Área Curricular Sistemas e Informática
Medellín, Colombia
2025

ACKNOWLEDGMENTS

This thesis marks the end of a very significant chapter in my life, and it would not have been possible without the support, guidance, and companionship of many people to whom I wish to express my deepest gratitude.

First and foremost, I thank God for granting me health, strength, and wisdom throughout this academic journey. His presence has accompanied me in moments of difficulty and has been my constant guide.

To my family, especially my parents, for their unconditional love, constant support, and words of encouragement at every stage of this path. Thank you for believing in me and for being my greatest motivation.

To my thesis advisor, Juan David Velásquez Henao, for his patience, academic guidance, and valuable insights that significantly enriched this project. His commitment to excellence inspired me to give my best.

To the professors of the National University of Colombia for sharing their knowledge and fostering in me a critical, ethical, and professional attitude toward learning.

To my classmates and friends, whose friendship, collaboration, and emotional support made this process more manageable. Thank you for the conversations, advice, and shared moments that I will always remember fondly.

Finally, I would like to thank the committee members for their valuable feedback, and all the people who, in one way or another, contributed to the completion of this work and to my professional development. To all of you, thank you!

ABSTRACT

Advancing Healthcare Analytics: A Thematic Review of Machine Learning, Health Informatics, and Real-world Data Applications

This study aims to map the conceptual and methodological landscape of healthcare analytics by identifying dominant thematic clusters, synthesizing key trends, and outlining translational challenges and research opportunities in the field. To achieve this, 2,281 Scopus-indexed publications were analyzed using unsupervised text mining and clustering techniques, with a focus on identifying recurring themes, methodological innovations, and gaps in the healthcare analytics literature across clinical, administrative, and public health contexts.

The analysis revealed eight dominant themes: intelligent systems for predictive healthcare, patient-centered health analytics, adaptive AI for clinical insights, demographic health analytics, digital mental health surveillance, ethical analytics for health surveillance, personalized care through data analytics, and AI-driven insights for outbreak response. Together, these clusters illustrate an ongoing transition toward real-time, multimodal, and ethically grounded analytics ecosystems. However, the field continues to face persistent challenges, such as data interoperability, algorithmic opacity, standardization of evaluation practices, and demographic bias.

The review also underscores emerging priorities that are likely to shape the next phase of development, including explainable AI, federated learning, and context-aware modeling, alongside ethical considerations related to data privacy and digital equity. From a practical perspective, key recommendations include the co-design of solutions with healthcare professionals, greater investment in infrastructure, and the deployment of real-time clinical decision support systems. Overall, healthcare analytics is positioned as a foundational pillar of learning health systems, carrying significant implications for translational research and the advancement of precision health.

Keywords

Healthcare analytics; health informatics; medical informatics; digital health; clinical decision support; mHealth.

RESUMEN

Avances en la analítica en salud: una revisión temática del aprendizaje automático, la informática en salud y las aplicaciones de datos del mundo real

Este estudio tiene como propósito mapear el panorama conceptual y metodológico de la analítica en salud, identificando los principales clústeres temáticos, sintetizando las tendencias clave y señalando los desafíos de transferencia y las oportunidades de investigación en el campo. Para ello, se analizaron 2.281 publicaciones indexadas en Scopus mediante técnicas no supervisadas de minería de texto y agrupamiento, con el objetivo de identificar temas recurrentes, innovaciones metodológicas y vacíos en la literatura sobre analítica en salud en contextos clínicos, administrativos y de salud pública.

El análisis reveló ocho temas dominantes: sistemas inteligentes para la atención sanitaria predictiva, analítica en salud centrada en el paciente, inteligencia artificial adaptativa para generar conocimientos clínicos, analítica de salud demográfica, vigilancia digital de la salud mental, analítica ética para la vigilancia en salud, atención personalizada a través de la analítica de datos e información impulsada por IA para la respuesta a brotes. En conjunto, estos clústeres reflejan una transición hacia ecosistemas de analítica en tiempo real, multimodales y con fundamentos éticos. Sin embargo, el campo aún enfrenta desafíos persistentes como la interoperabilidad de datos, la opacidad algorítmica, la estandarización de evaluaciones y los sesgos demográficos.

La revisión también resalta prioridades emergentes que probablemente orientarán la siguiente fase de desarrollo, entre ellas la inteligencia artificial explicable, el aprendizaje federado y la modelación consciente del contexto, junto con consideraciones éticas relacionadas con la privacidad de los datos y la equidad digital. Desde una perspectiva práctica, se proponen recomendaciones clave como el co-diseño de soluciones con profesionales de la salud, una mayor inversión en infraestructura y la implementación de sistemas de apoyo a la decisión clínica en tiempo real. En conjunto, la analítica en salud se posiciona como un pilar fundamental de los sistemas de salud en aprendizaje, con importantes implicaciones para la investigación traslacional y el avance de la salud de precisión.

Palabras clave

Analítica en salud; informática en salud; informática médica; salud digital; apoyo a la decisión clínica; mHealth.

CONTENT

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Healthcare analytics is a data-intensive and interdisciplinary field. It integrates computational methods, statistical modeling, artificial intelligence, and informatics to transform complex health data into actionable insights. These insights support decision-making at clinical, operational, and policy levels. The field is defined by its ability to analyze large-scale, high-dimensional, and irregular datasets. These include electronic health records, biomedical signals, population health indicators, and patient-generated data. By using such sources, healthcare analytics supports predictive modeling, risk stratification, and real-time monitoring [1], [2], [3], [4]. Healthcare analytics extends beyond clinical environments to include administrative and public health applications, enabling evidence-based interventions that promote health equity, precision medicine, and system-wide performance optimization [5], [6], [7], [8]. Methodologically, healthcare analytics relies on diverse learning systems such as deep learning, statistical inference, and hybrid models. These systems improve diagnostic accuracy, forecast patient outcomes, and support continuous monitoring in different care settings [9], [10], [11]. It also addresses key challenges such as data sparsity, interoperability, and model interpretability through reproducible frameworks, explainable AI, and privacy-preserving analytics [12], [13], [14]. The development of cloud-based platforms and IoMT infrastructures further supports the scalability and responsiveness of healthcare analytics, enabling dynamic cohort discovery and adaptive decision-making aligned with real-world needs [15], [16], [17]. From a conceptual perspective, healthcare analytics is not limited to tool deployment but also provides strategic frameworks for structuring research agendas, supporting clinical workflows, and guiding public health initiatives [18], [19], [20]. It fosters methodological innovation by integrating domain expertise into learning architecture and advancing reproducible, generalizable solutions for biomedical informatics [21], [22], [23]. As such, Healthcare analytics has emerged as a foundational pillar of modern biomedical informatics, providing scientifically rigorous and practically translatable insights that prioritize real-world applicability, conceptual coherence, and measurable translational outcomes.

While healthcare analytics has garnered substantial attention in recent years, the field lacks a comprehensive and methodologically grounded literature review that maps its global research landscape, identifies dominant thematic areas, and outlines current challenges and future directions. Existing reviews have primarily focused on specific subdomains, such as the application of machine learning for disease diagnosis and prediction [21], the classification of analytics into descriptive, predictive, and prescriptive types [7], or the evaluation of data mining techniques in clinical decision-making [22]. Although these contributions are valuable, they often provide a fragmented view, frequently limited to specific technologies, organizational levels, or application settings. Several studies have proposed frameworks or research agendas [23], and some have explored the commercial value of healthcare analytics [24] or its potential applications in domains such as precision medicine and public health [25]. However, most of these reviews do not provide an integrative analysis that captures the conceptual, methodological, and translational dimensions of the field.

Few studies have systematically examined how healthcare analytics research is distributed across global thematic clusters. Even fewer have critically assessed what these distributions mean for methodological development or practical implementation. While bibliometric reviews exist [26], they often stop at surface-level trend identification without a deeper analysis of how dominant topics intersect, evolve, or align with broader biomedical challenges. As systematic reviews have demonstrated, the field continues to confront and address challenges related to data quality, interoperability, legal constraints, and the ethical use of data while promoting generalizable methodologies aligned with evidence-based medicine [27], [28], [29]. To address these limitations, this review aims to apply text-mining techniques to a curated corpus of published literature to uncover recurring patterns, thematic structures, and methodological innovations in healthcare analytics. Through a critical examination of these findings, the review not only synthesizes the current state of knowledge but also identifies conceptual gaps, interdisciplinary synergies, and systemic barriers that affect the translation of data-driven methods into real-world clinical and public health contexts. This approach provides a novel contribution by offering a holistic and generalizable understanding of healthcare analytics as a foundational and evolving field in biomedical research.

Practical relevance and market scope. Recent industry reports highlight the magnitude of healthcare analytics as a rapidly expanding sector, with the global market valued at approximately USD 30 billion in 2022 and projected to exceed USD 80 billion by 2030, driven by the integration of electronic health records, predictive analytics, and AI-enabled monitoring systems [30], [31]. Beyond market growth, healthcare analytics has demonstrated tangible impacts, such as reducing hospital readmissions, improving chronic disease management, and generating substantial cost savings for health systems worldwide [32], [33]. These figures underscore both the economic significance and the transformative potential of healthcare analytics in addressing systemic challenges.

Scope, limitations, and contribution of this review. While comprehensive in its methodological and thematic synthesis, this review does not provide a disease-specific or country-specific analysis, nor does it incorporate non-indexed sources such as gray literature. Instead, its scope is to map the global research landscape, identify thematic clusters, and examine methodological innovations with an emphasis on their translational potential. By doing so, this work contributes a holistic and integrative perspective on healthcare analytics, bridging fragmented subdomain insights into a coherent framework that highlights both opportunities and systemic barriers. This review provides a foundation for advancing methodological innovation. It also helps align research with practical implementation and guides future agendas in biomedical informatics and public health.

The structure of this article is as follows: Section 2 presents an in-depth review of the relevant literature. Section 3 outlines the methodological approach adopted in the research. Section 4 reports the main results. Section 5 delivers a critical interpretation of the findings. Lastly, Section 6 concludes the paper.

Statement of Significance

Problem:
Despite the rapid expansion of healthcare analytics, there is no comprehensive, methodologically grounded review that maps the field's thematic structure, translational challenges, and methodological gaps.

What is already known:
Prior reviews typically focus on narrow subdomains—such as predictive modeling, descriptive analytics, or specific clinical applications—offering fragmented insights into the field's overall landscape.

What this paper adds:
This review uses text mining and clustering techniques to analyze 2,281 publications, identifying eight dominant themes in healthcare analytics. It provides a holistic synthesis of conceptual trends, technical innovations, and translational barriers while outlining future research directions, including federated learning, explainable AI, and privacy-preserving systems.

Who would benefit from the knowledge in this paper:
This paper will benefit biomedical informaticians, data scientists, clinicians, public health researchers, and policymakers seeking to design, evaluate, or implement data-driven methods in healthcare environments.

2 LITERATURE REVIEW

The recovered database has seven documents classified as reviews in healthcare analytics, listed in **Table 1**, with some bibliometric citation indicators.

***Table 1** Documents classified as reviews.*

Title	Authors	Rank Global Citations	Global Citations	Rank Local Citations	Local Citations	Year
Evolution of machine learning applications in medical and healthcare analytics research: A bibliometric analysis	Ajibade et al. [26]	1592	1	1592	0	2024
Machine Learning in Healthcare Analytics: A State-of-the-Art Review	Das et al. [21]	523	12	622	0	2024
Healthcare analytics—A literature review and proposed research agenda	Elragal et al. [23]	1348	2	1359	0	2023
Health analytics in business research: a literature review	Liu et al. [24]	758	7	241	1	2023
Precision Health Analytics With Predictive Analytics and Implementation Research: JACC State-of-the-Art Review	Pearson et al. [25]	184	37	387	0	2020
Health Analytics Types, Functions and Levels: A Review of Literature	Khalifa [7]	309	20	21	5	2018
A systematic review on healthcare analytics: Application and theoretical perspective of data mining	Islam et al. [22]	25	201	1	21	2018

Ajibade et al. [26] investigated the global development of machine learning in medical and healthcare research from 1994 to 2023 through a data mining analysis of Scopus and PubMed sources. The study reveals that funding has played a more pivotal role than international collaboration in shaping publication output. It identifies three core research themes, underscoring the expanding role of machine learning in chronic disease management, data security, and smart healthcare while highlighting future priorities in ethics and system integration.

Das [21] explores the growing role of machine learning in healthcare, emphasizing the variety of models, ranging from logistic regression and support vector machines to deep learning approaches, such as convolutional neural networks, for image analysis. Ensemble methods, such as Random Forest and Gradient Boosting, offer accuracy and reliability, while hybrid models enhance adaptability by merging deep learning and traditional techniques. The study provides insights into model selection, current challenges, and future directions in healthcare analytics.

Elragal et al. [23] emphasize the urgent need for continued research in healthcare analytics by reviewing how big data, AI, and machine learning have been applied to address healthcare challenges. The study bridges the fields of healthcare science and data science, identifying research gaps and proposing a future agenda for scholars, institutions, and policymakers. Offering a state-of-the-art overview, the paper supports both academic and practical efforts to advance analytics-driven solutions in modern healthcare systems.

Liu et al. [24] review the expanding role of health analytics in business research, emphasizing its commercial potential and value. By analyzing studies published in business journals, the authors outline how health data can inform diverse business domains, consolidate standard datasets and analytical methods, and address key research challenges. The review provides a roadmap for future inquiry, encouraging business scholars to explore health analytics for both practical applications and theoretical contributions.

Pearson et al. [25] highlight how predictive analytics can improve the management of complex conditions, such as heart, lung, blood, and sleep disorders. Advances in data science are enhancing the quality and availability of health data. The authors emphasize the need for implementation research to assess the effectiveness, risks, and policy implications of these tools. Drawing from a National Heart, Lung, and Blood Institute workshop, the review emphasizes aligning precision medicine with public health efforts.

The work of Khalifa [7] explores health analytics as a business-driven approach that merges business intelligence and big data to enhance healthcare outcomes. Through a literature review and qualitative analysis, the study identifies five analytics types—descriptive, diagnostic, predictive, prescriptive, and discovery—each contributing uniquely to healthcare improvement. Health analytics functions at operational, tactical, and strategic levels, demonstrating its broad impact on decision-making and its growing role in achieving the strategic goals of healthcare organizations.

Islam et al. [22] review healthcare analytics research from 2005 to 2016, using PRISMA guidelines, with a focus on subdomains, data sources, and methods. Most studies support the use of electronic medical records in clinical and administrative decisions, although the use of social media data is also increasing. Key gaps include limited prescriptive analytics and weak domain knowledge integration. The authors call for future research to address these issues and advance practical applications in healthcare analytics.

An integrative reading of existing reviews highlights both converging themes and notable differences across healthcare analytics research. Islam et al. [22] demonstrate that analytics has been predominantly applied to clinical and administrative decision-making, with an increasing use of social media data, but limited prescriptive applications. Pearson et al. [25] emphasize the importance of predictive analytics in managing complex conditions and underscore the need for implementation research to align precision medicine with public health. Khalifa [7] provides a functional classification of analytics types and organizational levels, illustrating their strategic role in healthcare decision-making. Complementing these perspectives, Das [21] and Ajibade et al. [26] document the rapid expansion of machine learning, including deep learning, ensemble, and hybrid approaches, while also revealing uneven global development that is more shaped by funding than collaboration. Liu et al. [24] extend the field into business research by emphasizing its commercial potential, and Elragal et al. [23] propose a research agenda that bridges healthcare and data science. ogether, these perspectives offer partial insights, yet they leave unanswered questions about the broader conceptual, methodological, and translational landscape that this review seeks to address.

3 MATERIALS AND METHODS

This paper uses the standard workflow for literature analysis [34], [35], [36], incorporating recent enhancements from informatics research [37], [38]:

1. Study design.
2. Data collection and preparation.
3. Data analysis and interpretation

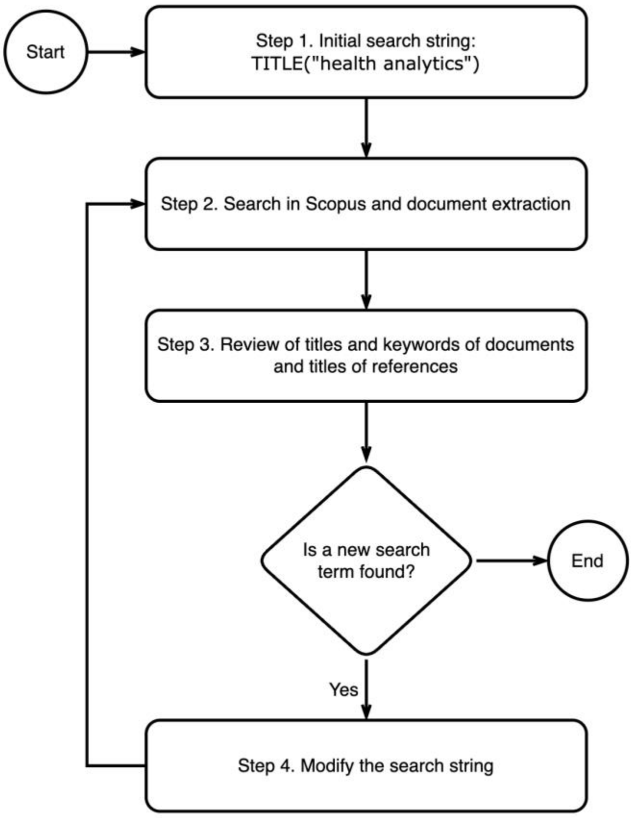
3.1 Study Design

Table 2 summarizes the parameters applied in this study. The first step of the search strategy, depicted in **Fig. 1**, involved retrieving records containing the term “health analytics” in their titles without applying any time constraints. Titles, author-provided keywords, and indexed keywords were reviewed manually and iteratively to uncover additional relevant terms and refine the use of search operators. Whenever a new term or operator appeared, it was added to the query, starting a new search cycle. Terms such as “medical data analytics” and “clinical analytics” were gradually incorporated. The process ended when no additional terms emerged. The final query string used in Scopus is shown in **Fig. 1**. Furthermore, the review was enhanced by performing both backward and forward citation tracking of key publications. Backward citation tracking entailed examining the reference lists of selected papers, while forward tracking identified subsequent works that cited those papers.

Fig. 2 shows the total number of documents retrieved. The search query ultimately returned 6,527 records during the identification phase. In the screening stage, 2,200 records—including editorials, corrections, letters, data articles, and brief notes—were excluded. Following, the inclusion and exclusion criteria listed in **Table 2** were applied. After this filtering, the final dataset consisted of 2,281 documents.

