Multifaceted Aging Clock (MAgiC): Navigating Alzheimer's and Dementia Pathways with Real-World Data-Driven Aging Markers using Explainable and Actionable AI

Md Kamruz Zaman Rana[1,2[0000-0003-0122-7672]] and Abu S.M. Mosa[1,2,*[0000-0002-8956-1466]]

[1]Institute for Data Science and Informatics, University of Missouri, Columbia MO 65203, USA
[2]Biomedical Informatics, Biostatistics and Medical Epidemiology, University of Missouri, Columbia MO 65203, USA
mrkfw@missouri.edu, mosaa@missouri.edu *Dissertation Supervisor

Abstract. The increasing elderly population and associated multimorbidity have spurred studies to use computational models for exploring the roles of epigenetic, cellular, metabolic, clinical, and social aging markers. This research recognizes the gap in a comprehensive predictive model utilizing vast real-world observational data against these diverse isolated biomarkers from disparate and rare data sources. This research aims to redefine aging beyond mere chronology by identifying clinical signals related to multidimensional biomarkers from structured and unstructured Electronic Health Records (EHR) and developing a multifaceted aging clock model. This model aims to identify and forecast the causal mechanisms impacting aging and disorders like Alzheimer's Disease and Related Dementia (ADRD), facilitating a shift to a multidimensional aging view. This patient-centered approach might uncover variances between data-driven and conventional knowledge, offering new biomarkers, reshaping medical paradigms, enhancing risk assessment, and leading to personalized interventions, thus enhancing EHR's role in aging management.

Keywords: Aging Clock, Multimorbidity, Causal GNN, Relational Deep Learning, Alzheimer

1 The Problem and Its Importance

1.1 Background

Aging is a multifaceted process that instigates extensive biological, physiological, and social alterations[1]. These alterations are essential in health research to understand the progression toward multimorbidity—multiple chronic conditions in an individual[2]. Studies have yielded the development of aging clocks, designed to estimate biological age rather than chronological age by analyzing various biomarkers associated with the aging process. Aging biomarkers are vital for revealing the functional and chronological aging states of various body systems, aiding in comprehending aging's

intricate influence on health[3]. Incorporating patient-specific care tailored to individual aging profiles and multimorbidity patterns can significantly enhance treatment effectiveness[4, 5]. Real-world observational data (RWD) aids in validating scientific theories by providing representative, real-life patient information, covering various aspects of health care, and offering evidence throughout disease progression[6].

1.2 Problem

The quest to understand and mitigate the effects of aging has led to the identification of numerous aging biomarkers across various biological systems. These biomarkers promise to unveil the intricate mechanisms of aging, providing insights that could lead to improved health outcomes and longevity. While a broad spectrum of biomarkers has been developed, their integration into a coherent framework that enhances predictive accuracy and clinical utility is lacking[7]. Current research often explores biomarkers in isolation or narrow contexts, failing to capture the comprehensive picture of an individual's aging process[3]. This siloed approach overlooks the interplay between different biomarkers and the systemic nature of aging. Moreover, the methodologies employed in studying aging biomarkers frequently encounter challenges, such as cohort selection bias inherent in cross-sectional studies[8]. Although longitudinal studies provide a valuable perspective by tracking changes over time, they are not without their own limitations, including high costs, extended durations, and the complexity of integrating their findings with other types of research. The field lacks a global strategy that aligns longitudinal data with cross-sectional and experimental findings to create a unified aging model. This integrated perspective is essential for uncovering composite biomarkers that provide a more accurate and clinically valuable prediction of aging trajectories and related health outcomes[9].

Electronic Health Records (EHR) provide a vast pool of RWD, and when fused with deep learning, they offer immense potential in detecting patterns in complex, extensive healthcare datasets[10, 11]. Yet, the full potential of utilizing biomarkers from EHR for clinical decision-making and research purposes has limitations due to data quality, incompleteness, and lack of standardization and interoperability[12, 13]. These include the need for advanced methodologies to harmonize and analyze diverse data types—ranging from clinical notes to laboratory results—ensuring that the integration contributes meaningfully to biomarker discovery and validation.

A significant disconnect exists between aging biomarker discovery and their clinical application. The slow translation of research into clinical practice is impeded by the absence of frameworks for validating and applying these biomarkers in patient care. Bridging this divide is crucial to fully harnessing biomarkers' potential in forecasting and addressing aging-related conditions.

1.3 Importance

According to the World Health Organization (WHO), Projected shifts in global demographics indicate that the population aged over 60 will climb from 12% to 22% by 2050[14]. Multimorbidity intensifies with age, seen in 30% of individuals aged 45-

64, 65% of those 65-84, and 82% in the 85 and older group[2]. Research has shown that multimorbidity is associated with an increased risk of mortality and low quality of life (QoL)[15]. This trend significantly intensifies the disease burden on healthcare systems worldwide, underscoring the need for innovative healthcare approaches tailored to understand the aging mechanism and its role in multimorbidity. Current research on multimorbidity shows a correlation with age, gender[16], early onset[17], socio-demographic, lifestyle, and psychological factors[18], highlighting the imperative for individualized healthcare strategies.

