Machine Learning and Longitudinal Biomarkers: Thematic Insights into Predicting Alzheimer Disease Progression

Sandip Basnet[1], Dr. Mandeep Kaur[2]
[1,2]*Department of Computer Science and Engineering*, *CT University,* Ferozpur Road Ludhiana, India

***Abstract*—Alzheimer's disease (AD) is the most prevalent cause of dementia globally, with more than 55 million individuals affected, and cases estimated to almost double by 2050. The heterogeneous course of AD poses major obstacles in early diagnosis, prognosis, and treatment planning. Whereas conventional diagnostic procedures depend on neuroimaging, cerebrospinal fluid biomarkers, and neuropsychological testing, these frequently cannot make predictions for early courses of disease. Machine learning (ML) advances have brought new possibilities for utilizing longitudinal biomarker data to model and predict AD progression with greater accuracy. This paper is a secondary data-based thematic analysis of the existing literature, datasets, and frameworks in AD prediction. Based on published research, especially using large-scale efforts like the Alzheimer's Disease Neuroimaging Initiative (ADNI), the article highlights five essential themes: (1) temporal dynamics of biomarker prediction, (2) missing and incomplete data management, (3) integration of multimodal biomarkers, (4) model interpretability and clinical acceptance, (5) generalizability and real-world application and (6) Clinical Integration and Deployment Challenges. Each theme highlights distinctive opportunities and challenges in the use of ML for predicting AD progression. The review points out how temporal models such as LSTM and Transformer networks have improved sequential biomarker interpretation and imputation and generative methods have reduced missing data problems. Integration of multimodal MRI, PET, CSF, and genetic markers has enhanced predictive accuracy, although transparency, overfitting, and clinical uptake issues continue to exist. The research concludes that explainable, multimodal, and federated ML models are the AD prognosis of the future, harmonizing computational accuracy with ethical and clinical considerations.**

***Keywords*—Alzheimer disease (AD), Machine Learning, Alzheimer's Disease Neuroimaging Initiative**

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that results in memory loss, cognitive impairment, and psychiatric disturbances. It is responsible for 60–80% of dementia in the world and causes significant emotional, social, and economic burdens on patients, caregivers, and the health system (WHO, 2022). Recent estimations indicate that over 55 million individuals around the globe live with dementia, and the number is projected to increase to 139 million by 2050 (WHO, 2021). The unknown pattern of AD development, where some patients deteriorate quickly while others stay stable, makes diagnosis and care planning challenging.

Early diagnostic routes mostly depend on clinical examinations, neuropsychological testing, neuroimaging (MRI, PET), and cerebrospinal fluid (CSF) examination of amyloid-β and tau proteins. Although these tools are useful, they trend to diagnose AD after the onset of symptoms and are not predictive of diseases trajectories. As a result, computational models have been increasingly favored that use biomarkers to predict disease progression earlier and more reliably.

Machine learning (ML), which is a subcategory of artificial intelligence, provides strong tools for processing high-dimensional and longitudinal data. RNNs, LSTM networks, and Transformer-based models are most suited for temporal data and make predictions about transitions from cognitively normal (CN) to mild cognitive impairment (MCI), and eventually to AD (Nguyen et al., 2020; Vafeidis et al., 2022). Moreover, explainable AI (XAI) techniques such as SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-Agnostic

Explanations) aim to bridge the loop between algorithmic prediction and clinician trust.

This research paper uses a secondary data collection method, examining previous research, datasets, and reports to let's examine more closely the way machine learning (ML) is revolutionizing the prediction of Alzheimer's disease (AD). Through an exploration of thematic analysis, we are able to distinguish between five overarching themes: temporal dynamics, missing data handling, merging multiple data types, interpretability, and generalizability. Collectively, these themes illuminated the possibilities and pitfalls of applying ML for predicting AD in everyday healthcare environments.

II. LITERATURE REVIEW

Alzheimer's disease (AD) is notoriously difficult to forecast because of its heterogeneity and the limitations of conventional diagnostic methods. In the last several decades, scientists have, more and more, resorted to machine learning (ML) approaches to categorize biomarker data and predict disease progression. Although the findings have been encouraging, the literature also identifies some of the methodological limitations, concerns over generalizability, as well as some doubt about how well these methods could be implemented in clinical practice. In the early years of computational research, traditional algorithms were used for multimodal biomarkers. For example, Zhang et al. (2011) integrated structural MRI, PET, and CSF biomarkers with Support

Vector Machines (SVM) and performed better than single-modality-based methods. Their results highlighted the value of the combination of imaging and biological measures, although linear classification constrained them from capturing the intricate, non-linear dynamics of disease. Subsequent research on top of this constructed ensemble models such as Random Forests and Gradient Boosting, which exploited the high-dimensional data yet were still challenged to incorporate the time dimension of longitudinal data (Li et al., 2015).