Table 2 *Parameters of the study.*

Parameter	Value
Database	Scopus.
Years of Analysis	All available.
Data Retrieval	April 7, 2025.
Search String	It is derived using an iterative construction method, which will be elaborated upon in the subsequent section.
Inclusion Criteria	Explicit mention of key terms related to health analytics in the title or abstract. Application of analytical techniques to health-related data. Use of structured or unstructured health data for analysis. Focus on data-driven decision-making in public health, clinical practice, or health services management. Use or development of technologies.
Exclusion Criteria	Did not include any of the key health analytics terms in the title or abstract. Focused on Administrative processes, healthcare policy, or workforce management without data analytics. Focused on Descriptive or conceptual discussions without analytic models or empirical data analysis. Purely qualitative studies. Articles related to Ethics, public policy, training, or organizational transformation in healthcare without analytics. Technology deployment or digital tools in healthcare that are not used for data analysis or health-related insights.



```
TITLE (
   "health analytics"
   OR "health data analytics"
   OR "healthcare analytics"
   OR "health care analytics"
   OR "health informatics"
   OR "public health analytics"
   OR "medical data analytics"
   OR "clinical analytics"
   OR "predictive health analytics"
)
OR AUTHKEY (
   "health analytics"
   OR "health data analytics"
   OR "healthcare analytics"
   OR "health care analytics"
   OR "health informatics"
   OR "public health analytics"
   OR "medical data analytics"
   OR "clinical analytics"
   OR "predictive health analytics"
)
```

***Figure 1** Search string design.*

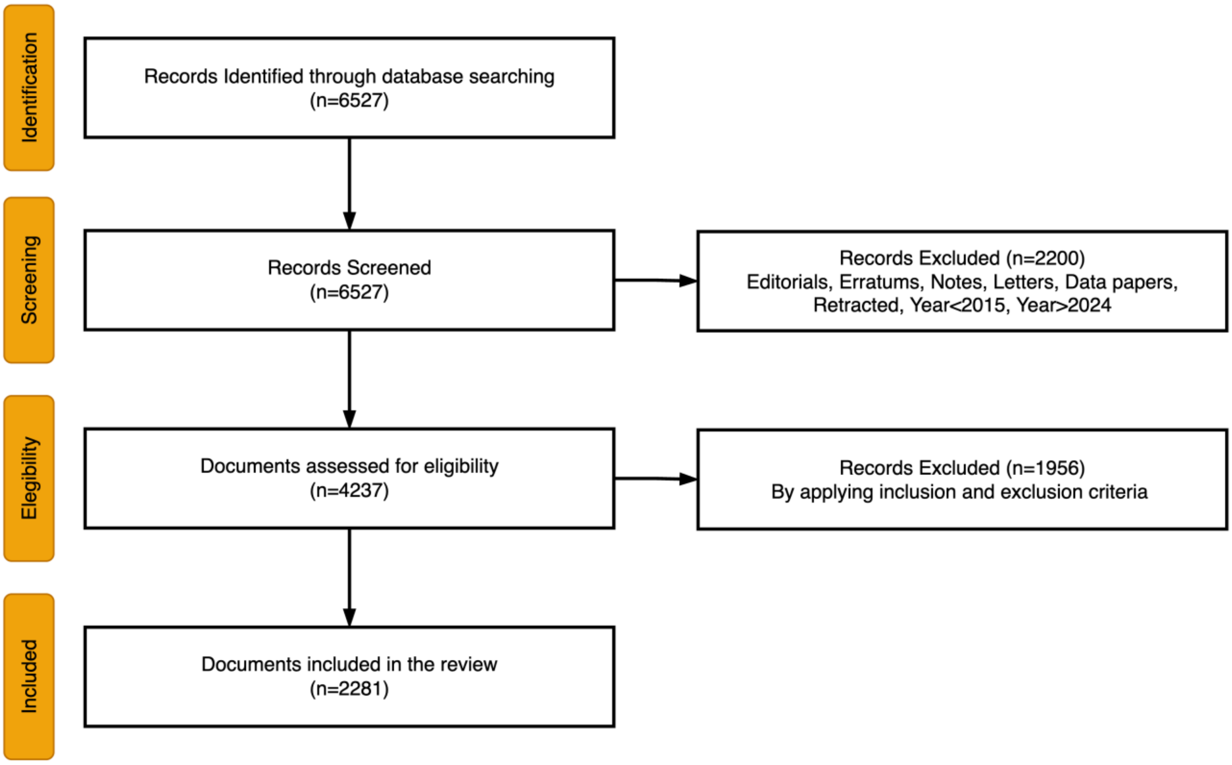


***Figure 2** The PRISMA flow chart.*

3.2 Data collection and preliminary preparation

Bibliographic data from Scopus were exported in CSV format, including titles, abstracts, keywords, authors, affiliations, sources, and references. Text processing combined automated methods with manual adjustments such as uppercasing text, standardizing American spelling, removing extra spaces, and unifying hyphenated terms. Noun phrases were extracted from titles and abstracts, and a new ‘descriptors’ column—combining noun phrases, author keywords, and index keywords—was created to identify dominant themes.

3.3 Data Analysis and Interpretation

This study employs text-mining techniques to uncover dominant themes in the health analytics literature using a tech-mining-based approach. Rather than relying solely on author or index keywords or applying topic modeling, the methodology focuses on analyzing a unified set of document descriptors. These descriptors combine author keywords, index keywords, and noun phrases extracted from titles and abstracts.

The process is divided into two main stages: descriptor cleaning and thematic clustering. **Fig. 3** presents a scheme of the process. Cleaning involves standardizing terms, converting abbreviations to their complete forms, and removing ambiguous or overly general descriptors using a curated stopwords list. This step is essential for improving the quality of the resulting thematic structure. Initially, 73,432 descriptors were collected—4,976 author keywords, 9,630 index keywords, and 65.851 noun phrases. After refinement, this was reduced to 57,790 descriptors, with some appearing in multiple categories.

For the second stage, a co-occurrence matrix was generated from the cleaned descriptors, and the Louvain algorithm was applied to detect thematic clusters by maximizing modularity. These clusters represent the dominant topics in the literature. Alternative clustering approaches in bibliometric research include the Leiden algorithm, hierarchical clustering, and topic modeling methods such as Latent Dirichlet Allocation (LDA). We selected Louvain because it offers a robust balance of scalability and interpretability for large co-occurrence networks, and its performance has been validated in numerous science mapping studies. It is essential to note that, although the Louvain algorithm yields non-overlapping partitions of the co-occurrence network, this does not imply rigid conceptual boundaries. In practice, many documents include descriptors spanning multiple clusters, and subdomains of healthcare analytics naturally overlap; the identified clusters should therefore be interpreted as dominant themes rather than mutually exclusive categories. All analyses were conducted using a custom Python pipeline that integrates descriptor cleaning, co-occurrence matrix construction, Louvain clustering, and visualization into a single, reproducible workflow. This integration, combined with the use of multiple descriptor sources (author keywords, index keywords, and automatically extracted noun phrases), distinguishes the present approach from existing bibliometric tools that typically implement these steps in isolation. A key feature of this approach is the integration of descriptor cleaning with clustering, which enables the simultaneous discovery of themes and resolution of conceptual redundancies. Each cluster was reviewed iteratively to ensure thematic coherence, eliminate noise, and consolidate synonyms. This method enhances the interpretability and precision of theme identification, making it a robust framework for literature reviews and science mapping in emerging research areas.

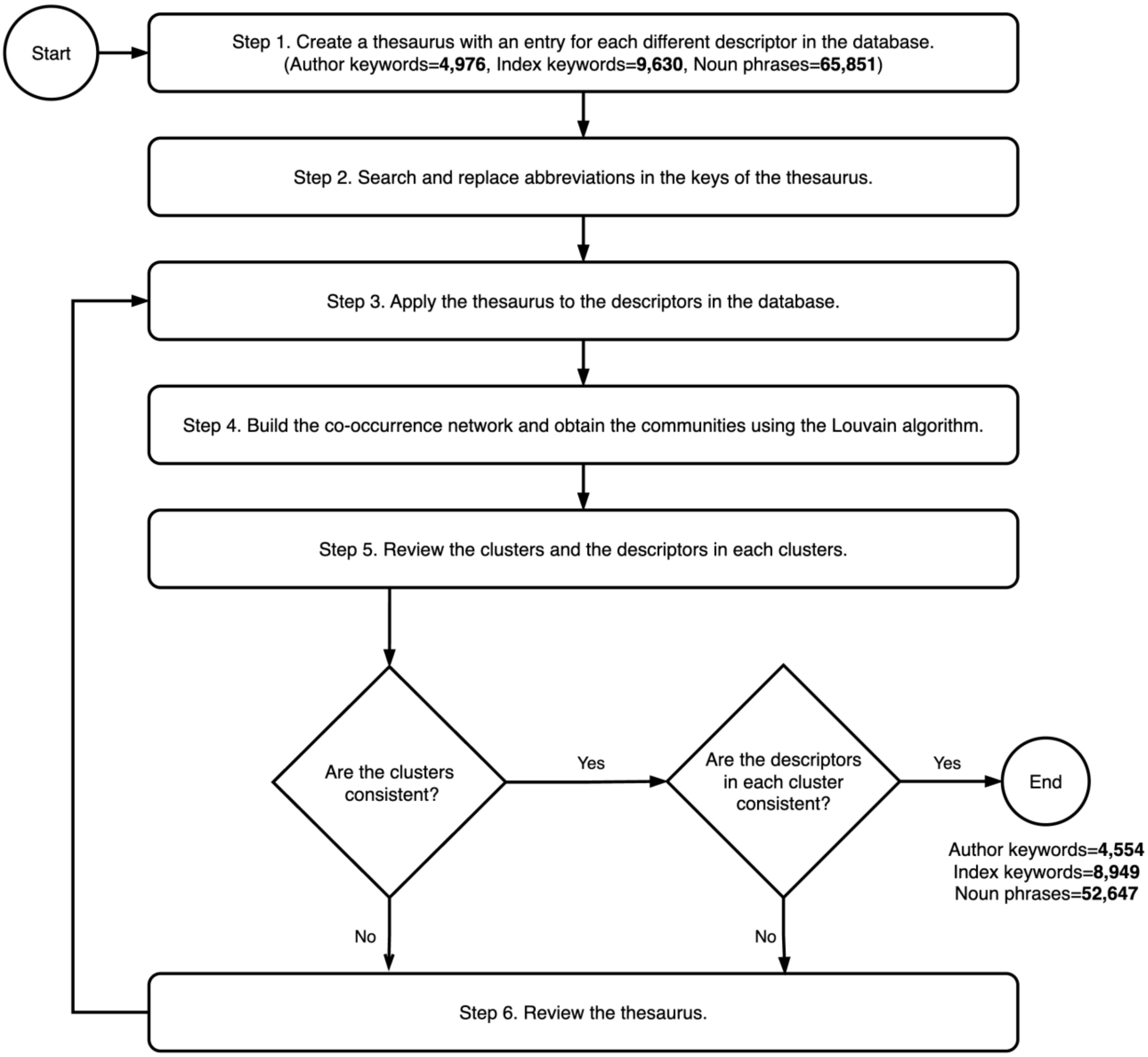


Figure 3 *Used methodology to obtain the dominant themes from database descriptors.*

4 RESULTS

4.1 General Dataset Description

The dataset comprises scientific publications spanning from 2015 to 2024, encompassing a total of 2,281 records and reflecting an annual growth rate of 34.57%. With an average document age of 3.87 years, the dataset is relatively recent. On average, each publication has received 14.75 citations, corresponding to 1.34 citations per year. The dataset draws from 978 distinct sources, with an average of 2.33 documents per source. The majority of contributions are research articles (1,199) and conference papers (721). A total of 10,194 authors contributed to this body of work, demonstrating a high degree of collaboration, as shown by an average of 5.24 authors and 5.57 co-authors per publication. Additionally, 25.47% of the documents involve international cooperation. These contributions originate from 3,919 institutions distributed across 106 countries.

4.2 Dominant Themes

Table 3 summarizes the eight thematic clusters derived using the methodology illustrated in **Fig. 3**. For this analysis, only descriptors with a minimum frequency of 60 occurrences were included, resulting in a total of 127 descriptors, which account for 99.8% of the dataset. The table displays the ten most frequently occurring descriptors, with entries organized by the total number of descriptors associated with each cluster. Within each cluster, the list of descriptors is sorted by their frequency of occurrence. Given that a single document can address multiple thematic areas concurrently, assigning each publication to a single theme is not feasible in this context. **Fig. 4** illustrates the co-occurrence network of descriptors, where node sizes correspond to the frequency of each descriptor within the dataset. The color intensity and thickness of the connecting edges reflect the strength of co-occurrence between descriptors. As anticipated, descriptors related to health informatics emerge as the most prevalent and exhibit strong interconnections.

Table 3 *Dominant thematic clusters*

Cluster Name	Num Terms	Percentage	Main Descriptors
Intelligent Systems for Predictive Healthcare	33	26.0	health and healthcare; artificial intelligence; big data; data mining; data analytics; decision making; healthcare system; health data; decision support system; insight
Patient-Centered Health Analytics	23	18.2	health informatics; patient; electronic health records; hospital; risk; outcome; diabetes; care; healthcare delivery; risk assessments
Adaptive AI for Clinical Insights	21	16.5	machine learning; forecasting or prediction; deep learning; learning systems; support vector machine; random forest; features; detection; predictive analytics; decision tree
Demographic Health Analytics	20	15.7	gender; adults; clinical study; epidemiology; age; cohort analysis; retrospective study; population; risk factors; participants
Digital Mental Health Surveillance	11	8.8	information; people; individuals; natural language processing; social media; topics; symptoms; mental health; internet; social networking
Ethical Analytics for Health Surveillance	10	7.9	medical informatics; public health; public health informatics; evaluation; information system; visualizations; data visualization; information processing; communities; access
Personalized Care through Data Analytics	6	4.8	disease; diagnoses; treatments; patient treatment; condition; patient outcomes
AI-Driven Insights for Outbreak Response	3	2.1	covid; pandemic; outbreaks

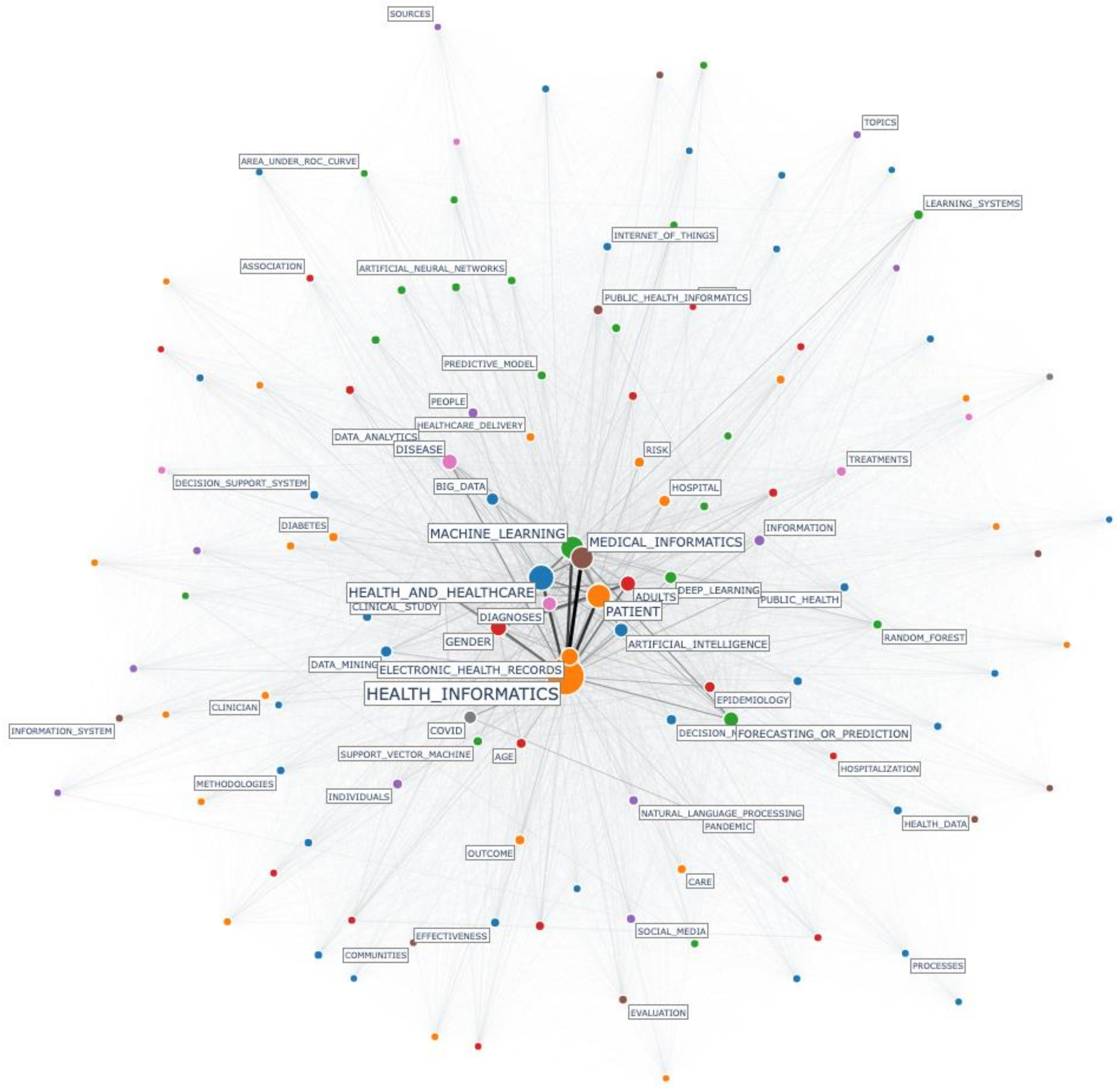


Figure 4 *Co-occurrence network of document descriptors. Node colors represent the eight thematic clusters, while edge widths and opacity indicate the strength of co-occurrence between descriptors. Node size is proportional to descriptor frequency. To improve readability, only the eight most frequent descriptors in each cluster are labeled, with larger font size indicating higher frequency. All other nodes are displayed without labels to minimize visual clutter while maintaining the full network structure.*

5 DISCUSSION

This section presents a critical analysis of the theme associated with each cluster, as listed in **Table 3**. While Section 4 presented the descriptive results of the clustering process, this section provides a critical interpretation of those findings, highlighting their methodological, conceptual, and translational implications. It is also important to note that while the Louvain algorithm yields non-overlapping partitions, these clusters should be interpreted as dominant themes rather than mutually exclusive domains, since many documents and subfields naturally intersect across methodological, clinical, and public health boundaries.