Composite aging biomarkers synthesize various biological parameters to provide a nuanced, holistic view of aging. They significantly surpass individual biomarkers in predicting age-related physiological changes and health outcomes. By integrating diverse data, they enhance predictive accuracy and clinical utility, aiding in early detection, monitoring of aging progression, and improved management of age-related disorders[3, 7, 9, 19–25]. Additionally, research shows that incorporating aging biomarkers into EHRs significantly boosts diagnostic precision, treatment personalization, and patient outcomes, offering a detailed view of individual health trajectories and aging impacts, which aids in early detection, targeted intervention, and monitoring of age-related health changes[26–28]. Leveraging AI and NLP to extract and utilize aging biomarkers from EHRs enables the efficient, accurate analysis of longitudinal, multi-dimensional health data across diverse populations, providing deeper insights into aging than isolated clinical studies can offer[29–31].

2 Goal and Research Questions

2.1 Development of an Aging Clock from the RWD

The first goal of the research is to identify and process the required structured and unstructured data from EHR to extract clinical and social signals associated with established aging biomarkers and develop a multifaceted aging clock. This step is pivotal in advancing our understanding of aging as a multifaceted phenomenon and setting the stage for clinical applications. This goal will answer the following research questions-

- What structured and unstructured data elements within EHRs are most relevant for identifying clinical and social signals associated with established aging biomarkers?
- How can we effectively process and integrate these varied data types from EHRs to construct a comprehensive and accurate aging clock?
- What methodologies or technologies are most effective in extracting and analyzing these data points from EHRs to ensure reliability and validity in the context of aging research?
- After identifying aging biomarker data within EHRs, which predictive methodologies should be utilized to construct an accurate and comprehensive multifaceted aging clock?

2.2 Clinical Validation, Fairness, and Utility Assessment Navigating ADRD Pathways

We will apply the developed aging clock model to assess and predict disease onset, progression, and multimorbidity patterns in Alzheimer's patients, demonstrating its utility.

- How can the aging clock model be applied to accurately assess and predict the progression of Alzheimer's disease and associated multimorbidity patterns in patients?
- What specific features of the aging clock model contribute most effectively to mapping the disease trajectory of Alzheimer's patients?
- What metrics will evaluate the aging clock model's predictive accuracy regarding Alzheimer's progression and multimorbidity patterns?

2.3 Deriving Actionable Insights for Alzheimer's Disease Interventions from the Aging Clock

Our final aim is to extract causal pathways from our aging clock, explain them to providers, and develop actionable strategies for preventing and managing Alzheimer's disease. This involves translating our model's findings into potential preventive measures, treatment options, and lifestyle changes directly impacting patient care and outcomes.

- How can causal pathways be extracted from the aging clock model to identify key factors influencing the progression of Alzheimer's disease?
- What methodologies can translate the aging clock model's findings into actionable preventive measures for Alzheimer's disease?
- What strategies can be designed to evaluate the effectiveness of the interventions derived from the aging clock model in managing Alzheimer's disease?

3 Approach and Methods

3.1 Approach for Development of an Aging Clock from the RWD

Method for Identifying Established Aging Biomarkers. A thorough review of the literature is underway to identify recognized and validated biomarkers of aging. Consultations with a panel of experts will be conducted to refine and ensure the accuracy of the biomarker list. This process also includes the identification of biomarkers relevant to Alzheimer's disease, which will support future efforts in model validation.

Method for Mapping Selected Biomarkers to EHR Data Elements. Translate each identified biomarker into corresponding EHR data elements by identifying relevant

clinical tests, procedures, laboratory values, medication records, and vital signs indicative of each biomarker's presence. Evaluate the routine clinical recording of these data points to assess the extraction feasibility and ensure prioritization based on data availability and reliability. Validate with clinical experts to confirm that the identified data accurately reflect the biomarkers' clinical manifestations.

Method for Automated Data Extraction. Implements data extraction algorithms tailored to EHR systems to retrieve corresponding to the identified biomarkers. This will involve a Generative Pre-trained Transformer (GPT) to parse and extract relevant information from unstructured data and data mining techniques for structured data.

Method for Loyal Cohort Selection. We will select patients to establish a loyal cohort based on the adequacy of biomarker-related data, which will be informed by the number of clinical events, duration, and visit frequency associated with these biomarkers.

