International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Integrating Social Determinants of Health into Predictive Models: Assessing How Dremio Can Aggregate Diverse Data Sources to Enhance Predictive Modeling in Healthcare

Saurabh Gupta

Abstract

This research suggests that SDOH can improve healthcare prediction models and mitigate health disparities. Predict at-risk individuals and improve health using non-clinical factors including money, education, housing, and care. Data isolation, privacy limitations, and information fusion make adoption challenging. The research emphasizes SDOH, sectoral coordination, and moral data preservation. Use modern data technology, streamline data transmission, and address SDOH's long-term healthcare impacts. Further research should focus on practical issues, enhance merging processes, and evaluate the reduction of healthcare inequalities in SDOH.

Keywords: Social Determinants of Health, Data isolation, information fusion, privacy limitations and modern data technology.

I. INTRODUCTION

Clinical decision-making, therapy tailoring, and resource management have improved using predictive modeling [1]. Quality and variety of data determine how well and accurately these models work. Previous healthcare prediction models used medical imaging, EHRs, and test results [2]. These important data sources rarely take into account changes in health outcomes. People know that socioeconomic determinants of health (SDOH), including poverty, education, housing, and healthcare, affect health outcomes and inequities [3]. These non-clinical data increase prediction models' accuracy, utility, and fairness [5].

Market, government, and community health records might include fragmented data [4]. Organizing and maintaining so much predictive modeling data is costly and complicated. A smart data technology like Dremio may help. A fast data lake engine, Dremio, mixes structured and unstructured data from many sources without alteration [6, 7]. Streamlining data integration makes SDOH easier to utilize in predictive models, increasing healthcare outcomes. This study analyzes how Dremio can enhance healthcare prediction models by combining healthcare and non-healthcare data like SDOH. The research demonstrates the relevance of social factors in healthcare analytics and investigates how sophisticated data platforms can overcome technological challenges when merging data types.

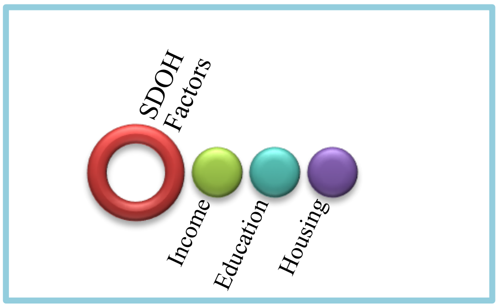

Figure 1 SDOH factors

a) Problem statement

Research reveals SDOH controls 80% of health outcomes. To ensure more accurate and fair assessments, include them [8]. SDOH data is disorganized, making prediction models difficult. It is available in many government and community health databases [8, 10]. The lack of proper SDOH data collection and organization makes real-time use problematic for healthcare workers and policymakers [11]. Improved technologies facilitate the swift integration of SDOH and clinical data. A data lake engine effortlessly combines structured and unstructured data. Few studies have examined how SDOH enhances Dremio forecasting algorithms, highlighting further investigation is needed.

b) Research Questions

- How can Dremio enhance healthcare prediction models by combining clinical and SDOH data?
- How does incorporating Dremio-compiled SDOH data influence healthcare model accuracy and predictability relative to clinical-only models?
- What issues arise when using Dremio to merge SDOH and clinical data for real-time healthcare prediction modelling?

c) Research objectives

- To evaluate Dremio's ability to acquire healthcare data from SDOH and other sources.
- To examine how Dremio-sourced SDOH data affects predictive healthcare model accuracy and performance.
- To investigate Dremio's technical problems and potential limitations for real-time data-based predictive healthcare modeling.

II. Literature review

Healthcare is increasingly using predictive modeling to enhance patient outcomes, resource management, and business efficiency. These models use EHRs, medical images, and test results to give crucial patient health data [12]. Most people feel clinical data does not disclose all health elements. Economic position, education, housing, and healthcare access are examples of socioeconomic determinants of health (SDOH) that have an impact on health outcomes [14]. According to recent research, SDOH affects 80% of health outcomes. We must incorporate them into more accurate and equitable prediction models [15].

SDOH-based healthcare approaches are flawed. SDOH is often disorganized, spread across several systems, and reported differently from clinical data, making integration problematic [13]. User data, community health records, and government systems all contain SDOH data, each with its own formats and standards [3]. Fragmented data is difficult to combine and use in real time. SDOH data often has mismatched locations and formats. This hampers healthcare workers' data use [11].

