Eureka Open Access Journals

Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)

ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

THEORETICAL FOUNDATIONS AND INTERNATIONAL EXPERIENCE OF USING BIG DATA TECHNOLOGIES IN THE HEALTH INSURANCE SYSTEM

Inakov Sh. A.

Majidov R. A.

School of Public Health, Tashkent State Medical University

Annotation

This article examines the theoretical foundations and international experience of applying Big Data technologies in the health insurance system. The effective use of Big Data in healthcare depends not only on the quality of analytical models, but also on the infrastructure for collecting, storing, processing, and integrating large-scale heterogeneous datasets. Modern healthcare generates massive volumes of data from electronic health records (EHR/EMR), laboratory diagnostics, medical imaging, insurance claims, pharmaceutical databases, and continuous streams from Internet of Things (IoT) devices. The rapid growth of global digital data highlights the increasing importance of advanced data-driven approaches for improving decision-making in health insurance. The study reviews key components of Big Data infrastructure, including cloud computing, distributed computing clusters (Hadoop, Apache Spark), and machine learning methods, and discusses their role in risk assessment, cost prediction, fraud detection, and personalized insurance services. Based on international practices, the article emphasizes that Big Data integration into health insurance can enhance efficiency, transparency, and quality of medical services through evidence-based management and proactive risk evaluation.

ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

Keywords: Big Data; health insurance; healthcare analytics; electronic health records (EHR/EMR); Apache Spark; Hadoop; cloud computing; machine learning; Internet of Things (IoT); risk assessment; predictive modeling; insurance claims data; fraud detection.

Introduction

In contemporary healthcare, the successful deployment of Big Data solutions is determined not only by the sophistication of analytical models but also by the underlying ecosystem for data acquisition, storage, interoperability, and integration. Dash et al. (2019) highlight that healthcare has become one of the most data-intensive and heterogeneous domains, bringing together electronic health records (EHR/EMR), laboratory outputs, medical imaging, insurance claims, pharmaceutical databases, and continuous streams produced by Internet of Things (IoT) devices. The rapid expansion of the global "digital universe"-from 130 EB in 2005 to 16 zettabytes in 2017 and a projected 40 zettabytes by 2020-further illustrates the accelerating growth of healthcare-related information. From an infrastructure perspective, Big Data environments are typically supported by cloud computing, distributed processing frameworks (e.g., Hadoop and Apache Spark), and machine learning capabilities. Evidence suggests that Apache Spark can outperform Hadoop MapReduce by an order of magnitude for selected analytical workloads, which is particularly relevant for real-time surveillance, monitoring, and forecasting applications. Moreover, continuous streams generated through IoT and mobile health (mHealth) technologies can facilitate early identification of diabetes progression from subclinical to clinical stages, strengthening proactive risk management strategies [6].

Beyond healthcare-specific applications, empirical studies also demonstrate the capacity of Big Data architectures to uncover latent patterns in large and complex datasets. Učkuronytė and Maknickienė (2025), for instance, analysed a dataset of 10,000 financial transactions and, due to technical constraints, retained 4,000

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

records for modelling. Using the UMAP algorithm, they identified 38 clusters, including three dominant clusters that effectively separated high-risk transactions. Within these clusters, transaction values ranged between USD 1-4 million, and more than 80% of the cases were associated with illicit sources. In addition, the presence of up to eight shell companies within a single transaction cluster highlighted the structural complexity of high-risk financial activity [29].

The scale and economic implications of Big Data are particularly evident in the United States healthcare system. Raghupathi and Raghupathi (2014) reported that U.S. healthcare digital data volumes reached 150 exabytes in 2011, with projections indicating a rapid transition toward the zettabyte scale if growth trends persist. Large integrated providers such as Kaiser Permanente reportedly manage approximately 26.5-44 petabytes of EHR and imaging data, which are actively used to support clinical decision-making processes. Economic estimates further suggest that advanced analytics could yield annual savings exceeding USD 300 billion, with around 8% attributed to reductions in national healthcare expenditures. Specifically, inefficiency reductions were projected at approximately USD 165 billion in clinical operations and USD 108 billion in research and development activities [23].