Liu et al. (2019) demonstrated that using Convolutional Neural Networks (CNNs) on MRI scans, combined with cognitive and clinical characteristics, greatly enhanced classifying individuals as cognitively normal, with Mild Cognitive Impairment (MCI), or diagnosed with AD in terms of accuracy. The study made a strong case for multimodal deep learning but employed aggressive preprocessing methods and computational power, and it is unclear if it will be feasible to implement it in practical health networks. Likewise, Khanna et al. (2021) employed attention networks to combine PET imaging and clinical information with greater sensitivity to the early phases of disease transition. While theoretically sound, their application of costly imaging modalities rules out use in low-resource environments. A second parallel research stream has investigated temporal modelling.

Alzheimer's onset is a dynamic process, naturally, so longitudinal information is more enlightening than cross-sectional static images. Nguyen et al. (2020) proposed time-sensitive Long Short-Term Memory (LSTM) networks for addressing variable lengths of time between patient consultations. Their performance was superior to conventional recurrent networks in demonstrating enhanced prediction of cognitive deterioration. LSTMs are still data-greedy and prone to overfitting, especially in small or imbalanced populations. New advancements using Transformer architectures have sought to address these disadvantages. Vafeidis et al. (2022) used a time-sensitive Transformer with attention, which indicated higher interpretability and accuracy than LSTMs. Nevertheless, as with most deep models, the black-box property of the model still exists, and thus concerns regarding trust and transparency are challenging in clinical use. Longitudinal databases are prone to having missing records due to patient dropout and non-uniform follow-ups for diagnosis. Simple imputation techniques like mean substitution warp variance and bias outcomes. More sophisticated techniques like Multiple Imputation by Chained Equations (MICE) have been used with reasonable success. Deep generative methods, specifically Generative Adversarial Imputation Nets (GAIN), presented a breakthrough by learning data distributions to impute missing biomarker values (Yoon et al., 2018). While these approaches enhance internal validity, they have in most cases been validated using benchmark collections like ADNI, raising issues about generalizability in actual clinical settings where missingness tends to be systematic rather than at-random.

One of the main strengths throughout the literature under review is the utilization of multimodal datasets, which capture AD's multi-factorial pathology. However, such datasets tend to be drawn from research organizations like ADNI or AIBL that are unrepresentative demographically. Models trained on these sources often perform poorly on external sets like NACC, and there is limited external validity. Furthermore, it is not common to have access to multimodal imaging and CSF biomarkers in regular clinical practice, lowering the usability of implementing such models at large.

Interpretability is another concern. Lundberg and Lee (2017) SHAP framework and Ribeiro et al. (2016) LIME technique has been used extensively to offer post-hoc explanations of model predictions. In AD research, SHAP values have identified hippocampal atrophy and tau protein concentration as key features, consistent with the established clinical findings (Lundberg and Lee, 2017). Although these approaches close the gap between the output of ML and medical reasoning, they are approximations and not clear representations of the model's decision process. Post-hoc interpretability is criticized to create false reassurance and therefore more inherently interpretable models might be necessary for clinical uptake.

Ultimately, generalizability and deployment issues are poorly explored. Most studies exhibit high accuracy within the ADNI dataset but experience dramatic performance reductions when applied to external cohorts (Khanna et al., 2021). Federated learning has been suggested as a possible solution, allowing distributed training institutions without centralization of patient data. This has the advantages of increased data diversity and protecting patient privacy but poses technical and regulatory complexities. Notably, the literature does not have enough translational studies on whether and how predictive models are implemented in clinical workflows, where cost, interpretability, and usability concerns may prove more important than marginal improvements in accuracy.

In general, the literature reveals significant progress in applying ML for predicting AD progression, that is, multimodal integration and temporal modelling. While it is seen some impressive advancements in research, many of these innovations still haven't made their way into clinical practice. It is still grappling with significant issues like managing missing data, ensuring external validity, aligning accuracy with interpretability, and tackling the hurdles of deploying solutions in real-world healthcare settings. The future of this field hinges less on new algorithms and more on blending methodological complexity with what's acceptable in clinical settings.