Smart data solutions like Dremio can solve these challenges. Dremio is a fast data lake engine that combines structured and unstructured data from various sources. Clinicians may use clinical data and SDOH to enhance prediction models [4]. Working in real time and transporting data as much as possible provides Dremio with a viable data aggregation method [15]. Many studies show that Dremio and other new data platforms accelerate data integration. Examining big, diversified datasets takes less time and resources [17].

However, the process of using Dremio to SDOH to predict models is new and has not yet been tested. SDOH data provide a fuller picture of a patient's health, enhancing prediction models [16]. Better models predict hospital return rates and chronic illness outcomes using clinical data [18, 19, and 20]. Most people realize SDOH is important for healthcare prediction models. Complex and complicated data can complicate operations. Dremio simplifies data integration and may work [21]. Further research is required to explore how these technologies address real-time data integration challenges and impact the accuracy and performance of healthcare prediction models.

III. MATERIALS AND METHODS

a) Research Design

This research uses secondary data to show how adding social determinants of health (SDOH) to prediction models improves health outcomes. The project aggregates clinical and SDOH data from different previous articles. The study evaluates how data collection methods affect secondary data prediction models.

b) Data Collection

This study employs secondary data sources, gathering data from previous research articles through various databases. These data on income, education, housing, employment, and healthcare access help explain how SDOH affects health.

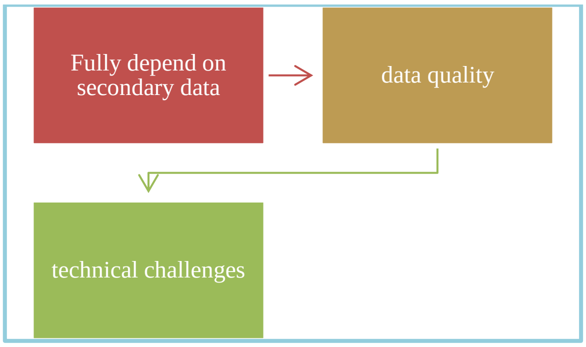

Figure 2 Data Collection Barriers

c) Inclusion and Exclusion Criteria

A systematic review choose relevant and high-quality publications and datasets using certain criteria. Include research that meets these criteria:

- Science-based study.
- Since 2023 and before
- Peer-reviewed journals.
- English

d) Extracting and assembling data

The current study extracts data from selected research and employs a single extraction technique to standardize data collection. In the summary of findings, qualitative analysis identifies themes, literature, and research needs. A standard data extraction form helps ensure data collection consistency. This reveals common themes, literature gaps, and research opportunities.

e) Ethical Consideration

This study's extensive public literature and data analysis reduce ethical hazards. The research follows systematic review ethics. Using just secondary data limits the study's scope. The organized grading and evaluation method reduces research bias due to analysis's subjectivity.

IV. Results

Enhanced prediction precision, variations in health outcomes, challenges in merging data, and sophisticated platforms emerged.

Theme 1: Accurate predictions

Several studies show SDOH improves healthcare model predictions. Studies found income, education, living security, and transportation more predictive than clinical data. SDOH models predicted hospital readmissions and chronic disease beginnings better in disadvantaged areas [22]. These results show the necessity of a comprehensive prediction method that includes patients' social and clinical characteristics.

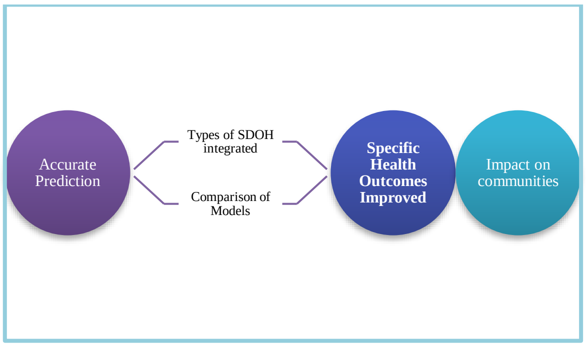

Figure 3 Accurate predictions

Theme 2: Health Care Results Variations

SDOH's involvement in eliminating healthcare disparities is also important. Studies show that SDOH factors especially race and ethnicity, socioeconomic status, and region impact health. Even with clinical features, low-income and disadvantaged groups had worse health outcomes [23]. Healthcare organizations must customize treatments for high-risk groups using SDOH data to enhance predictions and minimize health disparities.