Method for Modeling the Aging Clock. To model the aging clock, we will utilize a graph network-based deep learning approach, where clinical observations, patient data, and biomarkers are integrated with a visit timestamp to construct a temporal, heterogeneous graph. This model will leverage healthcare data's relational and temporal nature, exploring advanced methodologies such as Relational Deep Learning (RDL) and Causal-Based Graph Neural Networks (CausalGNN) to accommodate the data's dimensionality and complexity[32–34].

3.2 Approach for Clinical Validation, Fairness, and Utility Assessment Navigating ADRD Pathways

The model will be applied to longitudinal patient data to predict key milestones in the progression of Alzheimer's disease. We will use survival analysis techniques, such as Cox proportional hazards models or accelerated failure time models, to assess the model's ability to predict time-to-event outcomes related to Alzheimer's disease. Key predictors within the model will be identified using SHAP (SHapley Additive exPlanations) and permutation importance, pinpointing critical biomarkers influencing the disease trajectory. Model performance will be evaluated using Precision-Recall AUC, C-index, Brier Score, and other matrices.

3.3 Approach for Deriving Actionable Insights for Alzheimer's Disease Interventions from the Aging Clock

We will extract causal relationships from the aging clock model using advanced causal inference techniques, such as counterfactual reasoning and structural equation modeling. This process will identify which biomarkers and genetic markers are most influential in the progression of Alzheimer's disease. We will apply knowledge translation methodologies to convert the model's findings into practical preventive measures and treatment options. This includes developing guidelines for lifestyle

adjustments, dietary recommendations, and pharmacological interventions based on the identified causal factors.

Acknowledgments. This study is not funded.

Disclosure of Interests. The authors have no competing interests to declare relevant to this article's content.

References

1. Cannon ML (2015) What is aging? Disease-a-Month 61:454–459. https://doi.org/https://doi.org/10.1016/j.disamonth.2015.09.002
2. Skou ST, Mair FS, Fortin M, et al (2022) Multimorbidity. Nat Rev Dis Primers 8:48. https://doi.org/10.1038/s41572-022-00376-4
3. Bao H, Cao J, Chen M, et al (2023) Biomarkers of aging. Science China Life Sciences 2023 66:5 66:893–1066. https://doi.org/10.1007/S11427-023-2305-0
4. Roma MFB, Garção NIB, Covinsky KE, Aliberti MJR (2022) Delivering patient-centered care: outcomes of comprehensive geriatric assessment across healthcare settings. Geriatrics Gerontology and Aging 16:1–12. https://doi.org/10.53886/gga.e0220031
5. Liang H (2022) The Impact of Patient-Centered Care. Springer Nature Singapore, Singapore
6. Sun X (2022) Real-World Data and Real-World Evidence. In: Advanced Statistics in Regulatory Critical Clinical Initiatives. Chapman and Hall/CRC, Boca Raton, pp 105–132
7. Chen R, Wang Y, Zhang S, et al. (2023) Biomarking: Current state-of-art, challenges, and opportunities. MedComm – Future Medicine 2:. https://doi.org/10.1002/mef2.50
8. Nelson PG, Promislow DEL, Masel J (2020) Biomarkers for Aging Identified in Cross-sectional Studies Tend to Be Non-causative. The Journals of Gerontology: Series A 75:466–472. https://doi.org/10.1093/GERONA/GLZ174
9. Ader I, Pénicaud L, Andrieu S, et al (2021) Healthy Aging Biomarkers: The INSPIRE's Contribution. J Frailty Aging 10:313. https://doi.org/10.14283/JFA.2021.15
10. Spratt HM, Arbet J, Brokamp C, et al (2021) Lessons and tips for designing a machine learning study using EHR data. J Clin Transl Sci 5:e21. https://doi.org/10.1017/CTS.2020.513
11. Nadkarni GN, Fleming F, Mccullough JR, et al (2019) Prediction of rapid kidney function decline using machine learning combining blood biomarkers and electronic health record data. bioRxiv 587774. https://doi.org/10.1101/587774
12. Berger ML, Curtis MD, Smith G, et al (2016) Opportunities and challenges in leveraging electronic health record data in oncology. http://dx.doi.org/102217/fon-2015-0043 12:1261–1274. https://doi.org/10.2217/FON-2015-0043
13. Ryan D, Blakey J, Chisholm A, et al (2017) Use of electronic medical records and biomarkers to manage risk and resource efficiencies. Eur Clin Respir J 4:. https://doi.org/10.1080/20018525.2017.1293386