A. Comprative Review of AD prediction Methods:

A look at past studies reveals clear compromises among how complex a model is, how well it predicts, yet also how practical it works outside labs. Older techniques like SVMs or grouped models boosted diagnosis from multiple sources, however struggled tracking how illnesses evolve. Neural networks - like CNNs spotting patterns in space while Transformers handle timing details - not only scored higher in precision but sometimes made results easier to grasp, even if they needed more data plus heavier computing power.

Data completeness plus how well results apply elsewhere are still big issues. While generative imputation keeps key data traits, it usually doesn't perform well on outside groups where data is systematically absent. Even models built using detailed, multi-source data have trouble working reliably in everyday clinics. In total, upcoming research needs to focus more on time-based patterns, wider testing across sites, clearer explanations of outcomes, along with practical real-world use.

III. RESEARCH METHODOLOGY

1) Research Design: This study takes a qualitative approach, using thematic analysis of secondary data. Thematic analysis, as outlined by Braun and Clarke (2006), involves identifying, coding, and reporting patterns within the data. It's particularly effective for synthesizing findings from a wide array of secondary sources, including peer-reviewed literature, datasets, and global health reports.

2) Data Sources: The data is gathered secondary data from Alzheimer's Disease Neuroimaging Initiative (ADNI), Australian Imaging, Biomarkers and Lifestyle (AIBL), and National Alzheimer's Coordinating Center (NACC). Peer-reviewed research: Studies by Zhang et al. (2011), Nguyen et al. (2020), Garcia-Ribeiro et al. (2021), Liu et al. (2019), and others. Institutional documents: WHO dementia

statistics, NIH reports on AD biomarkers, and clinical guidelines.

3) Data Analysis: Thematic analysis is a qualitative method that helps identify patterns in data, allowing researchers to group findings into meaningful themes. In this study, it brings together various secondary research on predicting Alzheimer's, highlighting recurring issues like temporal modelling, missing data, and interpretability, while also critically connecting computational advancements with their clinical relevance. Thematic analysis helps integrate diverse evidence into conceptual themes that give equal weight to both computational and clinical implications. Unlike quantitative meta-analysis, which zeroes in on numerical results, this process emphasizes the richness of qualitative insights.

IV. ANALYSIS AND DISCUSSION

1) Temporal Dynamics in Biomarker Prediction: Longitudinal data follow disease development at numerous time points and are thus better than cross-sectional snapshots. Temporal dependence-management ML architectures such as LSTMs and Transformers have yielded promising results. Nguyen et al. (2020) demonstrated that time-aware LSTMs could predict patterns of cognitive decline better than standard RNNs, particularly when handling infrequent clinic visits. Similarly, Vafeidis et al. (2022) employed Transformer models with attention mechanisms to identify key time points in disease progression. These methods highlight the need for consideration of temporal dynamics to achieve valid prognosis.

2) Handling Missing and Incomplete Data: Longitudinal clinical data tend to have incomplete records due to patient dropout, irregular visitation, or test failure. Traditional imputation methods (mean substitution, k-NN) introduce bias, whereas advanced methods like Multiple Imputation by Chained Equations (MICE) and Generative Adversarial Imputation Nets (GAIN) achieve robustness (Yoon et al., 2018). For example, it has been demonstrated that GAIN is capable of imputing missing biomarker values more effectively by learning the data distribution. Handling missingness is therefore crucial to achieving robust model performance.

3) Multimodal Biomarker Integration: Alzheimer's pathology involves heterogeneous biological systems, and therefore multimodal data fusion is needed. Investigations integrating MRI, PET, CSF biomarkers, genetic data (e.g., APOE ε4), and neuropsychological test data have greater predictive capability (Zhang et al., 2011; Liu et al., 2019). Liu et al. (2019) introduced a hybrid model combining CNNs for imaging and fully connected layers for the clinical domain to enhance MCI-to-AD conversion prediction substantially. These multimodal methods emphasize the heterogeneity of AD and the need for comprehensive modelling.

4) Multimodal Biomarker Integration and Clinical Trust: Although deep learning models are highly accurate, their "black-box" design has hindered clinical uptake. Explainable AI (XAI) techniques such as SHAP and LIME have become powerful tools for interpretability (Lundberg and Lee, 2017; Ribeiro et al., 2016). SHAP analyses identify important biomarkers, such as hippocampal atrophy and abnormal tau levels that underpin model predictions, aligning computational results with known clinical knowledge. Attention mechanisms in Transformer models also increase interpretability by emphasizing diagnostically important time points (Garcia-Ribeiro et al., 2021).