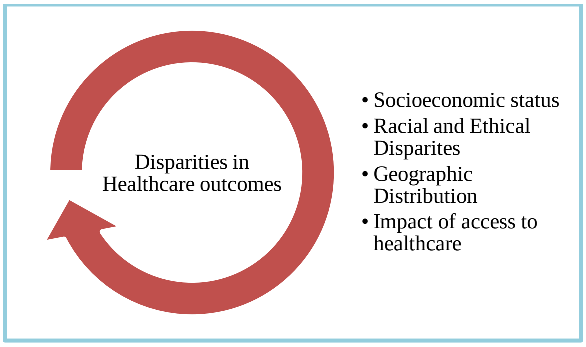

Figure 4 Health Care Results Variations

Theme 3: Challenges of data integration

Several investigations found broken data, inadequate SDOH data, and clinical-non-clinical integration [24]. For prediction models, healthcare data on stable housing and food poverty may not be available or standardized. Privacy concerns and HIPAA hampered the integration of private SDOH data into the healthcare system. These challenges suggest improved data exchange and privacy methods.

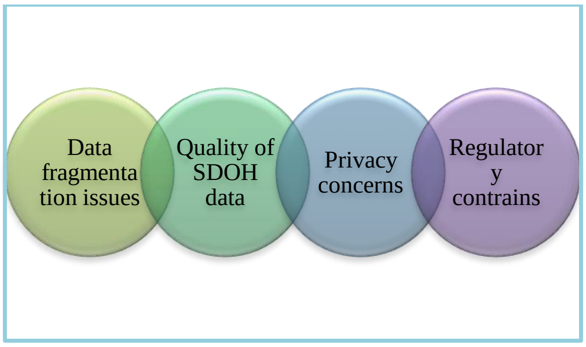

Figure 5 Challenges of data integration

Theme 4: High-tech data platform feature

Many studies recommended sophisticated data platforms that can handle structured and unstructured data to seamlessly integrate data sources [23]. Dremio, which provides unmodified real-time data transmission, enhances predictive modeling efficiency and scalability. These tools allowed healthcare businesses to use SDOH and clinical data to model non-clinical components and manage complicated health conditions.

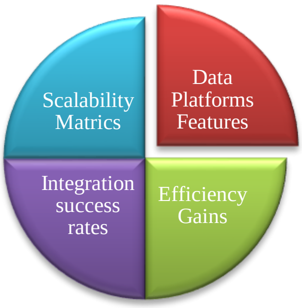

Figure 6 The Role of Advanced Data Platforms

V. DISCUSSION

This research found that SDOH improves health prediction systems. SDOH improves healthcare breakdown prediction and management. Data integration is hard. SDOH models outperform clinical data models. Non-clinical factors, including income, education, housing, and transportation, make high-risk health models more responsive [19]. Regional poverty and unstable housing can predict hospital readmissions better than clinical data [4]. These findings support Braveman and SDOH, which explain 80% of health outcome variability. This highlights their value in health care prediction models [5].

The research found SDOH enhances healthcare prediction models. When compared to clinical data-only models, factors such as income, education, and housing significantly improve prediction accuracy, especially for disadvantaged individuals [7]. This combination reduces healthcare inequities and assists the poor [9]. But data duplication and privacy problems limit it [6]. Real-time data integration is easier using Dremio, making models more scalable and efficient [10]. Privacy is a problem since HIPAA requires cautious SDOH data handling [14]. Research found that physical and social models improve health.

VI. CONCLUSION

Predictive models that integrate wealth, education, housing, and access to treatment are more accurate and help eliminate healthcare inequities, especially for poor groups. Benefits are clear, but data isolation, privacy issues, and technical barriers to database integration persist. Dremio and other advanced data tools can combine and analyze health and social data in real time, which is promising. Healthcare modeling may use SDOH, but it must address practical concerns and private data use. The study shows that healthcare must address physical and social issues to enhance results and fairness.

VII. RECOMMENDATIONS AND FUTURE RESEARCH

This research suggests various ways to enhance predictive healthcare models' social health component. Healthcare industries should standardize SDOH data collection and availability. Social services and healthcare should share more data. HIPAA-compliant SDOH data needs security. Standards are required for SDOH in prediction models, data separation, and healthcare-non-health sector interaction. SDOH raises permission and data ownership ethical issues that require more study. Research should concentrate on social concerns and communities.