In parallel, chronic disease management-particularly for diabetes-has increasingly shifted toward data-driven approaches, as conventional clinical follow-up alone cannot fully capture the multifaceted epidemiological and economic burden of the condition. Global Burden of Disease (GBD)-based modelling indicates that despite improvements in the diabetes care cascade (detection, treatment, and glycaemic control) between 2000 and 2023, major gaps persist: in 2023, nearly half of individuals living with diabetes worldwide remained undiagnosed, while only about one in five treated patients achieved optimal glycaemic control [25]. These findings reinforce the importance of transitioning from treatment-focused models toward strategies that prioritize early detection, continuous monitoring, and individualized risk assessment.

Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

Digital health tools, including mHealth applications, play a central role in this transition by enabling real-time capture of clinically meaningful indicators such as glucose dynamics, physical activity, dietary patterns, and medication adherence. Such data streams can enhance both clinical decision-making and insurance-related analytics. Nevertheless, evidence from systematic evaluations indicates that average user retention for mHealth applications remains modest (approximately 56%), suggesting that real-world effectiveness depends strongly on personalized design, passive/automated data capture, reminder systems, gamification elements, and integration with healthcare professionals and service pathways [13].

International evidence from health insurance systems consistently underscores the value of integrated, Big Data-enabled approaches for improving chronic disease control and system performance. In Indonesia, the implementation of a national universal health insurance scheme facilitated the consolidation of healthcare utilization data for millions of beneficiaries into a unified digital platform. This infrastructure has supported expenditure monitoring-particularly for diabetes and other non-communicable diseases (NCDs)-and strengthened the design of preventive strategies [1]. Similarly, experience from Saudi Arabia suggests that national NCD policies may remain suboptimal in the absence of robust Big Data-driven epidemiological surveillance and structured mechanisms for program evaluation. These findings reinforce the need to connect disease-specific datasets with health information systems and insurance registries; in particular, linking comprehensive surveillance architectures with electronic health records (EHRs) is critical for prevention, early case detection, and cost containment [10]. China's health insurance reforms provide further evidence that strategic purchasing, redesigned provider payment arrangements, and information technology-supported governance can enable large-scale coverage expansion. The resulting universal insurance model, covering more than 95% of the population, has enhanced financial protection and improved equitable access

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

to services for diabetes and other chronic conditions. Importantly, Big Data-supported primary healthcare (PHC)-based integrated service delivery has enabled more precise identification of high-risk groups and facilitated targeted resource allocation [30].

Taiwan offers an additional example of the analytical potential of integrated insurance data systems. By linking insurance registries with mortality databases and population surveys, researchers have developed high-performance cardiovascular risk scores, demonstrating that real-world clinical information can be leveraged to predict individual risk. This approach is particularly relevant for the early identification of diabetes-related cardiovascular complications and for directing preventive interventions toward vulnerable subgroups [5]. Evidence from Nepal also confirms that social health insurance can substantially influence healthcare utilization patterns: older adults with chronic diseases who were enrolled in insurance schemes reportedly used healthcare services almost four times more frequently. This highlights the operational interdependence between insurance coverage, data infrastructure, and clinical need, supporting the view that Big Data-enabled insurance platforms can serve as strategic tools for improving equity and strengthening financial protection in chronic disease management [21].