5.1 Theme 1: Intelligent systems for predictive healthcare

5.1.1 Cluster Definition

This thematic cluster defines healthcare analytics as a multidisciplinary and evolving ecosystem where artificial intelligence, machine learning, and data analytics converge to support predictive, adaptive, and patient-centered healthcare delivery across clinical, operational, and public health domains. It reflects a shift from retrospective analysis to real-time and anticipatory modeling by integrating heterogeneous health data—ranging from electronic health records, hospital-level surveys, and medical images to social media traces and wearable sensor data—into interpretable and ethically grounded decision-support systems [2], [39], [40], [41], [42], [43]. The cluster consistently positions AI-driven insight generation as central to the transformation of diagnostic, prognostic, and treatment pathways, as illustrated in use cases spanning obesity prevention, tuberculosis prognosis, depression detection, and neurological disease prediction [44], [45], [46], [47], [48], [49]. The growing reliance on ensemble learning, Bayesian models, and explainable AI not only enhances predictive accuracy but also addresses clinical uncertainty and promotes system-wide transparency and trust [14], [40], [50], [51]. Across studies, the interpretability and ethical deployment of computational tools emerge as non-negotiable design principles, ensuring usability in real-world clinical settings [52], [53], [54].

Additionally, the literature portrays healthcare analytics as an agent of systemic change, capable of revealing inefficiencies, optimizing workflows, and improving patient experience—whether through real-time surveillance during pandemics or process mining in surgical environments [39], [55], [56]. Notably, the cluster redefines healthcare analytics as a proactive infrastructure that scales with data complexity and institutional needs, integrating human-centered considerations such as emotional support, trust, and personalized care [40], [57], [58]. In doing so, it articulates a comprehensive vision of healthcare analytics as a scientifically grounded, operationally viable, and ethically responsive field that empowers both micro-level clinical decisions and macro-level system optimization [59], [60], [61], [62].

5.1.2 Current Trends

Current trends in healthcare analytics reflect a convergence toward real-time, personalized, and ethically aware data-driven systems that integrate artificial intelligence, machine learning, and natural language processing for improved clinical and operational decision-making. A dominant trajectory involves the widespread deployment of predictive models across domains such as diabetes prognosis, emergency admissions, stroke prediction, and mental health, often combining ensemble and hybrid strategies to enhance accuracy, robustness, and personalization [43], [63], [61], [41]. These models increasingly draw on diverse data modalities—ranging from imaging and physiological signals to social media and patient-reported outcomes—while leveraging IoT-enabled wearables and cloud infrastructures for continuous monitoring, early intervention, and remote diagnostics, particularly in resource-constrained environments [62], [64], [58], [65]. Another notable trend is the integration of large language models into clinical workflows, which improves documentation, communication, and the disambiguation of complex clinical texts while supporting adaptive patient care pathways [66], [67], [44]. The emergence of explainable and knowledge-aware frameworks underscores growing concerns for trust, accountability, and interpretability in high-stakes applications such as prenatal diagnostics, neurological disorders, and national surveillance systems [14], [44], [64]. Equally important is the shift toward federated learning and privacy-preserving infrastructures, which prioritize data sovereignty and equitable access while maintaining performance and scalability [53], [64]. Real-time spatiotemporal analysis and public sentiment tracking through infoveillance reflect the extension of healthcare analytics into public health and behavioral surveillance, enriching our understanding of system-level dynamics and user engagement [68], [69], [70]. Meanwhile, human-centered and interdisciplinary design approaches guide the development of usable and trustworthy interfaces, such as telehealth dashboards and personalized alerts, aligning analytical models with clinical contexts and patient preferences [57], [49]. Overall, the field is moving beyond static, retrospective analytics toward dynamic, explainable, and context-aware infrastructures that enable scalable, individualized, and ethically responsive healthcare solutions [52], [71], [54].

5.1.3 Challenges

Despite rapid progress, healthcare analytics still faces challenges that limit its integration into clinical practice and public health systems. Central among these is the persistent issue of model generalizability, as predictive systems often perform inconsistently when transferred across demographic groups, clinical environments, or geographic regions, particularly in the context of rare conditions and multimodal data inputs [51], [41], [72]. Data heterogeneity, imbalance, and poor annotation quality exacerbate this problem, particularly in underrepresented populations and emerging domains such as emotion recognition and mental health intervention [45], [47], [73]. The opacity of many AI models, particularly deep learning architectures, limits interpretability and undermines clinician trust, with explainable AI tools still insufficiently validated for real-world decision-making [14], [52], [48]. Simultaneously, technical infrastructures face obstacles in integrating data from IoT devices, EHRs, and national surveillance systems due to inconsistent interoperability standards, latency issues, and fragmented data governance protocols [39], [65], [74], [62]. Ethical and operational concerns persist across the board: federated learning systems grapple with privacy, high aggregation costs, and scalability bottlenecks in low-resource settings [53], [64]; public health dashboards and decision-support systems struggle with usability, transparency, and equitable access [57], [70]; and clinician engagement remains low due to misalignment with workflow and alert fatigue [75], [56].

Furthermore, systemic inefficiencies—such as the underutilization of process mining in surgical pathways and limited cross-disciplinary collaboration—impede the translational impact of healthcare analytics [55], [70]. Even high-performing ensemble and hybrid models often rely on curated datasets that fail to capture real-world variability, raising questions about the sustainability and fairness of AI-driven interventions [76], [77], [49]. Collectively, these challenges underscore that technical sophistication must be matched with contextual awareness, ethical compliance, and collaborative governance to ensure the equitable, scalable, and trustworthy deployment of healthcare analytics systems [59], [54], [58].

5.1.4 Research Opportunities and Future Directions

Future research in healthcare analytics is poised to transcend algorithmic development, moving toward adaptive, interpretable, and context-aware systems that serve both personalized medicine and public health objectives. A key direction lies in integrating explainable AI with predictive modeling to improve trust, transparency, and clinical decision-making—particularly in sensitive applications such as diagnostic imaging, critical care, and cancer diagnosis [78], [52], [63]. Hybrid learning frameworks that blend symbolic reasoning, clinical guidelines, and deep learning are increasingly advocated to enhance algorithmic transparency and contextual interpretability [59], [60], [79]. Multimodal large language models and dynamic predictive systems offer promising pathways for processing and synthesizing data from text, images, physiological signals, and wearable devices for real-time risk stratification and early interventions [66], [62], [61]. In parallel, longitudinal and personalized models that adapt to evolving patient trajectories—particularly for neurological and metabolic conditions—may improve early warning systems and individualized care [48], [49], [41]. Real-world validation across diverse healthcare contexts remains a critical research need to ensure the robustness, fairness, and generalizability of analytics models [63], [51]. Federated and cloud-based architectures must continue to evolve, addressing latency, data privacy, and computational efficiency to expand access in low-resource and underserved settings [64], [39], [53].

Meanwhile, socio-technical dimensions such as clinician collaboration, patient engagement, dashboard usability, and interdisciplinary validation are essential for ensuring ethical deployment and sustained impact [57], [70], [58]. Public health applications, including pandemic surveillance and mental health monitoring, benefit from combining generative models, uncertainty-aware analytics, and real-time informatics systems [68], [47], [54]. To achieve scalability and resilience, future systems must support dynamic retraining, feedback loops, and interoperability across infrastructures while integrating diverse health indicators—from genomic data to social media signals—for inclusive, proactive, and actionable healthcare analytics [73], [69], [65].

5.1.5 national-level contribution

Although most of the reviewed applications come from Global North contexts, these trends find a highly relevant parallel in Colombia and Latin America. In the case of the cluster on predictive health systems, their application is particularly pertinent to the management of cardiovascular diseases, as they represent the leading cause of mortality in the country according to the Instituto Nacional de Salud. The integration of electronic health records and administrative registries such as SISPRO or RIPS, together with physiological signals from monitoring devices, opens the possibility of developing predictive models for hospitalizations and readmissions in patients at cardiovascular risk. Similarly, advances in patient-centered analytics align with local initiatives in telemedicine and primary care in rural areas, where data analytics could optimize resource allocation, reduce inequities in access, and strengthen preventive care. These examples illustrate how global trends described in the literature can be contextualized and tested in Latin American settings, thereby expanding their relevance and applicability.

5.1.6 Recommendations to enhance practical value

To enhance the practical value of healthcare analytics, the cluster emphasizes a shift from isolated algorithmic development to collaborative, system-oriented integration that prioritizes clinical alignment, ethical governance, and user-centered design. Co-development paradigms involving clinicians, patients, and data scientists are repeatedly underscored as essential to ensure that predictive models are responsive to workflow demands and user needs, including minimizing documentation burdens and maximizing clinical relevance [56], [78], [75]. Embedding explainability-by-design into analytics systems enhances interpretability and fosters clinician trust, particularly when accompanied by transparent validation frameworks and traceable decision pathways [14], [76], [48], [60]. Standardization of performance reporting, benchmarking protocols, and interoperable ontologies is necessary for replicability, cross-platform learning, and regulatory compliance across diverse healthcare environments [63], [77], [65]. Investments in training interdisciplinary teams—including clinicians, informaticians, and policymakers—are crucial to bridging translational gaps and promoting the scalable deployment of analytics-informed care [70], [78]. Adaptive dashboards and real-time interfaces should be co-designed with stakeholders to ensure usability, especially under resource constraints and data uncertainty, as seen in pandemic contexts and mHealth applications [57], [39], [80].

Additionally, cloud-based infrastructures and latency-aware systems must be developed to support real-time analytics in remote and low-resource settings [62], [53]. Ethical AI governance frameworks must evolve to address bias, consent, transparency, and inclusive data strategies, ensuring equitable care across populations [64], [49], [54]. Sentiment analytics and social media monitoring should be harnessed to inform public health strategies and counter misinformation [69]. Finally, educational programs in data literacy and model interpretation should be institutionalized to empower healthcare professionals with the skills necessary to confidently and effectively implement data-driven decisions [58], [76]. These combined efforts underscore the need for healthcare analytics to transition from tool-centric innovation to integrated, equitable, and sustainable healthcare transformation.

5.2 Theme 2: Patient-Centered Health Analytics

5.2.1 Cluster Definition

This thematic cluster defines healthcare analytics as the convergence of health informatics, predictive analytics, and electronic health records (EHRs) to enhance patient care, improve clinical outcomes, and optimize healthcare delivery at both individual and population levels. At its core, the cluster reflects a transformation in biomedical and clinical practice enabled by scalable, AI-driven systems that support real-time and near-real-time risk assessments, disease prediction, and personalized treatment strategies [35], [47], [61]. Health informatics emerges as a multidisciplinary infrastructure that integrates data from electronic health records (EHRs), Internet of Things (IoT) devices, imaging systems, and physiological signals to enable tailored interventions in both chronic and acute care contexts, such as diabetes, sepsis, epilepsy, and tuberculosis [43], [44], [72], [49]. These systems not only predict patient-specific risks using ensemble and deep learning models but also facilitate early diagnosis and proactive care planning across hospital-based and distributed settings [73], [61], [58], [76]. The EHR remains central as a unifying data source, enabling both granular patient-level monitoring and macro-level public health surveillance [40], [62], [74].

Additionally, federated and fog-based infrastructures allow privacy-preserving analytics across institutional boundaries, expanding the reach of healthcare analytics to low-resource and latency-sensitive environments [53], [71]. Importantly, this cluster acknowledges the dual role of informatics systems in enhancing workflow efficiency and contributing to clinician burden, as seen in associations between EHR overload and professional burnout [56]. It also acknowledges the increasing importance of interpretability and semantic clarity in analytics-driven decision-making, particularly when utilizing large language models to navigate complex clinical terminology [14], [67]. Ultimately, this thematic cluster conceptualizes healthcare analytics as a dynamic, data-centric ecosystem that bridges personalized care with system-wide transformation, underpinned by real-world applicability, clinical integration, and algorithmic transparency [79], [48], [60].

5.2.2 Current Trends

Recent trends in healthcare analytics reveal a decisive convergence of machine learning, health informatics, and big data infrastructures to support real-time, scalable, and patient-centered decision-making. A dominant trajectory is the integration of AI and ensemble models into hospital and electronic health record (EHR) systems, enabling early warning systems and personalized risk prediction across diverse clinical contexts such as diabetes, sepsis, ALS, and emergency department admissions [43], [81], [46], [63], [51]. These models are increasingly built upon multimodal data inputs, including physiological signals, imaging, digital traces, and behavioral metrics, allowing for nuanced and context-aware decision support [82], [44], [47]. Cloud-based and IoT-enabled platforms are emerging as key enablers of continuous monitoring and remote diagnostics, particularly in resource-constrained and latency-sensitive environments [62], [65], [64]. In parallel, federated and fog computing architectures are being explored to protect privacy while maintaining performance and reducing communication overhead [53]. Large language models (LLMs) are expanding the scope of health informatics by supporting clinical document interpretation, abbreviation disambiguation, and natural language-based patient engagement [67], [54]. The rise of explainable AI (XAI) is particularly notable in high-stakes applications, such as cancer and neurological disease diagnosis, where clinical trust and decision transparency are crucial [52], [14], [48].

Furthermore, process mining and human-centered dashboard design highlight the increasing integration of analytics into hospital operations and public health interfaces, fostering greater usability and stakeholder engagement [55], [57]. The COVID-19 pandemic catalyzed the evolution of health surveillance dashboards, prompting the fusion of structured and unstructured data for dynamic public health monitoring and misinformation mitigation [68], [74], [69]. Finally, the emphasis on model generalizability, particularly through population-wide validation using datasets like the UK Biobank, reflects a growing interest in equitable and globally applicable healthcare solutions [83], [77]. Collectively, these developments define a thematic trajectory toward intelligent, adaptive, and ethically grounded healthcare delivery systems.

5.2.3 Challenges

Despite rapid technological advancement, healthcare analytics continues to face persistent challenges that constrain its real-world effectiveness, clinical adoption, and equity of impact. A central concern lies in the integration of AI-driven models with electronic health records (EHRs), particularly in settings with limited data coverage, inconsistent standards, and underdeveloped infrastructures, as seen in developing countries and across heterogeneous hospital systems [39], [68], [62]. This heterogeneity contributes to diminished model generalizability and risk assessment accuracy, especially for chronic and complex conditions such as diabetes, sepsis, and ALS [43], [49], [46]. While ensemble models and large language models (LLMs) offer advanced predictive capabilities, their implementation is hindered by issues of interpretability, contextual sensitivity, and performance variability across data sources [63], [54]. Post hoc explainability methods remain insufficient to bridge the gap between algorithmic output and clinician trust, particularly when models lack embedded domain knowledge or yield ambiguous results [14], [47].

Furthermore, ethical and regulatory concerns around patient consent, data privacy, and algorithmic transparency remain unresolved, even in federated frameworks where data security must be balanced with system performance [53], [78], [60]. Real-world deployment is further impeded by infrastructural constraints, such as fluctuating network latency in IoT-enabled systems and limited processing capacity in wearable devices [64], [62]. From a user-centered perspective, the failure to co-design decision support tools with clinicians contributes to poor usability, alert fatigue, and professional burnout, especially when systems add to documentation burdens [57], [56], [75]. Models developed for adolescent health, depression detection, or abbreviation disambiguation further illustrate how demographic bias, data sparsity, and semantic complexity can distort outcomes and limit reproducibility [84], [47], [67]. Finally, the lack of standardized evaluation frameworks across institutions hampers benchmarking and validation efforts, reinforcing a fragmented ecosystem [45]. Collectively, these limitations underscore the pressing need for robust data governance, interdisciplinary collaboration, and adaptable system design to ensure scalable, interpretable, and ethically sound healthcare analytics solutions.

5.2.4 Research Opportunities and Future Directions

This thematic cluster presents a robust and multifaceted research agenda focused on advancing healthcare analytics through methodological innovation, system-level integration, and ethical alignment. Central to future efforts is the development of adaptive, explainable, and generalizable AI models that synthesize multimodal data—including structured electronic health records (EHRs), unstructured clinical texts, radiological images, and patient-generated health data—for real-time risk prediction and personalized decision support [66], [44], [43], [65]. Hybrid architectures that combine traditional statistical methods with large language models and deep learning are increasingly recognized as essential for achieving both contextual reasoning and predictive robustness in high-stakes clinical environments [63], [21], [72]. The cluster also emphasizes the importance of interoperability, calling for data harmonization across hospital systems and federated infrastructures to scale predictive capabilities and facilitate cross-institutional collaboration [39], [53], [71]. Ethical imperatives—such as mitigating bias in disease prediction algorithms, ensuring demographic fairness, and enhancing transparency in model decision-making—remain at the forefront, especially in chronic disease monitoring and mental health analytics [77], [45], [47]. Research must also prioritize longitudinal validation frameworks to assess model performance over time and across diverse populations, with particular attention to adolescent health and underrepresented groups [84], [67]. New directions include incorporating emotion recognition, social determinants, and digital footprints into predictive models for holistic care assessments [82], [40], [47]. The use of human-centered design in dashboards, auto-documentation systems, and decision aids is crucial for enhancing usability, reducing clinician burnout, and supporting real-world adoption [57], [56], [74]. Additionally, process mining and blockchain integration offer promising avenues for procedural optimization and secure, auditable data sharing [55], [79]. Ultimately, advancing this cluster will depend on unifying technical scalability, ethical responsibility, and clinical relevance, ensuring that healthcare analytics evolves into an interpretable, inclusive, and sustainable component of healthcare systems [54], [79], [46].