14. Ageing and health. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Accessed 2 Apr 2024
15. Ratushnaya L, Gabelaia D, Petrova S, et al (2023) Investigation of the Causes and Ways of Reducing Injuries in Age-Related Patients with Multimorbidity. Journal of Complementary Medicine Research 14:66. https://doi.org/10.5455/jcmr.2023.14.02.10
16. Brijoux T, Woopen C, Zank S (2021) Multimorbidity in old age and its impact on life results. Z Gerontol Geriatr 54:108–113. https://doi.org/10.1007/S00391-021-01920-9
17. Nicholson K, Griffith LE, Sohel N, Raina P (2021) Examining early and late onset of multimorbidity in the Canadian Longitudinal Study on Aging. J Am Geriatr Soc 69:1579–1591. https://doi.org/10.1111/JGS.17096
18. Salive ME, Suls J, Farhat T, Klabunde CN (2021) National Institutes of Health Advancing Multimorbidity Research. Med Care 59:622–624. https://doi.org/10.1097/MLR.0000000000001565
19. Chan MS, Arnold M, Offer A, et al (2021) A Biomarker-based Biological Age in UK Biobank: Composition and Prediction of Mortality and Hospital Admissions. The Journals of Gerontology: Series A 76:1295–1302. https://doi.org/10.1093/GERONA/GLAB069
20. Schalkamp A-K, Lerche S, Wurster I, et al (2022) Machine learning-based personalized composite score dissects risk and protective factors for cognitive and motor function in elderly. medRxiv 2022.11.18.22282498. https://doi.org/10.1101/2022.11.18.22282498
21. Sagers L, Melas-Kyriazi L, Patel CJ, Manrai AK (2020) Prediction of chronological and biological age from laboratory data. Aging 12:7626–7638. https://doi.org/10.18632/AGING.102900
22. Bafei SEC, Shen C (2023) Biomarkers selection and mathematical modeling in biological age estimation. npj Aging 2023 9:1 9:1–9. https://doi.org/10.1038/s41514-023-00110-8
23. Swami U, Nussenzveig RH, Agarwal N (2020) Quest for Ideal Composite Biomarkers for Response to Immunotherapies. Clinical Cancer Research 26:5059–5061. https://doi.org/10.1158/1078-0432.CCR-20-2321/274651/AM/QUEST-FOR-IDEAL-COMPOSITE-BIOMARKERS-FOR-RESPONSE
24. Liu Z (2021) Development and Validation of 2 Composite Aging Measures Using Routine Clinical Biomarkers in the Chinese Population: Analyses From 2 Prospective Cohort Studies. The Journals of Gerontology: Series A 76:1627–1632. https://doi.org/10.1093/GERONA/GLAA238
25. Satoh K, Sato-Ueshima M, Kagami-Katsuyama H, et al (2023) Profiles of the composite biomarker specific for β-amyloid accumulation in the brain relevant to age and sex. medRxiv 2023.01.13.23284514. https://doi.org/10.1101/2023.01.13.23284514
26. Ubaida-Mohien C, Tanaka T, Tian Q, et al (2023) Blood Biomarkers for Healthy Aging. Gerontology 69:1167–1174. https://doi.org/10.1159/000530795
27. Piau A, Steinmeyer Z, Cesari M, et al (2021) Intrinsic Capacitiy Monitoring by Digital Biomarkers in Integrated Care For Older People (ICOPE). Journal of Frailty and Aging 10:132–138. https://doi.org/10.14283/JFA.2020.51/METRICS

28. Noaeen M, Amini S, Bhasker S, et al (2023) Unlocking the Power of EHRs: Harnessing Unstructured Data for Machine Learning-based Outcome Predictions. medRxiv 2023.02.13.23285873. https://doi.org/10.1101/2023.02.13.23285873
29. Priya S, Peluso A, Shekhar MC, et al (2023) Abstract 4191: Assessing performance of biomarker extraction from electronic health records: Data augmentation methods for a hierarchical self-attention network (HiSAN). Cancer Res 83:4191–4191. https://doi.org/10.1158/1538-7445.AM2023-4191
30. Kuan V, Fraser HC, Hingorani M, et al (2021) Data-driven identification of ageing-related diseases from electronic health records. Scientific Reports 2021 11:1 11:1–17. https://doi.org/10.1038/s41598-021-82459-y
31. Andrew NE, Beare R, Ravipati T, et al (2023) Developing a linked electronic health record derived data platform to support research into healthy ageing. Int J Popul Data Sci 8:. https://doi.org/10.23889/IJPDS.V8I1.2129
32. Wein S, Malloni WM, Tomé AM, et al (2021) A graph neural network framework for causal inference in brain networks. Scientific Reports 2021 11:1 11:1–18. https://doi.org/10.1038/s41598-021-87411-8
33. Wang L, Adiga A, Chen J, et al (2022) CausalGNN: Causal-Based Graph Neural Networks for Spatio-Temporal Epidemic Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence 36:12191–12199. https://doi.org/10.1609/AAAI.V36I11.21479
34. Fey M, Hu W, Huang K, et al Relational Deep Learning: Graph Representation Learning on Relational Databases