Beyond utilization effects, a growing body of research indicates that both the breadth of insurance coverage and the quality of benefit design are strongly associated with health outcomes and financial resilience among populations living with chronic conditions. Experimental and quasi-experimental studies over the last decade show that insurance expansion improves access to primary care, increases uptake of preventive services and diagnostic testing, and contributes to earlier detection of chronic diseases, including diabetes. In the United States, expansions of Medicaid and other public insurance programs have been linked to higher rates of diabetes diagnosis and nearly a twofold increase in medication use, suggesting that insurance can move diabetes from an under-detected "hidden

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

stage" into structured clinical management [12, 24]. At the same time, reducing financial exposure remains a core function of insurance in chronic disease contexts. Studies from China suggest that public health insurance meaningfully reduces outpatient-related financial pressure; however, private insurance schemes may not always fully remove barriers associated with inpatient treatment costs. These findings emphasize that insurance design, reimbursement rules, and depth of coverage are decisive determinants of the extent to which chronic disease-related financial burdens can be mitigated [20].

The integration of Big Data and artificial intelligence (AI) within health insurance systems has the potential to reshape not only clinical decision support but also the financial and administrative efficiency of healthcare delivery. Evidence from the U.S. healthcare system indicates that administrative spending reaches hundreds of billions of dollars annually, with substantial resources devoted to billing operations, insurance claims processing, documentation, and related accounting procedures rather than direct patient care. AI-driven platforms that combine EHRs, claims datasets, and clinical information within a unified digital ecosystem can reduce costs by automating billing workflows, streamlining prior-authorization procedures, and supporting risk-adjusted payment and compensation models. Estimates suggest that broad adoption of such systems could generate administrative savings of up to USD 168 billion per year [16]. In parallel, real-world Big Data infrastructures such as PIONEER in Europe and North America have operationalized large-scale analytics by harmonizing EHRs, registries, and insurance claims through the OMOP Common Data Model, enabling comparative assessment of comorbidities, medication safety, and complications across millions of patients. This provides a robust evidence base for implementing precision-oriented principles such as "the right patient - the right treatment - at the right time" in routine practice [16, 26]. Conceptual population health management models developed in the United States further indicate that integrating Big Data from EHRs, billing systems, pharmaceutical

Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

databases, and mHealth sources enables the evaluation of both individual clinical risk and the broader epidemiological and financial risk profile of an insurance portfolio. Collectively, these frameworks provide a theoretical basis for developing accurate, dynamic, and prediction-oriented insurance risk models for chronic diseases such as diabetes [8, 22].

Within this context, health insurance claims (HIC) data constitute a foundational element of Big Data ecosystems because they support near-complete population coverage, longitudinal follow-up, and joint assessment of clinical outcomes and economic consequences. A large retrospective analysis conducted in the Netherlands, involving claims data from more than 574,000 patients, showed that both physical and mental comorbidities significantly influenced healthcare utilization, demonstrating the feasibility of identifying individualized risk patterns using Big Data methodologies [7]. Evidence from Germany and other European settings similarly indicates that standardized coding systems-including ICD, ATC, and CPT-enable insurance datasets to be converted into high-resolution digital models for evaluating disease epidemiology, medication adherence, clinical outcomes, and cost-effectiveness. Such models are essential for evidence-based risk stratification and for strengthening prevention strategies in chronic conditions such as cardiovascular disease and diabetes [9]. Moreover, Big Data analyses aligned with real-world data (RWD) and real-world evidence (RWE) frameworks allow evaluation of medication effectiveness, adverse drug events, and healthcare spending in routine clinical practice through the linkage of EHRs, billing systems, and disease registries, thereby capturing real-life patterns that may be underrepresented in controlled clinical trials [10]. Finally, the increasing application of machine learning methods in population health analytics enhances predictive performance by integrating multifactorial, non-linear, and interdependent risk determinants. This strengthens the methodological basis for early identification of high-risk groups and supports more efficient, targeted resource allocation within health insurance systems [2].