5.2.5 Recommendations to enhance practical value

To enhance the practical value of healthcare analytics, this thematic cluster highlights the need for co-development strategies that integrate clinical expertise, ethical oversight, and technical innovation across the entire deployment pipeline. The co-design of AI models with clinicians ensures that outputs are interpretable, actionable, and embedded within real-world clinical workflows, addressing usability concerns and supporting clinician trust [40], [41], [63]. Investment in interoperable data infrastructure—particularly the expansion of electronic health record (EHR) coverage, integration with IoT and public health platforms, and support for federated data networks—is vital to enable scalable, low-latency risk assessments across institutions and geographies [39], [62], [71], [68]. Standardized validation protocols and transparent performance metrics, including those assessing model explainability, demographic parity, and longitudinal stability, are essential to support benchmarking and multi-site adoption [69], [48], [46], [45]. Policymakers and healthcare administrators should promote adaptive public health dashboards and decision-support systems that prioritize accessibility, usability, and personalized indicators for diverse populations [57], [74], [81]. Equally important are regulatory and ethical frameworks that govern the responsible use of AI, especially in areas involving surveillance, automated documentation, and mental health diagnostics, where concerns around bias, transparency, and patient consent remain significant [78], [79], [47], [67]. Institutions should also support AI-powered EHR optimization to reduce clinician burnout and enable more efficient documentation workflows [56]. Training healthcare professionals in AI literacy and interpretability is crucial to ensure that these systems function as augmentative tools rather than disruptive technologies [75], [54]. Finally, interdisciplinary collaborations—bridging informatics, behavioral science, and policy—will be crucial in tailoring AI interventions to specific healthcare settings and populations, thereby enabling ethically grounded and clinically impactful analytics applications [44], [70], [54]. These combined strategies define a practical roadmap for transitioning healthcare analytics from experimental promise to widespread, sustainable implementation.

5.3 Theme 3: Adaptive AI for Clinical Insights

5.3.1 Cluster Definition

This thematic cluster focuses on integrating machine learning and deep learning systems into healthcare analytics, highlighting their crucial role in transforming prediction, detection, and decision-making processes across both clinical and subclinical contexts. It encompasses a wide range of methodologies—including support vector machines, decision trees, random forests, neural networks, and ensemble models—that operationalize structured and unstructured data from electronic health records, wearable sensors, medical images, and digital traces to support diagnosis, risk stratification, and outcome forecasting [51], [43], [63], [81], [76], [49]. Applications span both individualized and population-level interventions, from predicting adolescent obesity using behavioral and anthropometric features [45] to detecting breast cancer through explainable mammography models [52] and forecasting emergency room admissions [63]. The emphasis on feature extraction and optimization—whether through manual selection or algorithmic techniques—underpins the analytical rigor of this domain [50], [77]. Notably, ensemble and hybrid models are favored for their adaptability in capturing the complexity and heterogeneity inherent in healthcare data, enabling systems to deliver accurate and context-aware predictions in dynamic environments [21], [47]. Real-time deployment frameworks increasingly rely on the fusion of historical and live data streams to inform proactive interventions, such as monitoring sepsis risk from physiological signals [43] or forecasting treatment response trajectories in mental health [41]. The cluster also reflects an evolution toward interpretable AI, as tools like SHAP, LIME, and occlusion maps are embedded into learning workflows to enhance transparency and clinical trust [14], [48]. Moreover, the use of predictive analytics extends beyond clinical settings into public health, where models contribute to forecasting healthcare expenditure [2] and detecting behavioral health indicators on social media [47], [82]. Collectively, this cluster marks a significant shift from retrospective, rule-based systems to scalable, personalized, and explainable predictive architectures that reshape healthcare delivery through data-driven intelligence [78], [60], [79].

5.3.2 Current Trends

Recent developments in healthcare analytics reflect a dynamic convergence of hybrid, ensemble, and deep learning models, all aimed at improving predictive accuracy, scalability, and real-time clinical utility. A dominant trend across the literature is the integration of multimodal data—ranging from medical images and physiological signals to behavioral logs and textual records—into adaptive frameworks that leverage the strengths of multiple algorithms, such as support vector machines, random forest, logistic regression, and deep neural networks [21], [50], [77], [51]. These multi-model systems are increasingly deployed to address complex prediction tasks, including stroke, heart disease, sepsis, adolescent obesity, and epilepsy [61], [45], [43], [65], [49]. Deep learning architectures, especially convolutional and recurrent neural networks, are widely applied for image analysis and temporal modeling, supporting use cases such as breast cancer diagnostics and 3D neuroimaging [52], [14], [48]. Simultaneously, hybrid systems that integrate deep learning with knowledge graphs or domain ontologies are gaining momentum, offering both improved performance and enhanced explainability, particularly in sensitive contexts such as depression detection and drug interaction modeling [59], [47]. Another salient trajectory involves federated learning and fog computing frameworks, which facilitate privacy-preserving, low-latency analytics across distributed IoMT environments [53], [64], [62]. In addition, retrieval-augmented large language models, such as GPT-4, are being evaluated for their ability to complement traditional predictive pipelines by reducing feature dimensionality while retaining interpretability [63], [83]. Studies also underscore the importance of feature selection and dimensionality reduction techniques in optimizing model complexity and supporting performance generalization across various health conditions and patient populations [72], [76]. Metaheuristic optimization strategies further enhance pipeline adaptability in resource-constrained or imbalanced data scenarios [50]. Across applications—from tuberculosis prognosis and diabetes forecasting to healthcare expenditure modeling—the shared emphasis is on building robust, interpretable, and cloud-integrated analytics systems that align technical sophistication with the operational demands of modern healthcare [44], [2], [81], [65].

5.3.3 Challenges

Despite significant advances in healthcare analytics, this thematic cluster continues to face structural, technical, and ethical limitations that hinder large-scale adoption and clinical integration. A core challenge lies in the generalizability of models trained on heterogeneous, imbalanced, or siloed datasets, which compromises their reliability when applied to new populations or care settings [73], [50], [43], [63]. Disparities in data quality, such as those observed in COVID-19 surveillance or abbreviation disambiguation, highlight the fragility of predictive systems built without harmonized and representative data sources [68], [67]. Deep learning models, while powerful, remain highly dependent on extensive labeled datasets and high-performance computing—resources not uniformly available, particularly in under-resourced or federated environments [64], [53], [65]. Although ensemble and hybrid models demonstrate improved performance, they introduce greater complexity and opacity, thereby exacerbating issues related to interpretability, clinical trust, and deployment transparency [52], [46], [76]. These concerns are particularly acute in high-stakes contexts such as sepsis prediction, depression detection, or breast cancer diagnosis, where black-box models fail to offer actionable explanations [52], [47], [48]. Compounding this are operational barriers, such as system latency, alert fatigue, and limited interoperability across EHR platforms and IoT devices, all of which challenge the real-time deployment of AI-driven solutions [75], [62], [65]. Ethical and regulatory issues also remain unresolved, including privacy concerns in data sharing, standardization of evaluation metrics, and the accountability of language models when applied without contextual grounding [78], [66], [63]. Moreover, reliance on static features and lack of dynamic model updating often result in reduced clinical relevance over time, especially in chronic disease monitoring [41], [49]. To address these interconnected limitations, future systems must prioritize interpretable architectures, robust validation frameworks, adaptive feature modeling, and equity-aware designs that account for demographic variability and infrastructural constraints [81], [83], [44], [72].

5.3.4 Research Opportunities and Future Directions

Several promising research pathways emerge from this thematic cluster, pointing toward a future of more adaptable, interpretable, and equitable healthcare analytics. First, hybrid ensemble frameworks—combining voting, stacking, and probabilistic reasoning—are being refined to improve the robustness and accuracy of disease prognosis, patient monitoring, and therapy response prediction, particularly when applied to multimodal and dynamic datasets [51], [43], [41]. A parallel direction focuses on lightweight deep learning architectures and edge-AI strategies tailored for IoMT and wearable devices, which can support real-time monitoring in low-resource or decentralized environments while addressing latency and energy constraints [64], [65], [53]. Knowledge-enhanced systems, including those built on domain ontologies or knowledge graphs, are increasingly deployed for pharmacological surveillance, depression detection, and public health forecasting, offering better contextualization and more interpretable decision pathways [59], [47], [44]. The inclusion of retrieval-augmented generation in large language models (LLMs), combined with structured numerical features, is also gaining traction for bridging the gap between unstructured and structured data paradigms in clinical applications [63].

Furthermore, multimodal learning architectures—fusing physiological, visual, behavioral, and demographic features—are emerging as practical tools for emotion recognition, comorbidity modeling, and treatment adjustment in complex care settings [82], [44], [21]. Explainable AI remains a foundational concern, with calls to embed local and global interpretability mechanisms into both ensemble and deep models to build clinician trust, particularly in high-stakes domains such as breast cancer detection and cardiovascular risk assessment [52], [14], [48]. Researchers are also exploring the integration of predictive systems into electronic health records (EHRs) and telemedicine platforms to enhance real-time decision support and optimize care delivery workflows [58], [48]. Finally, to address disparities in model performance across populations, ongoing work emphasizes fairness-aware algorithms, benchmarking across diverse cohorts, and data augmentation strategies, including generative models and oversampling, to confront data imbalance and ensure equitable health outcomes [76], [49], [72], [67]. These directions collectively signal a shift toward more contextual, scalable, and inclusive predictive healthcare infrastructures.

5.3.5 Recommendations to enhance practical value

To maximize the real-world impact of healthcare analytics, this thematic cluster emphasizes the need for predictive systems that are interpretable, scalable, and context-aware. Central to this effort is the co-design of modular, plug-and-play architectures that integrate seamlessly into existing hospital, telehealth, and public health infrastructures, allowing for rapid deployment and real-time intervention across diverse care environments [73], [81], [62], [65]. Developers should prioritize ensemble and federated learning frameworks that adapt to local data conditions while preserving privacy, especially in edge and wearable IoMT settings [64], [53]. Equally important is embedding these models into electronic health record systems and open-source platforms to enable continuous data ingestion and cross-institutional reproducibility [39], [71], [78]. Model explainability must be addressed through the integration of interpretability modules—such as those using domain-informed or SHAP-based techniques—to increase clinician trust, particularly in high-stakes domains like diagnostic imaging and sepsis prediction [52], [60], [43], [47], [14]. To ensure clinical utility, predictive tools must be validated longitudinally across diverse populations and geographies, as evidenced by performance discrepancies in obesity and cardiovascular risk prediction [83], [45]. These efforts should be supported by standardized benchmarking protocols, usability testing, and continuous evaluation of metrics like alert fatigue, time-to-intervention, and patient satisfaction [75], [58], [49]. Investments in clinician training and stakeholder-centered design can further align model interfaces—such as interpretable alerts, treatment recommendations, and emotion-aware systems—with clinical workflows and user expectations [82], [79], [77]. Finally, cross-disciplinary collaboration and policy incentives are vital to address ethical, regulatory, and deployment challenges, including fairness, data governance, and cost-effectiveness, ultimately ensuring that predictive analytics becomes a sustainable, equitable, and actionable component of healthcare delivery [66], [76], [48], [46].

5.4 Theme 4: Demographic Health Analytics

5.4.1 Cluster Definition

This thematic cluster centers on population-based healthcare analytics grounded in retrospective, cohort, and longitudinal clinical study designs, with a strong emphasis on demographic stratification—particularly gender, age, and epidemiological risk factors—as core predictors of health outcomes. Studies in this domain utilize large-scale participant data, including structured clinical records and unstructured patient-reported outcomes, to model disease prevalence, incidence, and progression in adult populations [14], [76], [48]. Predictive models in this cluster integrate demographic and clinical variables to support personalized risk assessment and targeted intervention strategies. For example, ensemble models for diabetes and stroke prognosis include age, gender, comorbidities, and physiological parameters as critical input features [51], [61]. Similarly, machine learning approaches to emergency admissions and breast cancer detection rely on both structured (e.g., hospital records) and unstructured (e.g., imaging, self-reports) datasets to refine accuracy and enable gender-specific or age-specific risk profiling [63], [85]. Mental health analytics, such as depression prediction among adolescents and young adults, highlight the importance of incorporating socioeconomic, behavioral, and regional variables into age-stratified models [84], [45]. This cluster also contributes to cardiovascular and obesity-related analytics, demonstrating how demographic segmentation improves generalizability across multicenter and global datasets [83], [45]. Beyond clinical forecasting, these studies inform public health policy by operationalizing large-scale informatics systems for population surveillance and the design of interventions. The application of deep learning, ensemble methods, and AI-enhanced epidemiological tools extends the scope of analysis from individual prognosis to population-level modeling, facilitating actionable insights in both clinical and community contexts [81], [46], [49]. In sum, this cluster is characterized by its systematic integration of demographic and clinical indicators, standardized epidemiological methods, and computational tools to enhance evidence-based, risk-oriented healthcare delivery and public health planning.

5.4.2 Current Trends

This thematic cluster encompasses the convergence of epidemiological methods, machine learning models, and population-level analytics to enhance understanding and prediction of health outcomes across diverse demographic strata. A prominent trend involves embedding age, gender, and comorbidity data into ensemble frameworks that support disease prediction, including diabetes, sepsis, cardiovascular conditions, and mental health disorders [51], [43], [63], [61], [77], [83]. These models leverage retrospective cohort studies, real-time monitoring systems, and wearable IoMT devices to contextualize risks, personalize interventions, and enhance generalizability across healthcare systems [64], [62], [65]. Increasingly, population health analytics integrates stratified cohort designs and age-adjusted metrics to capture intergenerational shifts and demographic disparities, especially in adolescent obesity and depression, where gender-specific responses are evident [45], [84], [41]. There is also a growing emphasis on combining structured clinical variables with self-reported and behavioral data, such as perceived health status and self-efficacy, which improves the explanatory power of models in chronic disease management and mHealth adoption studies [80], [58]. Deep learning and ensemble methods are commonly used to model dynamic risk, correct class imbalances, and identify early treatment non-responders, particularly in longitudinal data applications like Alzheimer's and depression progression prediction [48], [46], [41]. In parallel, health informatics platforms are incorporating user-centered visualizations and real-time feedback to accommodate the diverse needs of populations, including ethnicity imputation and the social determinants of health [57], [56]. The inclusion of spatiotemporal analytics—such as those applied in COVID-19 surveillance—signals an emerging interest in geo-epidemiological modeling that accounts for both time-based and location-based variability [68]. Collectively, these studies reveal a shift from traditional demographic control variables to active, explanatory components in predictive frameworks, reinforcing the role of healthcare analytics in precision public health and equitable intervention planning [40], [84], [49].

5.4.3 Challenges

Despite notable methodological advancements, this thematic cluster continues to grapple with persistent challenges related to data heterogeneity, demographic sensitivity, and real-world applicability. A central issue is the inconsistency and fragmentation of demographic variables—particularly gender and age—across datasets, which limits the generalizability of models and undermines the validity of cross-population comparisons [51], [83], [84]. While ensemble and deep learning models show promise in predicting outcomes like diabetes, cardiovascular risk, and obesity, their reliance on detailed and often unevenly distributed demographic inputs complicates their deployment in underrepresented or resource-constrained settings [51], [45], [64]. Retrospective and cohort-based studies, although valuable, often suffer from selection bias, recall bias, and inconsistent variable definitions, making it challenging to extract causal inferences or inform equitable intervention design [39], [77], [61]. Studies involving adolescents, women, or racially minoritized populations face additional barriers due to data sparsity, measurement error, or lack of standardization in cohort inclusion criteria [45], [84], [80].

Furthermore, integrating heterogeneous modalities—including wearable sensor data, self-reported attributes, and electronic health records—into coherent analytic frameworks remains a technical and conceptual challenge, particularly when transparency and interpretability are required for clinical adoption [62], [61], [41]. Large language models and AI-driven predictive systems still lack adequate contextual reasoning, reducing their robustness in dynamic, real-world healthcare environments [66], [63]. Ethical concerns also persist, including the potential misuse of sensitive demographic data, the underrepresentation of key subgroups, and the absence of standardized performance metrics that capture fairness and inclusivity [76], [56], [54]. Moreover, public-facing dashboards and cohort visualizations are often not optimized for usability or clarity, which limits their impact on public health communication and engagement [57]. Addressing these challenges will require interdisciplinary collaboration, standardized data protocols, and inclusive model validation strategies to ensure equitable and actionable healthcare analytics.

5.4.4 Research Opportunities and Future Directions

Emerging directions in this thematic cluster underscore the importance of integrating multimodal, longitudinal, and demographically sensitive data into predictive healthcare analytics. A growing consensus emphasizes refining wearable and IoMT technologies through federated learning to capture real-time physiological signals while preserving patient privacy, particularly among underrepresented populations segmented by age, gender, and socioeconomic status [64], [43], [76]. Studies increasingly prioritize hybrid cohort models that combine structured epidemiological data with unstructured sources such as patient narratives, social media expressions, and environmental indicators to generate deeper, context-aware risk insights [63], [69], [49]. Longitudinal research designs are gaining traction for tracking behavioral and clinical outcomes over time, revealing complex age–gender interactions in conditions such as depression, obesity, and cardiovascular risk [84], [45], [83]. Advances in explainable AI offer promising tools to enhance transparency and bridge the gap between model interpretability and clinical decision-making, especially when applied to cohort-based interventions and personalized informatics tools [14], [60], [40]. The development of generalizable models that maintain predictive accuracy across cohorts with diverse demographic profiles remains a critical objective, supported by cross-cohort validation and cloud-enabled, participant-centered data collection frameworks [72], [58], [65]. Research must also address the fragmentation of cohort datasets and improve stratification techniques to ensure equity in predictive modeling, particularly for women, older adults, and racially marginalized groups [37], [53]. Further opportunities lie in incorporating social determinants of health and self-reported metrics such as self-efficacy and digital health engagement to capture the behavioral and psychosocial dimensions of health disparities [80], [56]. Finally, the use of geo-epidemiological visualizations and real-time public health dashboards can support timely policy responses and personalized interventions based on spatial and temporal risk patterns [68], [74]. Collectively, these research directions point toward a more inclusive, adaptable, and evidence-driven future for healthcare analytics.

5.4.5 Recommendations to enhance practical value

To translate analytical innovations into actionable and equitable healthcare strategies, this thematic cluster calls for standardized, interoperable, and demographically inclusive frameworks. A key priority is the harmonization of demographic profiling protocols—particularly those capturing gender, age, and socio-environmental factors—to ensure comparability across retrospective studies, cohort analyses, and real-time monitoring systems [81], [46], [68]. Health systems must invest in unified data infrastructures that support cross-cohort linkage and longitudinal surveillance, thereby enabling predictive models to function across diverse clinical environments and patient populations [39], [48]. Ensemble and Bayesian models, while powerful, require enhanced interpretability through explainable AI techniques to be meaningfully integrated into clinical workflows and health policy applications [61], [73], [14]. Training clinical professionals in the ethical use of AI and transparent decision-support systems—especially in contexts involving gender-sensitive or age-specific care—is essential for ensuring trust and accountability [63], [66], [76]. To improve inclusion, data collection protocols should minimize missing demographic variables and enhance the representation of underrepresented groups such as older adults, gender minorities, and socioeconomically marginalized populations [57], [46], [80]. Embedding predictive tools, such as DeepHealthNet and GPT-4, within electronic health records and digital health platforms can facilitate early intervention, particularly when guided by age-adjusted and gender-informed metrics [45], [83], [77]. Policymakers should prioritize equity-driven funding and regulatory support for adaptive dashboards that provide real-time visualizations of stratified health risks, improving responsiveness in both public health and primary care settings [74], [58], [65]. Cross-sector collaboration among epidemiologists, data scientists, ethicists, and clinicians will be vital to bridging population-level analytics with personalized intervention pathways, translating complex models into context-aware, scalable, and ethically grounded healthcare solutions [84], [69], [56]. These coordinated efforts can ensure that healthcare analytics systems are not only technically sophisticated but also socially responsive and clinically transformative.