References

1. F. Amrollahi, S. P. Shashikumar, A. Meier, L. Ohno-Machado, S. Nemati, and G. Wardi, "Inclusion of social determinants of health improves sepsis readmission prediction models," J. Am. Med. Inform. Assoc., vol. 29, no. 7, pp. 1263–1270, 2022.
2. K. J. T. Craig, N. Fusco, T. Gunnarsdottir, L. Chamberland, J. L. Snowdon, and W. J. Kassler, "Leveraging data and digital health technologies to assess and impact social determinants of health (SDoH): a state-of-the-art literature review," Online J. Public Health Inform., vol. 13, no. 3, 2021.
3. Ebbehoj, M. Ø. Thunbo, O. E. Andersen, M. V. Glindtvad, and A. Hulman, "Transfer learning for non-image data in clinical research: a scoping review," PLOS Digit. Health, vol. 1, no. 2, p. e0000014, 2022.
4. V. Ehrenstein, H. Kharrazi, H. Lehmann, and C. O. Taylor, "Obtaining data from electronic health records," in Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 [Internet]. Agency for Healthcare Research and Quality (US), 2019.
5. S. C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, "Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines," NPJ Digit. Med., vol. 3, no. 1, p. 136, 2020.
6. Kelley, Public Health Evaluation and the Social Determinants of Health. Routledge, 2020.
7. S. Kino, Y. T. Hsu, K. Shiba, Y. S. Chien, C. Mita, I. Kawachi, and A. Daoud, "A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects," SSM-Popul. Health, vol. 15, p. 100836, 2021.
8. T. M. Krause, C. Schaefer, and L. Highfield, "The association of social determinants of health with health outcomes," Am. J. Manag. Care, vol. 27, no. 3, pp. e89–e96, 2021.

9. E. Kriegova, M. Kudelka, M. Radvansky, and J. Gallo, "A theoretical model of health management using data-driven decision-making: the future of precision medicine and health," J. Transl. Med., vol. 19, p. 1-12, 2021.
10. K. N. Lopez et al., "Addressing social determinants of health and mitigating health disparities across the lifespan in congenital heart disease: a scientific statement from the American Heart Association," J. Am. Heart Assoc., vol. 11, no. 8, p. e025358, 2022.
11. M. M. Malik, S. Abdallah, and M. Ala'raj, "Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review," Ann. Oper. Res., vol. 270, no. 1, pp. 287–312, 2018.
12. V. Mhasawade, Y. Zhao, and R. Chunara, "Machine learning and algorithmic fairness in public and population health," Nat. Mach. Intell., vol. 3, no. 8, pp. 659–666, 2021.
13. J. Phuong et al., "Extracting patient-level social determinants of health into the OMOP common data model," in AMIA Annu. Symp. Proc., vol. 2021, p. 989, 2021.
14. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," NPJ Digit. Med., vol. 1, no. 1, pp. 1–10, 2018.
15. M. J. Rattermann, A. Angelov, T. Reddicks, and J. Monk, "Advancing health equity by addressing social determinants of health: Using health data to improve educational outcomes," PLoS One, vol. 16, no. 3, p. e0247909, 2021.
16. S. Raza, E. Dolatabadi, N. Ondrusek, L. Rosella, and B. Schwartz, "Discovering social determinants of health from case reports using natural language processing: algorithmic development and validation," BMC Digit. Health, vol. 1, no. 1, p. 35, 2023.
17. K. Salman, J. M. Turek, and C. Donovan, "Community health and wellness fair: a strategy for assessment of social determinants of health, inclusion and engagement of newcomers," J. Community Health, vol. 46, no. 3, pp. 527–537, 2021.
18. J. L. Snowdon et al., "Evaluating social determinants of health variables in advanced analytic and artificial intelligence models for cardiovascular disease risk and outcomes: A targeted review," Ethn. Dis., vol. 33, no. 1, pp. 33–43, 2023.
19. C. White-Williams et al., "Addressing social determinants of health in the care of patients with heart failure: a scientific statement from the American Heart Association," Circulation, vol. 141, no. 22, pp. e841–e863, 2020.
20. R. Wrembel, "Data integration, cleaning, and deduplication: Research versus industrial projects," in Int. Conf. Inf. Integr. Web, Cham: Springer Nature Switzerland, 2022, pp. 3–17.
21. C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review," J. Am. Med. Inform. Assoc., vol. 25, no. 10, pp. 1419–1428, 2018.
22. M. Y. Yang, Evaluating the Impact of Social Determinants of Health on Prediction of Clinical Outcomes in the Intensive Care Unit (Doctoral dissertation, Massachusetts Institute of Technology), 2023.
23. M. Y. Yang, G. H. Kwak, T. Pollard, L. A. Celi, and M. Ghassemi, "Evaluating the Impact of Social Determinants on Health Prediction," arXiv preprint arXiv: 2305.00065, 2023.
24. Y. Zhang et al., "Assessing the impact of social determinants of health on predictive models for potentially avoidable 30-day readmission or death," PLoS One, vol. 15, no. 6, p. e0235064, 2020.