5) Generalizabitlity and Real-World Deployment: In spite of progress, concerns about generalizability persist. The majority of models are trained on the ADNI dataset, which does not necessarily reflect globally diverse populations. Experiments demonstrate that models tend to perform suboptimal when tested on AIBL or NACC datasets (Khanna et al., 2021). Federated learning of training models on distributed datasets without central data has the potential to address this issue (Sheller et al., 2020). Ethical factors involving patient privacy, data security, and fairness considerations continue to be crucial to clinical deployment.

6) Clinical Integration and Deployment challenges: Even though methodological advancements in the prediction of Alzheimer's disease (AD) through machine learning have been significant, clinical integration is still limited. A common challenge across the reviewed literature is the disconnect between the

models developed and their deployment in a healthcare context. Architectures like LSTMs (Nguyen et al., 2020) and attention-based ones (Garcia-Ribeiro et al., 2021; Vafeidis et al., 2022) excel on benchmarking datasets like ADNI. But they are rarely tuned for clinical pipelines, where real-time processing, cost-effectiveness, and interpretability are just as valuable as predictive performance. Khanna et al. (2021) point out this issue in their research on deep fusion of PET and clinical data, citing that while multimodal models have better sensitivity, their reliance on expensive and invasive modalities limits clinical scalability. Likewise, Liu et al. (2019) cite that their multimodal CNN-based architecture obtained high accuracy, but was only tested within controlled research settings. One major hurdle as faced is dataset bias. A lot of the research out there relies on the ADNI dataset, which, while comprehensive, doesn't really reflect the diversity of patients around the world (Zhang et al., 2011). This means that models might not perform as well in real-world settings outside of academia or the Western context. For clinical adoption, it's crucial that we can generalize findings to support a wide range of patient populations. Trust in clinical settings hinges on explainability. Although tools like SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) provide some transparency, they only offer approximations and don't fully capture how models think. It's essential for models to strike a balance between performance and interpretability to meet the ethical and practical needs of both clinicians and patients.

V. CORE FINDINGS AND DISCUSSIONS

The five themes have been identified really highlight the transformative yet challenging nature of machine learning in predicting Alzheimer's disease. Temporal models are a big leap forward in understanding disease dynamics, but they depend on high-quality longitudinal data, which is often hard to come by. Dealing with missing data is a major challenge, especially since dropout rates in dementia trials can exceed 30%. The promise of multimodal integration is gigantic, but it calls on us to reconcile different sources of data, which is complicated by technical as well as ethical considerations. Interpretability is not a nice-to-have; it's an ethical and scientific imperative. Without transparency, even the most accurate models may be rejected by clinicians and patients. Explainable AI can assist in establishing trust, but it must strike a balance between simplicity and accuracy. Problems with generalizability highlight the need to validate models across cohorts and ensure that predictions can be used in heterogeneous populations, accounting for variations in ethnicity, geography, and socio-economic status. In the future, we ought to develop hybrid approaches blending computational intelligence with expert domain knowledge, federated learning architectures providing generalized and privacy-preserving learning, and personalized prediction platforms examining unique biomarker paths. These technologies, if developed, can potentially revolutionize the detection and intervention of Alzheimer's disease early on.

VI. CONCLUSION AND FINAL OUTCOMES

The research analysed in detail how machine learning is revolutionizing the task of forecasting the progression of Alzheimer's disease, based on thematic analysis of secondary data. We've identified five major themes: temporal dynamics, missing data handling, multimodal fusion, interpretability, and generalizability, which reflect both the possibilities and limitations of this area of work.

Our analysis illustrates that although sequence models such as LSTMs and Transformers have enhanced forecasting performance, they are hindered by incompleteness in data and generalizability. Multimodal models integrating imaging, biochemical, genetic, and clinical information present a complete picture but need sophisticated harmonization techniques. Knowledge of these models is important to establishing trust in clinical environments, with explainable AI (XAI) techniques varying from computational output to medical rationale. In the end, issues of generalizability and ethical implications regarding privacy and fairness are still challenging

When implementing these models in clinical realities. Looking to the future, the way forward for Alzheimer's disease prognosis is interpretable, multimodal, and federated machine learning models that hit the sweet spot of accuracy, interpretability, and fairness. By coordinating computational innovation with clinical requirements, these models will be able to transform early detection, improve patient outcomes, and reduce the worldwide burden of Alzheimer's disease

REFERENCES

[1] Alsadhan, N.A., 2025. Image-based Alzheimer's disease detection using pretrained convolutional neural network models. *arXiv preprint* arXiv:2502.05815.