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

International literature increasingly indicates that the convergence of Big Data streams from health insurance systems, clinical information infrastructures, and population-based epidemiological registries with artificial intelligence (AI) and machine learning has become a major driver of healthcare transformation at clinical, economic, and strategic levels. In practice, healthcare data are produced not only within hospitals and physician practices, but also across insurance institutions, pharmaceutical supply chains, medical device ecosystems, and the broader life sciences sector. The integration of these multi-origin datasets enables descriptive, predictive, and prescriptive analytics, thereby supporting optimized decision-making both at the level of individual patients and across entire insurance portfolios [11]. Evidence from Iran illustrates this potential: analysis of national electronic prescription data comprising more than 84 million records enabled evaluation of diabetes medication policies, identification of regional disparities, detection of inappropriate prescribing patterns, and formulation of adaptive policies informed by Big Data evidence [16]. At the same time, research has clearly established that embedded ethnic, gender, and social inequalities within datasets may generate algorithmic bias in AI-driven health insurance models. Consequently, the development of equitable and generalizable risk models requires open and representative data sources, inclusive standardization practices, and transparent algorithmic design principles [12]. Furthermore, the linkage of population health Big Data with clinical and insurance datasets allows simultaneous assessment of disease patterns, social determinants of health, medication adherence, and financial risk within a single analytical framework. This provides a theoretical basis for designing high-accuracy, policy-relevant, and clinically actionable risk assessment platforms for chronic diseases such as diabetes [3].

However, effective implementation of Big Data in healthcare and insurance contexts depends not only on technological infrastructure, but also on legal, institutional, and methodological preparedness. Evidence from Bangladesh

ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

suggests that although electronic health data systems, clinical documentation, and digital health services are expanding rapidly, the absence of clearly defined legal frameworks governing data protection, secondary data use, and regulation of AI-based medical applications limits the full realization of Big Data benefits [6]. In contrast, Taiwan's mandatory national health insurance program, established in 1995, has consolidated nationwide medical claims into the National Health Insurance Research Database (NHIRD), enabling real-world Big Data-driven clinical, pharmacoeconomic, and policy decision-making. The NHIRD has supported hundreds of high-impact studies addressing medication effectiveness, cancer risks, drug-related expenditures, and health policy evaluation [8]. From a public health standpoint, the integration of Big Data with decision-analytic models has become central to risk assessment, disease forecasting, and resource optimization, with particular relevance for prevention-focused strategies in diabetes and other chronic diseases [14]. Nevertheless, such implementations must address critical challenges including privacy risks, algorithmic errors, and the potential amplification of social inequities. These risks require balancing through robust data governance structures and appropriate regulatory oversight. Therefore, the adoption of Big Data in health insurance should be conceptualized not merely as a technological upgrade, but as a strategic framework integrating clinical, economic, and legal dimensions to support the development of accurate, reliable, and sustainable diabetes risk assessment platforms.

Beyond efficiency gains, Big Data analytics is increasingly recognized as a strategic instrument for detecting health inequalities and structural weaknesses within healthcare systems. Population-level predictive models frequently reveal pronounced differences across geographic areas, social groups, and levels of healthcare access, indicating that without small-area analytics, substantial portions of the true disease burden may remain obscured. This is especially important for identifying high-risk and underserved subpopulations affected by chronic diseases such as diabetes [5]. In addition, advanced analytical platforms

Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

that combine multimedia Big Data with EHRs, IoT-generated streams, and insurance datasets enable real-time clinical decision support by facilitating early detection, improved classification, and individualized forecasting [13]. Predictive modelling approaches developed within Hadoop/MapReduce environments for diabetes have processed millions of records, including glucose measures, blood pressure indicators, prescription histories, and insurance claims, demonstrating the feasibility of identifying disease subtypes, forecasting complications, and informing optimal treatment pathways. This provides further confirmation of the applied value of Big Data-based risk stratification platforms [4]. The broader theoretical foundation of Big Data analytics in healthcare relies on linking clinical, epidemiological, and business-oriented datasets into a unified analytical pipeline, supporting personalized medicine and value-based insurance mechanisms through descriptive, predictive, and prescriptive modelling [1]. Finally, evidence from China’s commercial health insurance sector suggests that Big Data can be used to model disease risk, treatment pathways, and healthcare costs in ways that reduce information asymmetry, mitigate adverse selection and moral hazard, improve the fairness of premium-setting, and enhance operational efficiency [15].