5.5 Theme 5: Digital Mental Health Surveillance

5.5.1 Cluster Definition

This thematic cluster focuses on the intersection of mental health, digital expression, and computational analytics, where social media platforms and internet-based environments serve as primary sources of data for understanding psychological well-being. Digital traces—textual, behavioral, and relational—left by individuals on platforms like Twitter, forums, and social networking sites offer a rich, unsolicited stream of information reflecting emotional states, symptoms, attitudes, and mental health risks [82], [84], [69]. Natural language processing (NLP) and machine learning, including large language models (LLMs), enable the extraction and interpretation of mental health-related topics and behaviors from unstructured online content, facilitating scalable, real-time surveillance of public sentiment and individual distress signals [47], [54], [63]. Studies in this cluster demonstrate how age, gender, and behavioral patterns influence the expression of symptoms online, with applications ranging from depression detection to the design of internet-based therapeutic interventions [45], [41], [80]. The incorporation of mobile health platforms, user self-reporting, and early treatment data expands this domain from passive observation to active support and clinical decision-making, making social networking both a diagnostic space and a therapeutic interface [41], [80]. This approach signifies a paradigm shift from traditional, clinic-based mental health monitoring toward digital-first, context-aware, and population-wide analytics that leverage real-time, user-generated content [66], [63]. While offering expanded coverage and early detection capabilities, this model also raises methodological and ethical challenges, particularly around data privacy, validation of computational inferences, and the inclusivity of marginalized populations [64], [43]. Overall, the cluster defines an emerging interdisciplinary field where informatics, psychology, and public health converge to interpret mental health as a dynamically mediated phenomenon—shaped by individual expression, social interaction, and the information flows embedded in modern digital ecosystems.

5.5.2 Current Trends

This thematic cluster illustrates the growing convergence of mental health analytics, natural language processing, and digital ecosystems, where social media platforms serve both as data sources and intervention environments. Across recent studies, there is an evident shift from passive monitoring to active engagement, with digital traces—such as tweets, posts, and forum interactions—being mined for early detection of depression, anxiety, and emotional distress using advanced computational models, including LLMs and domain-informed neural architectures [47], [54], [66]. These models increasingly incorporate context-aware features, explainability mechanisms, and multimodal inputs (e.g., textual, visual, and physiological) to enhance their predictive accuracy and clinical relevance [82], [67], [51]. The application of generative AI tools, like ChatGPT, exemplifies the expansion of conversational agents into mental health education, self-help, and support services [66]. Concurrently, researchers are integrating social data with clinical datasets—such as electronic health records—to bridge the gap between formal diagnostics and informal symptom expression in online spaces [63]. Retrospective and longitudinal analyses increasingly target adolescent and young adult populations, especially females aged 20–24, where symptoms of depression and emotional burden are most prominent [84], [45]. Beyond individual risk assessment, public sentiment analysis from platforms like Twitter has been employed to track collective responses to health interventions, highlighting geographical and demographic disparities in perception and behavior [69]. mHealth technologies and internet-based therapies are being redefined by these trends, offering personalized and scalable alternatives to traditional treatment pathways [80], [41]. Importantly, these developments underscore a shift toward socially embedded, real-time mental health monitoring, where data ethics, model transparency, and interdisciplinary collaboration are essential for the practical implementation of healthcare systems [54], [66]. As such, the cluster represents a paradigm shift in healthcare analytics, transforming user-generated content into dynamic, interpretable, and actionable insights that support both early detection and patient-centered interventions in mental health care.

5.5.3 Challenges

Despite the rapid evolution of digital mental health analytics, this cluster faces persistent methodological, ethical, and translational challenges that constrain its practical implementation. A central issue is the contextual ambiguity of language used on social media, where users often express symptoms through irony, sarcasm, or culturally coded terms, making it difficult for NLP models to generate reliable interpretations without introducing significant error or hallucinated outputs [66], [63]. Additionally, model generalizability is undermined by demographic and linguistic biases embedded in training data, with systems frequently performing inconsistently across different age groups, cultural contexts, and socioeconomic backgrounds [84], [69], [54]. The limited availability of ground truth clinical labels and standardized validation protocols for online behavioral data further complicates the evaluation and robustness of predictive systems [63], [68]. Moreover, while mHealth tools and digital platforms offer opportunities for proactive intervention, their success is often constrained by user engagement patterns, digital literacy, and the inherent limitations of self-reported data [80], [41]. From an ethical standpoint, the passive mining of public online behavior to infer mental health status raises significant concerns around consent, autonomy, and the stigmatization of individuals whose expressions may be taken out of context [82], [54]. Regulatory frameworks remain underdeveloped, and interdisciplinary collaboration among computational scientists, clinicians, and ethicists is still insufficient to ensure the responsible application of these tools [66].

Furthermore, the opacity of deep learning models—particularly those used in emotion recognition and depression detection—poses challenges for explainability and stakeholder trust, especially in clinical settings where decisions must be justified and transparent [47], [82], [67]. The reliance on unstructured digital traces without robust integration into clinical workflows may also result in fragmented or misleading interpretations that do not translate into meaningful health outcomes. Ultimately, while this cluster holds transformative potential for population-scale mental health surveillance, its sustainability depends on overcoming limitations in data quality, ethical governance, representativeness, and clinical interoperability.

5.5.4 Research Opportunities and Future Directions

Future research in mental health analytics must advance both technically and ethically to realize the full potential of integrating social media data, structured clinical information, and multimodal signals. A central priority is the development of explainable, ethically aware AI systems capable of detecting mental health indicators from informal digital content while addressing issues of interpretability, fairness, and consent [60], [47], [54]. Improving contextual embeddings and domain-specific adaptation of large language models (LLMs) is crucial for enhancing their sensitivity to cultural nuances and informal symptom expression, particularly in underrepresented populations [63], [84]. Longitudinal datasets derived from social media activity and mHealth platforms offer an opportunity to track temporal changes in mental health states and personalize interventions, as shown in studies of depression prevalence among young adults [84], [41]. Multimodal integration—combining text, images, behavioral logs, and biometric data—could lead to more robust models of mental health that bridge the gap between clinical precision and real-world relevance [82], [62]. There is also strong potential for hybrid systems that fuse unstructured internet data with structured clinical records, enhancing both predictive accuracy and clinical utility [54], [63]. From an ethical standpoint, future systems must embed standards for de-identification, consent-by-design, and user intent detection to mitigate risks of misuse and privacy violations [66], [54]. Social media should not only be treated as a data source but as an active platform for intervention, enabling just-in-time support through chatbots, psychoeducational content, and real-time feedback [82], [66]. Additionally, research should prioritize localized, culturally adaptive models and dashboards that democratize access to insights, facilitating tailored public health messaging and community-level interventions [69], [67]. Cross-lingual adaptation and interdisciplinary collaboration will be essential to developing scalable, inclusive, and socially responsive systems that align with evolving norms in healthcare delivery, mental health care, and digital ethics [54], [80].

5.5.5 Recommendations to enhance practical value

To enhance the practical value of this thematic cluster, a multi-faceted strategy must be pursued that integrates technological refinement, ethical safeguards, clinical alignment, and interdisciplinary collaboration. Central to this effort is the development of explainable and interpretable models that clarify which symptoms, behaviors, or linguistic patterns trigger detection, ensuring that outputs are actionable for both clinicians and patients [47], [54]. Embedding such models into clinical workflows, including telehealth platforms and decision support systems, will facilitate the real-world utility of mental health analytics, particularly in under-resourced or remote settings [63], [41]. To achieve this, standardized benchmarks and ethically sourced datasets representing diverse populations must be created alongside model validation protocols that account for cultural, linguistic, and demographic variability [84], [82]. Cross-sector partnerships among data scientists, mental health professionals, public health officials, and social media platforms are crucial for establishing robust data acquisition ethics, annotation guidelines, and user consent frameworks [63], [66]. Real-time dashboards that visualize topic trends, sentiment fluctuations, and early warning signals can inform public health strategies, vaccine campaigns, and mental health outreach efforts [74]. These tools must prioritize intuitive design and incorporate metadata such as platform type, posting time, and regional factors to contextualize insights. Personalized feedback systems based on social media analysis, integrated into mHealth or self-help platforms, can transform passive symptom monitoring into proactive care and engagement [80], [41].

Meanwhile, institutional investment in training programs for health professionals—covering the interpretation and application of NLP-driven analytics—will be crucial for adoption. Finally, funding agencies and journals should support interdisciplinary research that integrates informatics, behavioral science, and clinical psychiatry, enabling the development of scalable, equitable, and ethically responsible mental health analytics tools responsive to evolving societal needs [63], [67]. These recommendations are crucial for transitioning this cluster from theoretical innovation to a cornerstone of patient-centered mental health care infrastructure.

5.6 Theme 6: Ethical Analytics for Health Surveillance

5.6.1 Cluster Definition

This thematic cluster is defined by the convergence of medical informatics, public health informatics, and information systems to advance data-driven, community-accessible, and ethically grounded healthcare solutions. Central to this domain is the integration of information processing, data visualization, and evaluation frameworks that support real-time monitoring, personalized medicine, and informed public health decision-making across diverse healthcare infrastructures. From EHR-based clinical tools [63] and cloud-enabled IoT platforms for real-time monitoring [62], [65] to open-source systems like HADES for causal-effect estimation [71] and fog computing environments for safeguarding sensitive data [53], the studies demonstrate a commitment to developing interoperable, privacy-aware systems. These systems extend to predictive models for sepsis [43], diabetes [51], ALS prognosis [46], and seizure detection [72], underscoring the use of ensemble learning and machine learning techniques for individualized and population-level prediction. Equally, platforms such as public health dashboards [74], [57], mHealth tools [80], and social media-based information dissemination [69] emphasize the importance of accessibility and visual communication in empowering both professionals and the general public. The cluster increasingly aligns technical innovation with usability through human-centered design [57], interdisciplinary collaboration [70], and explainable models [14], [58], ensuring equitable access to and interpretation of health data. Biomedical informatics emerges not only as a technological conduit but also as a translational science, bridging the gap between complex data analytics and practical, inclusive healthcare interventions [60]. As illustrated in applications ranging from neurological diagnostics [48] to sexual health surveillance [57], the field promotes an ecosystem approach that links real-time health monitoring with clinical utility and policy relevance. Ultimately, this cluster reframes healthcare analytics as a multidimensional infrastructure that supports timely diagnosis, cost-efficient care, treatment personalization, and community-informed decision-making through the intelligent fusion of computational capabilities and public health imperatives.

5.6.2 Current Trends

This thematic cluster reflects a robust convergence of artificial intelligence, machine learning, federated architectures, and cloud computing within healthcare analytics, driven by the imperative to create personalized, predictive, and explainable health systems. Ensemble learning models and hybrid frameworks are increasingly applied across conditions such as diabetes [51], [81], cardiovascular disease [83], adolescent obesity [45], and sepsis [43], demonstrating a sustained focus on precision in clinical prediction [73], [61], [50]. Infrastructure, such as fog-enabled and IoT-integrated platforms [53], [65], enables remote patient monitoring and decentralized analytics while addressing challenges in data privacy and interoperability [64], [62]. The proliferation of large language models in tasks such as medical documentation, abbreviation disambiguation, and public health strategy simulation exemplifies the growing reliance on generative AI in biomedical informatics [66], [67], [54]. Parallel to this, the use of dashboards and real-time electronic surveillance systems—particularly in pandemic contexts [39], [74]—has evolved from descriptive visualizations to adaptive, interactive interfaces that inform clinical decision-making and public policy [57], [58]. The increasing interest in explainable AI techniques, including LIME and SHAP, underscores efforts to enhance the interpretability and trustworthiness of predictive systems in sensitive domains, such as breast cancer diagnosis and chronic disease management [14], [52], [76], [46].

Additionally, the integration of sentiment analysis and social media mining into health analytics offers novel avenues for tracking public opinion and behavioral responses to interventions [69]. A growing emphasis accompanies these developments on user-centered design, model transparency, and system-level evaluation frameworks that bridge technical innovation with clinical and societal utility [44], [55], [82]. Collectively, the cluster delineates a shift toward an interdisciplinary, real-time, and human-centric healthcare paradigm, where computational intelligence meets public health needs through interactive visualization, context-aware decision support, and ethical informatics.

5.6.3 Challenges

Despite promising advancements, this thematic cluster continues to face persistent and multifaceted challenges that hinder the full integration of AI-driven systems in healthcare analytics. A dominant issue is the limited contextual understanding of large language models, particularly in complex clinical environments where ambiguous expressions, imbalanced data, and domain-specific knowledge play crucial roles in decision-making [66], [47], [67]. The “black-box” nature of deep learning models contributes to skepticism among clinicians, especially in high-risk areas like diagnostic imaging and neurological monitoring, where model interpretability and validation protocols are still underdeveloped [78], [14], [48], [54]. Compounding these issues are data interoperability constraints across electronic health record systems and federated platforms, which affect both clinical surveillance and real-time public health responses [39], [71], [62], [65]. Furthermore, resource limitations—such as low computational power, latency, and communication costs—hamper the scalability of IoMT and cloud-integrated platforms, especially in under-resourced or fluctuating network settings [53], [64], [79]. The heterogeneity of datasets, coupled with missing or low-quality demographic data, poses challenges to the generalizability and fairness of models, particularly in underserved populations [57], [46], [72].

Additionally, systems designed for emotion recognition and visualization often struggle to align analytics outputs with real-world stakeholder needs due to inadequate design thinking and cross-sector validation [57], [82], [70]. The growing use of social media data also raises concerns about misinformation, consent, and the misinterpretation of public sentiment in digital spaces [69]. Finally, the field lacks standardized evaluation metrics, ethical frameworks, and deployment guidelines, resulting in inconsistent performance and limited translational impact from simulation to real-world implementation [75], [76], [68], [60]. Addressing these systemic, methodological, and infrastructural gaps is essential to achieve trustworthy, equitable, and sustainable healthcare analytics.

5.6.4 Research Opportunities and Future Directions

Future directions in healthcare analytics call for the development of interdisciplinary, adaptive, and ethically grounded information systems that integrate AI techniques with domain-specific knowledge to support personalized, predictive, and community-responsive healthcare. Combining deep learning with knowledge graphs and generative models offers promising opportunities in clinical tasks such as drug interaction prediction and clinical language processing [59], [67]. To ensure scalability, equity, and trustworthiness, research must prioritize the design of explainable AI (XAI) frameworks that support interpretability, particularly in sensitive contexts such as mental health and rare diseases [63], [14], [54]. The integration of multimodal data—including text, imaging, sensor streams, and genetic information—through federated and cloud-edge architectures enables real-time analytics while preserving privacy and reducing latency in low-resource settings [64], [65]. Simultaneously, the advancement of adaptive ensemble models for chronic and acute conditions such as ALS or stroke must be accompanied by rigorous longitudinal evaluation across diverse cohorts to ensure generalizability and clinical relevance [46], [61], [41]. Public health dashboards should evolve into modular, human-centered platforms incorporating automated explanations, sentiment tracking, and equity-focused metrics to inform decision-making in underserved communities [55], [57], [69].

Additionally, research should address standardization of data models, interoperability, and evidence-based design through implementation trials and cross-platform validation [68], [62], [70]. The fusion of cohort-specific modeling, visual analytics, and emotion-aware feedback into decision support systems further enhances both technical accuracy and user trust [47], [82], [48]. Finally, interdisciplinary collaborations among informatics, biostatistics, behavioral science, and public health policy are essential to align technological innovation with ethical stewardship and sustainable health system integration [71], [60], [79].

5.6.5 Recommendations to enhance practical value

To enhance the practical utility of healthcare analytics, this thematic cluster emphasizes the importance of interoperable platforms, human-centered design, and equity-driven implementation frameworks that can be scaled across diverse clinical and public health contexts. Building modular architectures aligned with cloud-based infrastructures, EHR systems, and IoT devices is essential to support real-time data processing and integration in resource-limited settings [62], [65], [58]. Investments in training programs must equip both clinicians and public health professionals with competencies in AI literacy, data ethics, and interpretation of visual outputs, bridging technical advances with practical decision-making [60], [76], [56]. Embedding explainable AI into development pipelines—particularly in sensitive domains such as depression detection and diagnostic imaging—can enhance transparency and foster trust in ML-assisted decisions [47], [67], [78]. Equally important is the use of interactive dashboards with standardized visualization principles, enabling non-expert users, policymakers, and clinicians to access timely and actionable insights for population health management [74], [46], [82]. Collaborative models such as federated learning and open-source platforms like HADES should be promoted not only for technical scalability but also for ensuring privacy-preserving analytics and community-focused evaluation principles [71], [53], [39]. To address concerns of bias, model accountability, and misinformation, governance frameworks must require performance reporting across diverse populations, integrate sentiment tracking for crisis communication, and involve stakeholders in co-design and validation processes [66], [69], [58]. Standardized evaluation frameworks, grounded in implementation science, are also necessary to assess the real-world impact of ensemble learning models and predictive dashboards [75], [61], [50]. Finally, this cluster advocates for a more substantial alignment between interdisciplinary training, regulatory science, and policy engagement to ensure that healthcare analytics innovations are usable, trustworthy, and capable of supporting equitable transformation at scale [70], [40], [54].