[2] Durga Supriya, H.L., Thomas, S.M. and Sowmya Kamath, S., 2024. A multimodal approach integrating convolutional and recurrent neural networks for Alzheimer's disease temporal progression prediction. *In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, pp.5207–5215.

[3] Garcia-Ribeiro, S., Monteiro, A. & Marreiros, G., 2021. Attention-enhanced LSTM networks for Alzheimer's diagnosis from longitudinal data. Computer Methods and Programs in Biomedicine, 129, p.104139. https://doi.org/10.1016/j.cmpb.2021.104139

[4] Khanna, S., Varma, S. & Bhattacharya, S., 2021. Deep learning-based fusion of PET and clinical data for Alzheimer's progression prediction. Journal of Alzheimer's Disease, 81(4), pp.1353–1365. https://doi.org/10.3233/JAD-201394

[5] Li, F., Tran, L., Thung, K.H., Ji, S., Shen, D. & Li, J., 2015. A robust deep model for improved classification of AD/MCI patients. Computer Methods and Programs in Biomedicine, 122(2), pp.372–381.

[6] Liu, M., Zhang, J., Adeli, E. & Shen, D., 2019. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Transactions on Biomedical Engineering, 66(8), pp.2374–2385. https://doi.org/10.1109/TBME.2018.2889362

[7] Lundberg, S.M. & Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS). [online] Available at: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html [Accessed 21 Jun. 2025].

[8] Nawaz, A., Anwar, S.M., Liaqat, R., Iqbal, J., Bagci, U. and Majid, M., 2021. Deep convolutional neural network based classification of Alzheimer's disease using MRI data. arXiv preprint arXiv:2101.02876.

[9] Nguyen, M., He, T., An, L., Alexander, D.C., Feng, J. & Yeo, B.T., 2020. Predicting Alzheimer's disease progression using deep recurrent neural networks. NeuroImage, 222, p.117203. https://doi.org/10.1016/j.neuroimage.2020.117203

[10] Prakash, D., Madusanka, N., Bhattacharjee, S., et al., 2021. Diagnosing Alzheimer's disease based on multiclass MRI scans using transfer learning techniques. Current Medical Imaging, 17(12), pp.1460–1472.

[11] Ribeiro, M.T., Singh, S. & Guestrin, C., 2016. "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135–1144. https://doi.org/10.1145/2939672.2939778

[12] Tufail, A.B., Ma, Y.K., Kaabar, M.K.A., Rehman, A.U., Khan, R. and Cheikhrouhou, O., 2021. Classification of initial stages of Alzheimer's disease through PET neuroimaging modality and deep learning: Quantifying the impact of image filtering approaches. Mathematics, 9(23), p.3101.

[13] Vlontzou, M.E., Athanasiou, M., Dalakleidi, K.V., Skampardoni, I., Davatzikos, C. and Nikita, K., 2025. A comprehensive interpretable machine learning framework for mild cognitive impairment and Alzheimer's disease diagnosis. Scientific Reports, 15(1), p.8410.

[14] Vafeidis, D., Costantinides, F., Perna, D. & Rittman, T., 2022. Time-aware transformer-based model for Alzheimer's disease progression prediction. IEEE Transactions on Medical Imaging. https://doi.org/10.1109/TMI.2022.3150383

[15] Winchester, L.M., Harshfield, E.L., Shi, L., et al., 2023. Artificial intelligence for biomarker discovery in Alzheimer's disease and dementia. Alzheimer's & Dementia, 19(12), pp.5860–5871. Yang, M., Li, X., Gong, B., Chen, X. and Yuan, G., 2025. A multi-label deep learning model for detailed classification of Alzheimer's disease. Actas Españolas de Psiquiatría, 53(1), pp.89–99.

[16] World Health Organization (WHO), 2021. Global status report on the public health response to dementia. Geneva: World Health Organization.

[17] Yoon, J., Jordon, J. & van der Schaar, M., 2018. GAIN: Missing data imputation using generative adversarial nets. In: Proceedings of the 35th International Conference on Machine Learning (ICML). [online] Available at: https://proceedings.mlr.press/v80/yoon18a.html [Accessed 21 Jun. 2025].

[18] Zhang, D., Wang, Y., Zhou, L., Yuan, H. & Shen, D., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage, 55(3), pp.856–867.