Conclusion

Overall, the reviewed evidence demonstrates that Big Data technologies are becoming a core enabler of more efficient, predictive, and prevention-oriented health insurance systems, particularly in the management of chronic diseases such as diabetes. The integration of heterogeneous data sources-including EHR/EMR, laboratory and imaging outputs, IoT/mHealth streams, and health insurance claims-supports advanced descriptive, predictive, and prescriptive analytics, enabling earlier risk identification, improved monitoring, and more targeted allocation of resources. International experience further indicates that universal insurance platforms and linked registries can strengthen cost control, equity, and

Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)
ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

financial protection, while real-world data frameworks expand the evidence base beyond traditional clinical trials. At the same time, sustainable implementation requires not only scalable infrastructure (cloud and distributed computing) but also strong data governance, interoperability standards, and regulatory mechanisms to mitigate privacy risks and algorithmic bias. Therefore, developing transparent and clinically actionable Big Data-driven risk assessment platforms represents a strategic priority for improving both the clinical outcomes and economic sustainability of health insurance systems.

References

1. Batko K., Ślęzak A. The use of Big Data Analytics in healthcare. // Journal of big data. 2022. № 1 (9). C. 3.
2. Bowe A. K. [и др.]. Big data, machine learning, and population health: predicting cognitive outcomes in childhood // Pediatric Research. 2023. № 2 (93). C. 300-307.
3. Bu D. D. [и др.]. Achieving Value in Population Health Big Data // Journal of General Internal Medicine. 2020. № 11 (35). C. 3342-3345.
4. Eswari T. [и др.]. Predictive methodology for diabetic data analysis in big data // Elsevier.
5. Ghadirian F. Big Data-Derived Population Studies and Mental Health Disparities // Journal of Psychosocial Nursing and Mental Health Services. 2025. № 3 (63). C. 3.
6. Hassan S. [и др.]. Big data and predictive analytics in healthcare in Bangladesh: regulatory challenges. // Heliyon. 2021. № 6 (7). C. e07179.
7. Howard S. W. [и др.]. A retrospective big data study using healthcare insurance claims to investigate the role of comorbidities in receiving low vision services // frontiersin.org. 2024. (4). C. 1264838.

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Health Sciences & Medical Innovation (EJHSMI)

ISSN 2760-4942 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/5

8. Janjua H. M. [и др.]. A of analytics and B of big data in healthcare research: Telling the tale of health outcomes research from the eyes of data // American Journal of Surgery. 2024. (230). С. 105-107.
9. Krefting J. [и др.]. Use of big data from health insurance for assessment of cardiovascular outcomes. // Frontiers in artificial intelligence. 2023. (6). С. 1155404.
10. Magalhães T., Dinis-Oliveira R. J., Taveira-Gomes T. Digital Health and Big Data Analytics: Implications of Real-World Evidence for Clinicians and Policymakers // International Journal of Environmental Research and Public Health. 2022. № 14 (19).
11. Mehta N., Pandit A., Shukla S. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study // Journal of Biomedical Informatics. 2019. (100).
12. Norori N. [и др.]. Addressing bias in big data and AI for health care: A call for open science // Patterns. 2021. № 10 (2).
13. Razzak I., Eklund P., Xu G. Improving healthcare outcomes using multimedia big data analytics. // Neural computing & applications. 2022. № 17 (34). С. 15095-15097.
14. Wang T. H., Tsai Y. T., Lee P. C. Health big data in Taiwan: A national health insurance research database // Journal of the Formosan Medical Association. 2023. № 4 (122). С. 296-298.
15. Wu J. [и др.]. The challenge of healthcare big data to China’s commercial health insurance industry: evaluation and recommendations // Springer. 2022. № 1 (22). С. 1189.
16. Wu J. [и др.]. Leveraging nationwide epidemiological big data for diabetes modeling and disease surveillance: an innovative approach for advancing public health policy in // Springer. 2025. № 1 (18). С. 1189.