5.7 Theme 7: Personalized Care through Data Analytics

5.7.1 Cluster Definition

This thematic cluster reflects a comprehensive integration of machine learning, predictive analytics, and digital technologies to transform disease diagnosis, treatment, and outcome prediction within healthcare systems. At its core, the cluster emphasizes a shift from static, reactive care to adaptive and anticipatory models that rely on both structured and unstructured health data to personalize patient treatment plans and monitor clinical conditions in real time [41], [62], [65]. By deploying AI-enhanced decision support systems, ensemble learning models, and real-time monitoring platforms, healthcare providers can predict clinical deterioration and tailor interventions to individual patient trajectories across a variety of conditions—including epilepsy [72], Alzheimer’s disease [48], diabetes [81], cardiovascular disease [83], breast cancer [52], and adolescent obesity [45]. The increasing role of large language models in medical documentation and communication workflows reinforces the relevance of computational systems in daily clinical practice [66], while wearable devices and mHealth platforms enable continuous feedback loops and self-management strategies for chronic conditions [70]. Furthermore, digital diagnostics now incorporate multimodal data—from radiology, microbiology, behavioral traces, and even sentiment analysis—to enhance accuracy and enable dynamic assessment of evolving patient conditions [44], [47], [82]. Across these developments, the cluster demonstrates that patient outcomes are increasingly understood through a multidimensional lens encompassing not only clinical endpoints but also psychological well-being, treatment satisfaction, and real-time physiological feedback [75], [52], [82]. This evolving landscape is grounded in ethical design, explainability, and scalability [76], making predictive healthcare both personalized and accountable. Ultimately, the cluster represents a paradigmatic shift in healthcare analytics—one that bridges computational intelligence and clinical insight to construct flexible, patient-centered care systems responsive to individual and population-level health dynamics [73], [78], [49].

5.7.2 Current Trends

Recent developments in healthcare analytics underscore a decisive shift toward integrating machine learning, deep learning, and large language models for enhanced disease prediction, early diagnosis, and individualized treatment pathways. Ensemble learning and hybrid models are increasingly outperforming single-model approaches, providing improved accuracy in predicting chronic conditions such as diabetes [51], stroke predisposition [61], and ALS [46] while also enabling multi-disease forecasting through optimized data pipelines [50], [73]. Real-time monitoring systems powered by wearable devices, cloud computing, and IoT platforms are becoming central to managing acute conditions like sepsis [43] and supporting scalable, remote healthcare infrastructures [58], [65], [62]. The integration of multimodal patient data—from radiological and microbiological inputs to behavioral and social media traces—marks a paradigm shift toward holistic patient modeling, as seen in tuberculosis care [44] and depression detection via mobile health (mHealth) platforms [41], [47], [80]. In parallel, large language models are transforming biomedical informatics by enhancing the precision of diagnostic tasks, including clinical abbreviation disambiguation [67], and are rivaling traditional risk scoring systems in disease forecasting [63], [83]. These trends are accompanied by growing attention to explainable AI, particularly in domains demanding high interpretability, such as fetal ultrasound classification [14], Alzheimer's detection [48], and breast cancer diagnosis [52]. Privacy-preserving approaches, such as federated learning [53], are gaining prominence for decentralized and secure analytics, while frameworks like HADES support scalable observational research on electronic health records [71]. Overall, the convergence of computational intelligence, clinical decision support, and patient-centered technologies reflects a broader movement toward predictive, transparent, and context-aware healthcare systems, with real-world applications extending across high-burden, chronic, and acute diseases [21], [82], [54]. These innovations collectively redefine how diagnoses and treatments are tailored, monitored, and adjusted, making data-driven personalization a foundational component of modern healthcare delivery.

5.7.3 Challenges

Despite significant technological advances, this thematic cluster continues to face substantial challenges that hinder the widespread clinical adoption of healthcare analytics. Central among these is the lack of interpretability in machine learning models, which reduces clinician trust, particularly in high-stakes applications like prenatal imaging and diagnostic decision-making [14], [60]. The black-box nature of deep learning systems remains unresolved despite the development of consistently applied explainable AI frameworks. Furthermore, data-related limitations such as imbalance, scarcity, and heterogeneity across populations impede model training, especially for underrepresented conditions or rare diseases [52], [72], [46]. Real-world deployment of AI models continues to suffer from performance degradation compared to controlled environments, reflecting the complexity of dynamic clinical settings and highlighting persistent gaps in implementation science [63], [75]. Ethical concerns—including patient privacy, data governance, and accountability in automated decisions—remain inadequately addressed in large-scale systems [66], [54], particularly when using wearable or federated learning frameworks in resource-constrained environments [53], [43]. Infrastructure demands such as high computational costs, network latency, and EHR integration further limit scalability and interoperability [62], [65]. In practice, even when predictive models are robust, systemic barriers such as clinician burnout due to EHR overload [56] , variable response to digital interventions [41], and context-dependent mHealth efficacy [80] can reduce their impact. Specific challenges in clinical abbreviation disambiguation [67], drug-drug interaction prediction [59], and treatment failures in chronic disease care [44] illustrate the limitations of current AI-driven strategies in addressing nuanced clinical realities. Fragmentation of health data, inconsistent coding standards—as seen in COVID-19 surveillance [39]—and lack of universal evaluation metrics undermine standardization efforts across the board. Collectively, these persistent obstacles call for interdisciplinary collaboration, human-centered design, and rigorous validation protocols to ensure that data-driven healthcare solutions are not only technologically sophisticated but also ethically grounded, clinically practical, and contextually responsive.

5.7.4 Research Opportunities and Future Directions

Future research in healthcare analytics should prioritize the development of interoperable, ethically aligned, and context-aware systems that integrate multimodal data—including imaging, physiological signals, electronic health records (EHRs), and patient-reported outcomes—to support real-time, individualized care. Emphasis must be placed on refining large language models for clinical use by incorporating contextual grounding, explainability, and concept disambiguation techniques, especially for high-stakes decision-support tasks involving disease diagnosis and treatment planning [63], [67], [60]. Hybrid modeling approaches that merge statistical, deep learning, and attention-based frameworks show promise in managing emotionally and physiologically complex conditions [82], while patient-specific ensemble learning can be extended beyond diabetes to predict outcomes across diverse chronic diseases [51], [81]. Federated and split learning architectures—such as FedSL—enable privacy-preserving analytics in distributed IoMT ecosystems, offering scalability in low-resource settings [64], [65]. Simultaneously, cloud-IoT integration and blockchain-backed data systems present viable paths for secure, scalable patient monitoring and long-term outcome tracking [79]. There is also a need to investigate early treatment predictors, such as therapeutic alliance and symptom change, to personalize interventions in mental health care [41]. Tools such as HADES facilitate transparent observational research at scale, supporting generalizability and reproducibility across cohorts [71]. In diseases with high variability—such as tuberculosis, Alzheimer's, and ALS—future efforts must incorporate adaptive ensemble methods and longitudinal modeling to reflect dynamic disease trajectories [44], [48], [46]. Addressing cohort effects in adolescent or gender-specific conditions like depression and obesity remains essential for equitable model performance [84]. Equally, expanding mHealth research to examine personal and environmental moderators, such as self-efficacy and provider engagement, will be critical to enhancing intervention uptake [80]. Ultimately, bridging prediction and action will require the design of explainable, human-centered systems that are aligned with clinical workflows and policy frameworks and validated through real-world trials and interdisciplinary collaboration [75], [70], [76].

5.7.5 Recommendations to enhance practical value

To enhance the practical impact of healthcare analytics in disease diagnosis and patient treatment, a multifaceted approach centered on co-design, clinical integration, and scalability is essential. Collaborative development of predictive tools with clinicians ensures that models align with real-world workflows and improve usability in decision-making contexts [63], [77]. Investments in explainable AI are necessary to build trust, particularly in high-stakes environments such as hospital admissions and chronic disease management [52], [76]. Cloud-based and wearable infrastructures should be optimized for low-latency, continuous monitoring, especially in primary care and resource-limited settings [62], [43], [65]. Ethical AI deployment requires standardized frameworks for data sharing, privacy preservation, and regulatory compliance, including federated models and open-access platforms [71], [53], [39].

Additionally, implementation science frameworks must be applied to anticipate barriers and facilitators across diverse healthcare environments [75]. Equity-focused design mandates that models account for cohort-specific disease burdens—such as gender and adolescent disparities in depression and obesity—and integrate social determinants of health [84], [45]. Interventions should promote patient self-efficacy and incorporate dynamic features, such as lifestyle feedback and emotional state monitoring, to personalize care [82], [81], [80]. To mitigate clinician burnout and support workforce sustainability, AI-driven automation should be applied to EHR workflows in a manner that enhances rather than burdens clinical tasks [56]. Education and training programs for healthcare professionals are crucial for enhancing AI literacy and promoting its adoption [78], [60]. Finally, robust evaluation metrics—including real-world validation and transparency in algorithmic decision-making—must be embedded throughout model development pipelines to ensure accountability and reliability [61], [79], [72]. These strategies, underpinned by interdisciplinary collaboration and patient-provider co-design, are necessary to transition healthcare analytics from experimental models to actionable, equitable, and sustainable systems that improve patient outcomes across diverse contexts.

5.8 Theme 8: AI-Driven Insights for Outbreak Response

5.8.1 Cluster Definition

The thematic cluster defined by “COVID,” “pandemic,” and “outbreaks” captures the emergence of healthcare analytics as a critical infrastructure for managing large-scale health emergencies through real-time surveillance, predictive diagnostics, and decentralized response mechanisms. Central to this cluster is the integration of Electronic Health Records (EHRs) into outbreak monitoring systems, exemplified by near real-time surveillance tools that facilitated the tracking of confirmed and suspected COVID-19 cases and guided resource allocation in developing countries [39]. Equally pivotal is the role of distributed architectures, such as fog computing and federated learning models like FedHealthFog, which balance energy efficiency and latency sensitivity while safeguarding data privacy [53]. These systems reflect a broader shift from centralized to edge-centric models capable of processing high-velocity, heterogeneous data in outbreak scenarios. Complementing these efforts, public health dashboards have become essential tools, incorporating interface design, data visualization, and infrastructure logic to communicate actionable insights during crises [74]. Predictive models have been deployed across remote patient monitoring platforms for conditions like sepsis and chronic diseases, supported by wearable devices and cloud-based analytics [43], [58], [65]. COVID-19 analytics has also expanded to include spatial and temporal uncertainty analysis in epidemiological datasets [68], sentiment analysis of public discourse surrounding vaccines [69], and human-centered design in public-facing digital tools [57]. The pandemic has accelerated the use of AI in diagnostic tasks, emphasizing the development of interpretable and explainable models to enhance clinical trust in critical decision-making [14], [48]. Furthermore, the crisis catalyzed the expansion of telehealth, smart care platforms, and remote monitoring systems, extending care beyond traditional clinical settings [70]. Collectively, this cluster reflects an interdisciplinary convergence of health informatics, epidemiology, and data science, unified by the goal of enabling proactive, transparent, and equitable responses to global health crises through agile, technology-enabled healthcare analytics frameworks.

5.8.2 Current Trends

Recent developments in healthcare analytics reveal a systemic transformation driven by the need to manage pandemics through real-time, scalable, and intelligent infrastructures. Central to this evolution is the integration of hospital-level and national data systems, exemplified by EHR-based surveillance platforms that enable automatic stakeholder notifications and inform critical logistics such as bed allocation and resource deployment during COVID-19 outbreaks [39]. Simultaneously, cloud-enabled wearable devices are playing an increasingly important role in monitoring real-time symptoms and predicting acute conditions, such as sepsis [43]. Meanwhile, federated learning architectures like FedHealthFog significantly reduce latency and energy demands, ensuring sustainability and responsiveness in decentralized environments [53]. The convergence of these systems with IoT and AI enables predictive modeling even in the presence of incomplete datasets, shifting the paradigm from retrospective analysis to proactive outbreak surveillance [62]. Public dashboards, designed with attention to user interface and infrastructure, have become essential tools for disseminating information and guiding decision-making under pressure [74]. Efforts to quantify uncertainty in pandemic data, such as comparisons among CDC, JHU, and NYT datasets [68], highlight the demand for transparency and data reliability. Parallel trends involve integrating sentiment analysis from social media platforms to address vaccine hesitancy and inform public health messaging [69]. Moreover, the pandemic accelerated interdisciplinary collaboration, as public health agencies, informaticians, and policymakers jointly deployed AI-powered tools for smart care and early detection in various domains, including heart failure and diabetes [70], [81]. This urgency also catalyzed the adoption of LLMs to support patient interaction, documentation, and diagnostic reasoning in real-time clinical contexts [54]. The resulting frameworks emphasize usability, ethical alignment, and context-aware design, aiming to maintain clinical efficacy under conditions of uncertainty [57], [14]. Ultimately, the COVID-19 pandemic served not only as a catalyst for digital innovation but also as a defining moment that reoriented healthcare analytics toward resilience, inclusivity, and adaptive, AI-enhanced public health strategies.

5.8.3 Challenges

Despite promising advancements, this thematic cluster continues to face substantial challenges that constrain its real-world impact during pandemics. Foremost among these are barriers to data integration and standardization, with inconsistencies in formats and reliability across global datasets undermining coordination and policy-making efforts [68], [74]. While EHR-based surveillance systems have enabled automated notifications and resource allocation [39], their effectiveness is limited by uneven infrastructure, data-sharing agreements, and gaps in population-level coverage. In edge and fog computing contexts, the high communication cost and computational demands of federated learning models strain low-resource environments, complicating rapid deployment during emergencies [53]. Wearable-based real-time monitoring systems, though promising, raise concerns over scalability, data reliability, and privacy, especially in underserved settings where network instability and limited connectivity persist [43], [62].

Additionally, the integration of heterogeneous data streams—ranging from physiological sensors to clinical records and social media—into coherent predictive systems remains technically unresolved [58], [65]. Public trust issues compound these challenges: AI models often lack interpretability, which reduces clinician confidence and hinders adoption in high-stakes contexts, such as prenatal diagnostics or seizure detection [14], [48], [72]. Although interpretability tools have emerged, standard practices for explanation and validation are still lacking, especially when applying large language models to sensitive health data [76], [54]. Moreover, real-time analytics require robustness under fluctuating data quality, yet many models perform suboptimally when confronted with noisy, imbalanced, or sparse outbreak data [81], [72]. Misinformation on social media, reflected in public vaccine sentiment [69], further complicates analytics-driven communication strategies. Finally, despite the rapid adoption of digital dashboards and telehealth systems, usability issues and practitioner burden remain unresolved, with dashboards often launched under constraints of missing data, limited stakeholder input, and insufficient public health agency capacity [57], [70]. These systemic and technical limitations underscore the urgent need for ethically aligned, explainable, and interoperable analytics frameworks that can operate across diverse clinical, infrastructural, and sociopolitical contexts.

5.8.4 Research Opportunities and Future Directions

Future directions in pandemic-oriented healthcare analytics call for advancing integration, adaptability, and interpretability across systems and methodologies. A significant opportunity lies in optimizing integrated data infrastructures for outbreak detection and regional responsiveness, particularly through scalable analytics platforms that accommodate real-time data from Internet of Things (IoT) devices, mobile applications, and electronic health records (EHRs) [39], [62]. These systems must function across institutional and national boundaries, necessitating interoperability standards to prevent failures in cross-context generalization. Future research should focus on developing hybrid, plug-and-play AI systems that combine adaptive learning with rule-based logic, enabling faster deployment during emergent scenarios while reducing false positives and enhancing transparency [53], [76]. Incorporating uncertainty quantification and synthetic datasets can further strengthen pandemic preparedness by simulating high-stress conditions [68].

Furthermore, integrating sociocultural factors and behavioral analytics into surveillance models will improve public engagement and system adherence, ensuring analytics solutions are both technically robust and socially grounded. There is also room for innovation in real-time dashboards that embed predictive algorithms and interactive interfaces for dynamic outbreak management [74]. Simultaneously, the proliferation of LLMs opens new research avenues in clinical documentation synthesis, mental health monitoring, and misinformation detection, supporting harmonized, explainable diagnostics and public health communication [54], [69]. Building interpretability directly into AI architectures—rather than relying solely on post-hoc explanations—will foster trust in high-stakes environments, such as diagnostics and clinical triage [14]. Addressing data heterogeneity, especially across underserved populations and noisy outbreak datasets, remains essential to ensure equitable model performance and reliability [72]. Additionally, the strategic use of federated frameworks should incorporate dynamic node selection and privacy-aware model updates, especially under infrastructure constraints and data protection mandates [53]. Lastly, interdisciplinary collaboration across public health, informatics, clinical science, and ethics will be critical to co-design scalable, real-time, and ethically governed analytics ecosystems capable of supporting both outbreak mitigation and long-term healthcare resilience [70], [57].

5.8.5 Recommendations to enhance practical value

To maximize the practical value of healthcare analytics in outbreak contexts, systems must be designed for real-time access, scalability, and stakeholder engagement while addressing interoperability, interpretability, and infrastructural diversity. A foundational step involves embedding real-time alerting and decision-making systems within clinical workflows, supported by interoperable cloud platforms and standardized data-sharing protocols to enhance responsiveness across institutions [39], [58], [65]. Harmonizing metadata standards and resolving inconsistencies—such as those seen in COVID-19 case and death datasets—is crucial for coherent analysis and cross-platform deployment [68]. Equally, dashboards and telehealth platforms should adhere to human-centered design principles tailored to the specific needs of various users, including policymakers, frontline clinicians, and the general public, with adaptable visualizations and localized interfaces [74], [57]. The wide-scale deployment of wearable-based predictive analytics and edge computing infrastructures must be supported through incentives or subsidies in low-resource environments [43], [62] . Simultaneously, standardized federated learning protocols should be developed to reduce energy and communication overhead while ensuring privacy-preserving collaboration, particularly in geographically fragmented systems [53]. Training programs for healthcare professionals and IT staff on interpreting AI-driven outputs, managing federated systems, and mitigating misinformation are critical to bridging the gap between data insight and clinical action [69], [70].

Furthermore, model transparency must be embedded at the design stage through the use of inherently interpretable algorithms coupled with real-world validation strategies that account for class imbalance, data heterogeneity, and latency constraints [14], [72], [81]. Collaborative governance frameworks should guide the ethical deployment of large language models (LLMs) and cloud-based systems, ensuring alignment with regulatory standards and minimizing bias and privacy risks [54]. Lastly, strategic alliances among academia, public health institutions, and industry can accelerate tool validation, enabling controlled pilots before full-scale deployment and enhancing public trust in outbreak-driven AI systems. Collectively, these actions ensure that analytics platforms evolve from experimental pilots to embedded, resilient systems that can support timely, equitable, and transparent pandemic responses.

6 Synthesis and Research Opportunities

6.1 Cross-Cutting Trends Across Themes

Across the eight thematic clusters, several cross-cutting trends become apparent. First, the field is converging toward real-time, multimodal, and patient-centered analytics, where electronic health records, imaging, physiological signals, and digital traces are combined to enable anticipatory and personalized care. Second, there is a growing emphasis on explainability, fairness, and trust, as evident in adaptive AI, demographic analytics, ethical frameworks, and mental health surveillance, all of which stress the need for transparency in high-stakes contexts. Third, the integration of public health and clinical perspectives illustrates an evolution from localized prediction to population-scale monitoring, as seen in pandemic analytics, demographic studies, and digital surveillance. Ultimately, these themes collectively underscore healthcare analytics as a socio-technical ecosystem, where methodological advances are inextricably linked to issues of equity, usability, and governance. This convergence underscores the transition from fragmented innovations to integrated infrastructures that align with broader biomedical priorities.

6.2 Conceptual and Methodological Gaps

Despite rapid advances, important conceptual and methodological gaps persist across healthcare analytics. A recurring issue is model generalizability, as predictive systems often underperform when applied to new populations, institutions, or demographic groups, limiting their translational value. Data heterogeneity and imbalance—spanning EHRs, imaging, wearable data, and social media—complicate integration and undermine fairness, particularly for underrepresented populations. Many studies rely on retrospective or curated datasets, which hinders reproducibility and real-world applicability. From a methodological perspective, interpretability remains an insufficiently addressed issue: while explainable AI techniques are increasingly adopted, they are inconsistently validated and rarely integrated into clinical workflows. Moreover, the field lacks standardized evaluation protocols that account for usability, scalability, and ethical dimensions, resulting in fragmented benchmarks. Finally, the conceptual framing of healthcare analytics is often siloed, with limited theoretical integration across clinical, public health, and socio-technical domains. Addressing these gaps is critical to building robust, equitable, and sustainable analytics infrastructures.

6.3 Underexplored Application Areas

While healthcare analytics has achieved notable progress in domains such as chronic disease management, diagnostic imaging, and outbreak response, several application areas remain underexplored. Low-resource and rural healthcare systems are often absent from empirical studies, despite their pressing need for scalable and affordable analytics solutions. Similarly, mental health and behavioral health analytics, though growing, lack robust longitudinal validation and integration with clinical care pathways. Applications that incorporate social determinants of health, including socioeconomic status, education, and environmental exposures, remain limited, thereby constraining efforts to address health equity. Additionally, the potential of analytics for preventive care and lifestyle interventions—such as nutrition, physical activity, and health literacy—remains underutilized compared to disease-focused applications. Ultimately, policy and regulatory domains have yet to fully benefit from analytics-driven insights, particularly in evaluating health system performance, resource allocation, and population-level interventions.

Additionally, Health Economics and Outcomes Research (HEOR), which integrates clinical and administrative data for cost-effectiveness and outcome evaluations, did not emerge as a distinct cluster in our analysis. This omission reflects the keyword scope of the search and the clustering process, which prioritized descriptors explicitly associated with healthcare analytics. Nonetheless, HEOR represents a valuable underexplored domain where analytics-driven approaches could strengthen decision-making in resource allocation, equity assessment, and health system performance. Expanding research into these areas could enhance both societal impact and translational value.

In the Latin American context, and particularly in Colombia, these underexplored areas become especially relevant. The absence of empirical studies in middle-income health systems limits the ability to validate models developed in Global North settings. In Colombia, databases such as SISPRO and RIPS represent valuable sources of information for designing analytics that address local challenges such as system fragmentation, inequities in access, and the growing burden of non-communicable chronic diseases, especially cardiovascular conditions. Likewise, the implementation of analytics in telemedicine programs and primary care services in rural areas offers a concrete opportunity to reduce regional gaps and optimize resources. These scenarios demonstrate that expanding research into Latin America would not only enable the validation of methodologies in diverse populations but also increase the social impact and relevance of healthcare analytics in contexts where the need for scalable and sustainable solutions is critical.

6.4 Opportunities for Integration and Theoretical Advancement

Current trends in healthcare analytics reveal significant opportunities to integrate fragmented approaches and advance theoretical foundations. A key priority is the fusion of multimodal data streams—spanning clinical records, genomics, behavioral data, and social media—into unified frameworks that enable holistic patient and population-level insights. Similarly, combining statistical modeling, machine learning, and domain knowledge can enhance interpretability and improve clinical trust, moving beyond purely data-driven pipelines. Theoretical progress is also needed to conceptualize healthcare analytics not only as a set of tools but as a socio-technical system, where human factors, governance, and ethical considerations are intrinsic to design. Integrative approaches, such as federated and collaborative learning architectures, open pathways for scalable, privacy-preserving systems that bridge institutional and national boundaries. By aligning methodological innovations with theories of precision medicine, equity, and health systems resilience, the field can transition toward more cohesive, theory-informed, and practically actionable frameworks.

6.5 Strategic Research Agenda for the Field

Building on the trends, gaps, and opportunities identified, a strategic research agenda for healthcare analytics should focus on five key priorities. First, equity and inclusivity must be central, ensuring that models generalize across diverse populations and settings, particularly in underrepresented and low-resource environments. Second, explainability and trustworthiness require stronger integration of interpretable mechanisms and validation frameworks into both clinical and public health workflows. Third, advancing real-world implementation research is crucial for evaluating scalability, usability, and impact, thereby moving beyond proof-of-concept studies toward system-wide adoption. Fourth, the development of interoperable infrastructures—including standardized ontologies, shared benchmarks, and federated architectures—will enable collaboration across institutions and nations while preserving privacy. Finally, interdisciplinary and cross-sector partnerships should guide the ethical governance, policy alignment, and clinical integration of analytics, ensuring that innovation remains socially responsive and clinically relevant. Pursuing these directions will position healthcare analytics as a foundational enabler of precision medicine, resilient health systems, and equitable public health transformation.

Finally, it is necessary to situate this agenda within the Latin American and Colombian context, where health systems face particular challenges of fragmentation, uneven coverage, and inequitable access. The application of healthcare analytics in the region has been concentrated in isolated projects of telemedicine, epidemiological surveillance (such as dengue, chikungunya, and COVID-19), and cardiovascular risk programs in some health insurance providers (EPS); however, there is still a lack of systematic studies evaluating the scalability and sustainability of these initiatives. Incorporating regional examples into the strategic agenda would strengthen the relevance of healthcare analytics in low-resource settings, improve interoperability among national information systems (such as SISPRO and RIPS in Colombia), and promote models that integrate social determinants of health in vulnerable populations. In this way, the field could move toward more inclusive and resilient analytics infrastructures that align with the goals of equity and sustainability in Latin America.

6.6 Practical and Policy Implications

To ensure that the proposed strategic agenda transcends the academic domain and offers actionable contributions, it is necessary to translate these priorities into concrete guidelines for key stakeholders. For hospitals and healthcare providers, analytics can be integrated into electronic health records and triage systems to anticipate high-risk cases (e.g., cardiovascular events), optimize resource allocation, and strengthen quality of care monitoring. For the research community, the establishment of shared benchmarks, open repositories, and collaborative evaluation protocols would enable reproducibility, regional cooperation, and capacity-building in data science for health. For public health agencies and policymakers, advanced analytics can guide epidemiological surveillance, equitable distribution of resources, and continuous monitoring of health inequalities, aligning with national information systems such as SISPRO and RIPS in Colombia.

By articulating these guidelines, healthcare analytics becomes not only a research field but also a practical tool that informs clinical decision-making, supports evidence-based policymaking, and strengthens health system resilience. This dual orientation—scientific and practical—ensures that innovation in healthcare analytics contributes directly to addressing the pressing needs of hospitals, researchers, and public health authorities in both Colombia and Latin America.

7 LIMITATIONS OF THE STUDY

This study has several limitations that should be acknowledged. First, although the dataset was extensive—comprising 2,281 Scopus-indexed publications—reliance on a single database may have excluded relevant works indexed in other repositories, particularly regional or non-English contributions, which could bias the thematic structure. Second, text-mining and clustering methods, while systematic, are sensitive to preprocessing choices such as descriptor cleaning, stopword removal, and normalization. Ambiguities in terminology and underrepresentation of emerging concepts may have influenced the resulting clusters. Third, the interpretation of thematic clusters, although grounded in systematic descriptors, inevitably involves a degree of subjectivity. The process of labeling, consolidating, and framing themes reflects the researcher's judgment, which may differ from alternative readings of the data. Fourth, the synthesis emphasizes conceptual and methodological patterns rather than systematically evaluating clinical outcomes, implementation studies, or policy-level impacts, which would require complementary approaches such as systematic or scoping reviews. Fifth, bibliometric and text-mining methods tend to amplify dominant and highly cited themes, potentially overlooking innovative but less visible contributions. Moreover, the omission of Health Economics and Outcomes Research (HEOR) as a distinct cluster reflects the scope of our keyword strategy and descriptor clustering, which emphasized terms explicitly associated with healthcare analytics; while conceptually related, HEOR remains underrepresented in this review. Finally, healthcare analytics is a rapidly evolving field. Breakthroughs in generative AI, multimodal learning, and real-world deployment are emerging at a pace that may outstrip the coverage of this review, particularly for publications appearing after december 2024.

8 CONCLUSIONS

This review offers a comprehensive synthesis of the healthcare analytics landscape, conducted through a data-driven thematic analysis of 2,281 scholarly publications. The eight identified clusters—ranging from predictive systems and demographic modeling to ethical surveillance and outbreak response—highlight the multifaceted and interdisciplinary nature of the field. Specifically, the eight thematic clusters can be summarized as follows: (1) Predictive Systems, focused on developing AI-driven tools for forecasting clinical outcomes; (2) Demographic Modeling, centered on population-level risk stratification and resource allocation; (3) Ethical Surveillance, addressing privacy, governance, and trust in digital health monitoring; (4) Outbreak Response, analyzing epidemic patterns and supporting preparedness; (5) Mental Health Monitoring, leveraging digital platforms for early detection and support; (6) Adaptive AI for Personalized Care, advancing individualized treatment pathways; (7) Equity and Access in Healthcare Analytics, ensuring fairness across diverse populations; and (8) Integrated Decision-Support Infrastructures, enabling real-time, multimodal, and explainable healthcare solutions. Importantly, these clusters do not operate in isolation but intersect along shared trajectories: adaptive AI, patient-centered care, and demographic analytics converge on personalization and equity, while ethical surveillance, mental health monitoring, and outbreak response collectively address societal-scale challenges of trust, resilience, and governance.

Together, they also illustrate the field’s evolution from retrospective, siloed applications toward integrated, real-time, and explainable infrastructures. This trajectory aligns healthcare analytics with broader biomedical priorities, including precision medicine, health equity, and public health preparedness.

These themes reflect the maturation of healthcare analytics from isolated technical innovations toward a cohesive methodological ecosystem, marked by increasing integration of multimodal data, AI-enabled inference, and stakeholder-aligned decision support. Our findings underscore how healthcare analytics is transitioning into a foundational layer of biomedical informatics, offering scalable and adaptive solutions for clinical decision-making, population health management, and precision care delivery.

Despite this progress, persistent challenges limit the complete translation of analytics into practice. Interoperability constraints, algorithmic opacity, data representation inequities, and insufficient regulatory frameworks continue to undermine the reliability and generalizability of data-driven tools. Ethical concerns related to patient privacy, transparency, and accountability further complicate deployment, particularly in high-stakes environments such as digital mental health and pandemic surveillance. Addressing these limitations requires sustained methodological innovation, including federated learning, explainable AI, and privacy-preserving infrastructures, alongside cross-sector collaboration to align technical development with clinical and public health priorities.

Future research should prioritize the co-design of analytics systems with domain experts, foster equity-aware modeling across diverse populations, and institutionalize transparent evaluation protocols to ensure accountability and transparency. Building inclusive, real-time, and context-aware analytic ecosystems will be essential for operationalizing the promise of healthcare analytics. Ultimately, by mapping the thematic, methodological, and translational dimensions of this expanding field, this review lays a foundation for the next generation of research and practice. It positions healthcare analytics not merely as a set of tools but as a strategic enabler of learning health systems and a critical driver of global biomedical innovation.

Although this study is based on an international body of literature, it is important to note that most of the reviewed applications and developments originate from Global North contexts. Consequently, there is a need to expand research into Latin American settings, particularly in Colombia, where health systems face specific challenges related to data heterogeneity, limited interoperability, and inequities in access to services. Incorporating local examples would not only increase the contextual relevance of healthcare analytics but also generate evidence to inform public policy and clinical management in low- and middle-resource environments. Future research should therefore validate predictive models in Colombian cohorts, explore the potential of telemedicine in rural regions, and develop analytics applied to locally burdensome diseases such as cardiovascular and infectious conditions (e.g., dengue, chikungunya). This would enable a more equitable transition from international innovation to solutions adapted to regional realities.

REFERENCES

[1] I. Ognjanovic, «Healthcare data analytics», *Stud. Health Technol. Inform.*, vol. 274, pp. 122-135, 2020, doi: 10.3233/SHTI200672.

[2] J. Wang, Z. Qin, J. Hsu, y B. Zhou, «A fusion of machine learning algorithms and traditional statistical forecasting models for analyzing American healthcare expenditure», *Healthc. Anal.*, vol. 5, 2024, doi: 10.1016/j.health.2024.100312.

[3] S. Naik, P. Kumar, S. Saha, S. D. Bairagya, D. Rawat, y S. K. Baliarsingh, «Predictive Healthcare Analytics: A Multidisease Approach using Logistic Regression», presentado en 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024, 2024. doi: 10.1109/ICCCNT61001.2024.10725194.

[4] C. K. Leung, T. H. Daniel Mai, N. D. Thong Tran, y C. Y. Zhang, «Predictive Analytics to Support Health Informatics on COVID-19 Data», presentado en BIBE 2021 - 21st IEEE International Conference on BioInformatics and BioEngineering, Proceedings, 2021. doi: 10.1109/BIBE52308.2021.9635556.

[5] S. Mishra, M. Alharbi, H. K. Tripathy, B. Sahoo, y A. Alkhayyat, *Healthcare Analytics and Advanced Computational Intelligence*. en Healthcare Analytics and Advanced Computational Intelligence. 2024, p. 223. doi: 10.1201/9781032624891.

[6] E. Heitkemper, S. Hulse, B. Bekemeier, M. Schultz, G. Whitman, y A. M. Turner, «The Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) Dashboard for Health Equity in Rural Public Health: Usability Evaluation», *JMIR Hum. Factors*, vol. 11, 2024, doi: 10.2196/51666.

[7] M. Khalifa, «Health Analytics Types, Functions and Levels: A Review of Literature», *Stud. Health Technol. Inform.*, vol. 251, pp. 137-140, 2018, doi: 10.3233/978-1-61499-880-8-137.

[8] D. Kyriazis *et al.*, «The CrowdHEALTH project and the hollistic health records: Collective wisdom driving public health policies», *Acta Inform. Medica*, vol. 27, n.º 5, pp. 369-373, 2019, doi: 10.5455/aim.2019.27.369-373.

[9] M. Trigka y E. Dritsas, «Improving Cardiovascular Disease Prediction With Deep Learning and Correlation-Aware SMOTE», *IEEE Access*, vol. 13, pp. 44590-44606, 2025, doi: 10.1109/ACCESS.2025.3549417.

[10] D. S. Saleh, M. S. Othman, W. O. Khudhur, y E. A. Aljabarti, «Effective Techniques for Handling Missing Values in Thyroid Disease Diagnosis: A Comparative Analysis», *Appl. Comput. Intell. Soft Comput.*, vol. 2025, n.º 1, 2025, doi: 10.1155/acis/2766701.

[11] S. Kukreti *et al.*, «Enhancing Disease Prediction through Random Forests in Healthcare Analytics», presentado en Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2024, 2024, pp. 1693-1699. doi: 10.1109/IC3I61595.2024.10828927.

[12] S. Li, M. Iqbal, A. K. Bashir, y X. Wang, «Privacy Preservation in Artificial Intelligence-Enabled Healthcare Analytics», *Hum.-Centric Comput. Inf. Sci.*, vol. 15, pp. 1-15, 2025, doi: 10.22967/HCIS.2025.15.016.

[13] L. Bohlen, J. Rosenberger, P. Zschech, y M. Kraus, «Leveraging interpretable machine learning in intensive care», *Ann. Oper. Res.*, vol. 347, n.º 2, pp. 1093-1132, 2025, doi: 10.1007/s10479-024-06226-8.

[14] A. Harikumar, S. Surendran, y S. Gargi, «Explainable AI in Deep Learning Based Classification of Fetal Ultrasound Image Planes», presentado en Procedia Computer Science, 2024, pp. 1023-1033. doi: 10.1016/j.procs.2024.03.291.

[15] M. A. Saeed y M. A. Saeed, «Real-Time Diabetes Detection Using Machine Learning and Apache Spark», presentado en 4th International Conference on Emerging Smart Technologies and Applications, eSmarTA 2024, 2024. doi: 10.1109/eSmarTA62850.2024.10638853.

[16] K. Zheng, H.-R. Chua, M. Herschel, H. V. Jagadish, B. C. Ooi, y J. W. L. Yip, «Exploiting Negative Samples: A Catalyst for Cohort Discovery in Healthcare Analytics», presentado en Proceedings of Machine Learning Research, 2024, pp. 61287-61320. [En línea]. Disponible en: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85198295164&partnerID=40&md5=7403836867210701732006609bee7659

[17] P. Whig, S. Kouser, A. B. Bhatia, R. R. Nadikattu, y P. Sharma, «Enabling trustworthy data sharing and collaborative insights in healthcare analytics», en *Secure Big-data Analytics for Emerging Healthcare in 5G and beyond: Concepts, paradigms, and solutions*, 2024, pp. 145-167. doi: 10.1049/PBHE063E_ch8.

[18] R. Elragal, A. Elragal, y A. Habibipour, «Healthcare Analytics: Conceptualizing a Research Agenda», presentado en Procedia Computer Science, 2024, pp. 1678-1686. doi: 10.1016/j.procs.2024.06.345.

[19] M. Aizpurua, T. Galluzzi, y S. Esteban, «Creating and implementing a professional development program to improve organizational capacity for health analytics in a public health setting»,

presentado en Studies in Health Technology and Informatics, 2020, pp. 1231-1232. doi: 10.3233/SHTI200377.

[20] J. Ayles, M. D. C. C. de Lima, y N. Gupta, «Public sector health analytics capacity before and after Covid-19: A case study of manager perspectives in New Brunswick, Canada», *Int. J. Popul. Data Sci.*, vol. 9, n.º 1, 2024, doi: 10.23889/ijpds.v9i1.2370.

[21] S. Das, S. P. Nayak, B. Sahoo, y S. C. Nayak, «Machine Learning in Healthcare Analytics: A State-of-the-Art Review», *Arch. Comput. Methods Eng.*, vol. 31, n.º 7, pp. 3923-3962, 2024, doi: 10.1007/s11831-024-10098-3.

[22] M. S. Islam, M. M. Hasan, X. Wang, H. D. Germack, y M. Noor-E-alam, «A systematic review on healthcare analytics: Application and theoretical perspective of data mining», *Healthc. Switz.*, vol. 6, n.º 2, 2018, doi: 10.3390/healthcare6020054.

[23] R. Elragal, A. Elragal, y A. Habibipour, «Healthcare analytics—A literature review and proposed research agenda», *Front. Big Data*, vol. 6, 2023, doi: 10.3389/fdata.2023.1277976.

[24] Q. Liu, M. Yu, B. Xiong, Z. Cai, P. Zhang, y C.-W. Tan, «Health analytics in business research: a literature review», *J. Manag. Anal.*, vol. 10, n.º 3, pp. 447-476, 2023, doi: 10.1080/23270012.2023.2209863.

[25] T. A. Pearson *et al.*, «Precision Health Analytics With Predictive Analytics and Implementation Research: JACC State-of-the-Art Review», *J. Am. Coll. Cardiol.*, vol. 76, n.º 3, pp. 306-320, 2020, doi: 10.1016/j.jacc.2020.05.043.

[26] S.-S. M. Ajibade *et al.*, «Evolution of machine learning applications in medical and healthcare analytics research: A bibliometric analysis», *Intell. Syst. Appl.*, vol. 24, 2024, doi: 10.1016/j.iswa.2024.200441.

[27] M. Khalifa, «Challenges of Health Analytics Utilization: A Review of Literature», presentado en Studies in Health Technology and Informatics, 2019, pp. 55-58. doi: 10.3233/SHTI190015.

[28] S. Gold *et al.*, «Clinical Concept Value Sets and Interoperability in Health Data Analytics», *AMIA Annu. Symp. Proc. AMIA Symp.*, vol. 2018, pp. 480-489, 2018.

[29] M. Khalifa y M. Househ, «Utilizing Health Analytics in Improving the Performance of Hospitals and Healthcare Services: Promises and Challenges», en *Lecture Notes in Bioengineering*, 2021, pp. 23-39. doi: 10.1007/978-3-030-67303-1_3.

[30] «Healthcare Analytics Market Size | Industry Report, 2033». Accedido: 12 de septiembre de 2025. [En línea]. Disponible en: https://www.grandviewresearch.com/industry-analysis/healthcare-analytics-market

[31] «Healthcare Analytics Market Growth, Drivers, and Opportunities», MarketsandMarkets. Accedido: 12 de septiembre de 2025. [En línea]. Disponible en: https://www.marketsandmarkets.com/Market-Reports/healthcare-data-analytics-market-905.html

[32] S. Agboola *et al.*, «Health Care Cost Analyses for Exploring Cost Savings Opportunities in Older Patients: Longitudinal Retrospective Study», *JMIR Aging*, vol. 1, n.º 2, p. e10254, ago. 2018, doi: 10.2196/10254.

[33] T. Hawkins, «Corewell Health Study Determines Keys to Reducing Hospital Readmissions», Spectrum Health Newsroom. Accedido: 12 de septiembre de 2025. [En línea]. Disponible en: https://newsroom.spectrumhealth.org/corewell-health-study-determines-keys-to-reducing-hospital-readmissions/

[34] M. Aria y C. Cuccurullo, «bibliometrix: An R-tool for comprehensive science mapping analysis», *J. Informetr.*, vol. 11, n.º 4, pp. 959-975, 2017.

[35] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, y W. M. Lim, «How to conduct a bibliometric analysis: An overview and guidelines», *J. Bus. Res.*, vol. 133, pp. 285-296, 2021, doi: https://doi.org/10.1016/j.jbusres.2021.04.070.

[36] M. J. Page *et al.*, «The PRISMA 2020 statement: an updated guideline for reporting systematic reviews», *BMJ*, vol. 372, p. n71, mar. 2021, doi: 10.1136/bmj.n71.

[37] S. Staras *et al.*, «Using a clinical workflow analysis to enhance eHealth implementation planning: tutorial and case study», *JMIR MHealth UHealth*, vol. 9, n.º 3, p. e18534, 2021.

[38] K. S. Taylor, K. R. Mahtani, y J. K. Aronson, «Summarising good practice guidelines for data extraction for systematic reviews and meta-analysis», *BMJ Evid.-Based Med.*, vol. 26, n.º 3, pp. 88-90, 2021.

[39] A. Sheikhtaheri, S. M. Tabatabaee Jabali, E. Bitaraf, A. TehraniYazdi, y A. Kabir, «A near real-time electronic health record-based COVID-19 surveillance system: An experience from a developing country», *Health Inf. Manag. J.*, vol. 53, n.º 2, pp. 145-154, 2024, doi: 10.1177/18333583221104213.

[40] A. Al Nuairi, M. C. E. Simsekler, A. Qazi, y A. Sleptchenko, «A data-driven Bayesian belief network model for exploring patient experience drivers in healthcare sector», *Ann. Oper. Res.*, vol. 342, n.º 3, pp. 1797-1817, 2024, doi: 10.1007/s10479-023-05437-9.

[41] L. Hammelrath, K. Hilbert, M. Heinrich, P. Zagorscak, y C. Knaevelsrud, «Select or adjust? How information from early treatment stages boosts the prediction of non-response in internet-based

depression treatment», *Psychol. Med.*, vol. 54, n.º 8, pp. 1641-1650, 2024, doi: 10.1017/S0033291723003537.

[42] A. Manocha, S. K. Sood, y M. Bhatia, «Edge intelligence-assisted smart healthcare solution for health pandemic: a federated environment approach», *Clust. Comput.*, vol. 27, n.º 5, pp. 5611-5630, 2024, doi: 10.1007/s10586-023-04245-x.

[43] J. A. J. Singh, J. Gnanasoundharam, M. Birunda, G. Sudha, S. P. Maniraj, y C. Srinivasan, «Wearable Sepsis Early Warning Using Cloud Computing and Logistic Regression Predictive Analytics», presentado en 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2024, 2024. doi: 10.1109/ICRITO61523.2024.10522250.

[44] A. Sambarey *et al.*, «Integrative analysis of multimodal patient data identifies personalized predictors of tuberculosis treatment prognosis», *iScience*, vol. 27, n.º 2, 2024, doi: 10.1016/j.isci.2024.109025.

[45] J.-H. Jeong, I.-G. Lee, S.-K. Kim, T.-E. Kam, S.-W. Lee, y E. Lee, «DeepHealthNet: Adolescent Obesity Prediction System Based on a Deep Learning Framework», *IEEE J. Biomed. Health Inform.*, vol. 28, n.º 4, pp. 2282-2293, 2024, doi: 10.1109/JBHI.2024.3356580.

[46] F. Papaiz, M. E. T. Dourado, R. A. de Medeiros Valentim, R. Pinto, A. H. F. de Morais, y J. P. Arrais, «Ensemble-imbalance-based classification for amyotrophic lateral sclerosis prognostic prediction: identifying short-survival patients at diagnosis», *BMC Med. Inform. Decis. Mak.*, vol. 24, n.º 1, 2024, doi: 10.1186/s12911-024-02484-5.

[47] W. Zhang, J. Xie, Z. Zhang, y X. Liu, «Depression Detection Using Digital Traces on Social Media: A Knowledge-aware Deep Learning Approach», *J. Manag. Inf. Syst.*, vol. 41, n.º 2, pp. 546-580, 2024, doi: 10.1080/07421222.2024.2340822.

[48] G. Owusu, X. Wang, y J. Sun, «Longitudinal healthcare analytics for early detection and progression of neurological diseases: A clinical decision support system», presentado en Proceedings of the Annual Hawaii International Conference on System Sciences, 2024, pp. 3297-3306. [En línea]. Disponible en: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85199771167&partnerID=40&md5=d31126afc2bee3d2a28e5a4fe87dc405

[49] B. P. P. Kumar y H. M. Manoj, «Comparative Assessment of Machine Learning Models for Predicting Glucose Intolerance Risk», *SN Comput. Sci.*, vol. 5, n.º 7, 2024, doi: 10.1007/s42979-024-03259-5.

[50] A. S. Zamani, A. H. A. Hashim, A. S. A. Shatat, M. M. Akhtar, M. Rizwanullah, y S. S. I. Mohamed, «Implementation of machine learning techniques with big data and IoT to create effective prediction models for health informatics», *Biomed. Signal Process. Control*, vol. 94, 2024, doi: 10.1016/j.bspc.2024.106247.

[51] N. Al-Shanableh *et al.*, «Advanced ensemble machine learning techniques for optimizing diabetes mellitus prognostication: A detailed examination of hospital data», *Data Metadata*, vol. 3, 2024, doi: 10.56294/dm2024.363.

[52] D. K. Gurmessa y W. Jimma, «Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review», *BMJ Health Care Inform.*, vol. 31, n.º 1, 2024, doi: 10.1136/bmjhci-2023-100954.

[53] S. S. Tripathy *et al.*, «FedHealthFog: A federated learning-enabled approach towards healthcare analytics over fog computing platform», *Heliyon*, vol. 10, n.º 5, 2024, doi: 10.1016/j.heliyon.2024.e26416.

[54] H. Yu *et al.*, «Large Language Models in Biomedical and Health Informatics: A Review with Bibliometric Analysis», *J. Healthc. Inform. Res.*, vol. 8, n.º 4, pp. 658-711, 2024, doi: 10.1007/s41666-024-00171-8.

[55] F. Rabbi, D. Banik, N. U. I. Hossain, y A. Sokolov, «Using process mining algorithms for process improvement in healthcare», *Healthc. Anal.*, vol. 5, 2024, doi: 10.1016/j.health.2024.100305.

[56] Y. Wu, M. Wu, C. Wang, J. Lin, J. Liu, y S. Liu, «Evaluating the Prevalence of Burnout Among Health Care Professionals Related to Electronic Health Record Use: Systematic Review and Meta-Analysis», *JMIR Med. Inform.*, vol. 12, 2024, doi: 10.2196/54811.

[57] B. Ansari y E. G. Martin, «Integrating human-centered design in public health data dashboards: Lessons from the development of a data dashboard of sexually transmitted infections in New York State», *J. Am. Med. Inform. Assoc.*, vol. 31, n.º 2, pp. 298-305, 2024, doi: 10.1093/jamia/ocad102.

[58] S. Chintala, «Iot and Ai Synergy: Remote Patient Monitoring for Improved Healthcare», presentado en 4th International Conference on Innovative Practices in Technology and Management 2024, ICIPTM 2024, 2024. doi: 10.1109/ICIPTM59628.2024.10563530.

[59] H. Luo *et al.*, «Drug-drug interactions prediction based on deep learning and knowledge graph: A review», *iScience*, vol. 27, n.º 3, 2024, doi: 10.1016/j.isci.2024.109148.

[60] C. K. Leung, «Biomedical Informatics: State of the Art, Challenges, and Opportunities», *BioMedInformatics*, vol. 4, n.º 1, pp. 89-97, 2024, doi: 10.3390/biomedinformatics4010006.

[61] U. Islam, G. Mehmood, A. A. Al-Atawi, F. Khan, H. S. Alwageed, y L. Cascone, «NeuroHealth guardian: A novel hybrid approach for precision brain stroke prediction and healthcare analytics», *J. Neurosci. Methods*, vol. 409, 2024, doi: 10.1016/j.jneumeth.2024.110210.

[62] V. Dankan Gowda, S. M. Chaithra, S. S. Gujar, S. F. Shaikh, B. S. Ingole, y N. S. Reddy, «Scalable AI Solutions for IoT-based Healthcare Systems using Cloud Platforms», presentado en 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings, 2024, pp. 156-162. doi: 10.1109/I-SMAC61858.2024.10714810.

[63] B. S. Glicksberg *et al.*, «Evaluating the accuracy of a state-of-the-art large language model for prediction of admissions from the emergency room», *J. Am. Med. Inform. Assoc.*, vol. 31, n.º 9, pp. 1921-1928, 2024, doi: 10.1093/jamia/ocae103.

[64] W. Ni, H. Ao, H. Tian, Y. C. Eldar, y D. Niyato, «FedSL: Federated Split Learning for Collaborative Healthcare Analytics on Resource-Constrained Wearable IoMT Devices», *IEEE Internet Things J.*, vol. 11, n.º 10, pp. 18934-18935, 2024, doi: 10.1109/JIOT.2024.3370985.

[65] J. G. Jeslin, K. Vijayalakshmi, C. C. Vignesh, G. Suresh, G. V. Kosuri, y S. Murugan, «Predicting Patient Disease Progression with Cloud-based Decision Trees and IoT Data Integration», presentado en 2nd International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2024 - Proceedings, 2024, pp. 1040-1045. doi: 10.1109/ICSSAS64001.2024.10760687.

[66] X. Meng *et al.*, «The application of large language models in medicine: A scoping review», *iScience*, vol. 27, n.º 5, 2024, doi: 10.1016/j.isci.2024.109713.

[67] M. Hosseini, M. Hosseini, y R. Javidan, «Leveraging Large Language Models for Clinical Abbreviation Disambiguation», *J. Med. Syst.*, vol. 48, n.º 1, 2024, doi: 10.1007/s10916-024-02049-z.

[68] D. Halpern, Q. Lin, R. Wang, S. Yang, S. Goldstein, y M. Kolak, «Dimensions of uncertainty: a spatiotemporal review of five COVID-19 datasets», *Cartogr. Geogr. Inf. Sci.*, vol. 51, n.º 2, pp. 200-221, 2024, doi: 10.1080/15230406.2021.1975311.

[69] A. A. Mir y R. Sevukan, «Sentiment analysis of Indian Tweets about Covid-19 vaccines», *J. Inf. Sci.*, vol. 50, n.º 5, pp. 1308-1320, 2024, doi: 10.1177/01655515221118049.

[70] S. C. Christopoulou, «Machine Learning Models and Technologies for Evidence-Based Telehealth and Smart Care: A Review», *BioMedInformatics*, vol. 4, n.º 1, pp. 754-779, 2024, doi: 10.3390/biomedinformatics4010042.

[71] M. Schuemie *et al.*, «Health-Analytics Data to Evidence Suite (HADES): Open-Source Software for Observational Research», presentado en Studies in Health Technology and Informatics, 2024, pp. 966-970. doi: 10.3233/SHTI231108.

[72] E. Ali, M. Angelova, y C. Karmakar, «Epileptic seizure detection using CHB-MIT dataset: The overlooked perspectives», *R. Soc. Open Sci.*, vol. 11, n.º 6, 2024, doi: 10.1098/rsos.230601.

[73] L. R. Namamula y D. Chaytor, «Effective ensemble learning approach for large-scale medical data analytics», *Int. J. Syst. Assur. Eng. Manag.*, vol. 15, n.º 1, pp. 13-20, 2024, doi: 10.1007/s13198-021-01552-7.

[74] R. Rabiei, P. Bastani, H. Ahmadi, S. Dehghan, y S. Almasi, «Developing public health surveillance dashboards: a scoping review on the design principles», *BMC Public Health*, vol. 24, n.º 1, 2024, doi: 10.1186/s12889-024-17841-2.

[75] A. H. Van Der Vegt *et al.*, «Systematic review and longitudinal analysis of implementing Artificial Intelligence to predict clinical deterioration in adult hospitals: What is known and what remains uncertain», *J. Am. Med. Inform. Assoc.*, vol. 31, n.º 2, pp. 509-524, 2024, doi: 10.1093/jamia/ocad220.

[76] D. Mendhe, A. Dogra, P. S. Nair, S. Punitha, K. S. Preetha, y S. B. G. T. Babu, «AI-Enabled Data-Driven Approaches for Personalized Medicine and Healthcare Analytics», presentado en Proceedings of 9th International Conference on Science, Technology, Engineering and Mathematics: The Role of Emerging Technologies in Digital Transformation, ICONSTEM 2024, 2024. doi: 10.1109/ICONSTEM60960.2024.10568722.

[77] S. Rasheed, G. K. Kumar, D. M. Rani, M. V. V. P. Kantipudi, y M. Anila, «Heart Disease Prediction Using GridSearchCV and Random Forest», *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 10, 2024, doi: 10.4108/eetpht.10.5523.

[78] M. Khalifa y M. Albadawy, «AI in diagnostic imaging: Revolutionising accuracy and efficiency», *Comput. Methods Programs Biomed. Update*, vol. 5, 2024, doi: 10.1016/j.cmpbup.2024.100146.

[79] P. Bhambri, S. Rani, y M. Fahim, *Computational Intelligence and Blockchain in Biomedical and Health Informatics*. en Computational Intelligence and Blockchain in Biomedical and Health Informatics. 2024, p. 334. doi: 10.1201/9781003459347.

[80] S. Elkefi, «Access and usage of mobile health (mHealth) for communication, health monitoring, and decision-making among patients with multiple chronic diseases (comorbidities)», *IISE Trans. Healthc. Syst. Eng.*, vol. 14, n.º 3, pp. 179-192, 2024, doi: 10.1080/24725579.2023.2267085.

[81] G. Prabhakar, V. R. Chintala, T. Reddy, y T. Ruchitha, «User-cloud-based ensemble framework for type-2 diabetes prediction with diet plan suggestion», *E-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 7, 2024, doi: 10.1016/j.prime.2024.100423.

[82] M. M. Islam, S. Nooruddin, F. Karray, y G. Muhammad, «Enhanced multimodal emotion recognition in healthcare analytics: A deep learning based model-level fusion approach», *Biomed. Signal Process. Control*, vol. 94, 2024, doi: 10.1016/j.bspc.2024.106241.

[83] C. Han *et al.*, «Evaluation of GPT-4 for 10-year cardiovascular risk prediction: Insights from the UK Biobank and KoGES data», *iScience*, vol. 27, n.º 2, 2024, doi: 10.1016/j.isci.2024.109022.

[84] C.-H. Yang *et al.*, «Global, regional and national burdens of depression in adolescents and young adults aged 10–24 years, from 1990 to 2019: findings from the 2019 Global Burden of Disease study», *Br. J. Psychiatry*, vol. 225, n.º 2, pp. 311-320, 2024, doi: 10.1192/bjp.2024.69.

[85] M. M. Eltoukhy, M. Elhoseny, K. M. Hosny, y A. K. Singh, «Computer aided detection of mammographic mass using exact Gaussian–Hermite moments», *J. Ambient Intell. Humaniz. Comput.*, vol. 15, n.º 1, pp. 1139-1147, 2024, doi: 10.1007/s12652-018-0905-1.