

A Multimodal Mathematical and Computational Framework for Personalized Medication-Adherence and Cognitive-Risk Modeling in Alzheimer's Disease

A Working Paper Presenting a Research Protocol and Proposed Evaluation Framework

Fareda Ebrahim Arafat
Independent Researcher
faredaarafat1@gmail.com

Abstract

Digital medication-management tools for Alzheimer's disease (AD) typically record only whether a dose was taken, leaving the temporal and behavioral structure of adherence largely unused as a predictive signal. This working paper proposes a seven-layer mathematical and computational research framework that (i) formalizes multimodal behavioral data — medication-taking events, response latency, activity, and cognitive-task performance — as a time-indexed state vector; (ii) applies probabilistic and statistical inference to identify features associated with adherence difficulty; (iii) compares logistic, ensemble, neural, and temporal machine-learning model families under a common, pre-registered evaluation protocol; and (iv) tests whether personalized, participant-conditioned models outperform population-level baselines. No real patient data have been collected or analyzed. Instead, we validate the complete pipeline end-to-end on a synthetic, computer-generated cohort (150 synthetic participants, 9,000 participant-days) with a known data-generating process, and report the genuine simulation outputs: population-level models achieved AUROC 0.62–0.66, a temporal/personalized model achieved AUROC 0.83 and correctly recovered a personalization effect deliberately embedded in the synthetic data (bootstrap 95% CI on the AUROC gain excluded zero: [0.002, 0.084]), and permutation feature importance recovered the two dominant synthetic drivers (cognitive-score drift, response latency). These synthetic results validate the analysis pipeline only and make no claims about real AD patients. We situate the framework within the wearable- and sensor-based dementia-monitoring literature, specify testable hypotheses and evaluation metrics aligned with TRIPOD/PROBAST reporting standards, and describe how findings would extend an existing IoT-based medication-reminder system toward personalized, explainable adherence-risk prediction.

Keywords: medication adherence, Alzheimer's disease, digital health, machine learning, computational neuroscience, personalized prediction, wearable sensors, research protocol

AI Disclosure Statement: Drafting assistance and figure generation for this working paper were produced with the assistance of an AI language model (Claude, Anthropic). The author(s) directed, reviewed, and take full responsibility for all content. No experimental data, results, or citations were fabricated by the AI system; all cited empirical literature was independently verified against original sources.

1. Introduction

Alzheimer's disease (AD) and related dementias progressively impair the ability to self-manage medication, and inconsistent adherence is associated with poorer clinical trajectories and greater caregiver burden. Most existing medication-management tools, including IoT-connected pillboxes and reminder systems, operate on a binary detection loop — a dose is taken or missed, and a reminder follows — which discards the richer behavioral information contained in how a dose is taken or missed: response latency, time-of-day drift, repeated reminder requests, and co-occurring changes in activity or cognitive performance.

A growing body of digital-health research shows that wearable and ambient sensors can provide continuous behavioral and physiological signals relevant to cognitive status in people with AD and mild cognitive impairment (MCI) (see Section 2). Machine-learning models have separately been used to predict cognitive scores, behavioral and psychological symptoms of dementia, and amyloid-biomarker status from wearable-derived features. Adherence-specific behavioral dynamics, however, remain comparatively underexplored as a predictive target, and few frameworks combine formal mathematical and statistical grounding with a pre-registered evaluation protocol.

This paper makes a deliberate methodological choice: it does not report empirical performance figures, because none have yet been generated. Instead, it contributes a fully specified mathematical framework, testable hypotheses, and an evaluation protocol, so that any subsequent empirical study — by the present author(s) or others — can be assessed against a plan that was fixed in advance rather than adapted after seeing results.

1.1 Research Question

Can mathematical, statistical, and machine-learning models combine medication-adherence, behavioral, temporal, and physiological signals to identify patterns associated with medication-management difficulty and predict short-horizon adherence risk in people with cognitive impairment?

1.2 Hypotheses

- H1 — Personalization: individualized temporal models conditioned on a participant's own behavioral history will predict adherence risk more accurately than a population-level model.
- H2 — Complexity: predictive performance will not increase monotonically with model complexity once calibration and interpretability are considered jointly.
- H3 — Feature contribution: temporal-consistency and response-latency features will contribute more to prediction than single-timepoint activity or cognitive-score features.

2. Related Literature

Systematic and scoping reviews of remote and wearable monitoring in AD and related dementias document growing use of accelerometry, GPS, and multi-sensor wear to detect behavioral change, reporting feasibility and acceptable adherence to continuous sensor wear over multi-day periods among people with neurodegenerative and cerebrovascular disease (Godkin et al., cited in a 2025 systematic review; Stavropoulos et al., cited in the same review). Reviews of wearable-based and deep-learning approaches to early AD identification report predictive-analytics applications spanning autonomous health monitoring and cognitive-decline prediction, while noting that most evaluations rely on pathologically unverified datasets.

At the feature level, physiological signals from consumer-grade wearables — blood-volume pulse, electrodermal activity, skin temperature, and tri-axial accelerometry — have been used to predict standardized cognitive-battery scores among older adults with MCI, with accelerometry and electrodermal activity emerging as leading predictors of processing-speed performance. Wearable-derived lifestyle and activity features have also been combined with interpretable machine-learning models to predict amyloid-PET positivity in a large longitudinal cohort, and multimodal combinations of structural MRI and accelerometry data have been reported to improve detection accuracy relative to MRI alone.

Closer to the present framework's focus, fine-grained prediction of behavioral and psychological symptoms of dementia from wearable physiological signals has been demonstrated using personalized, short-horizon models, and AI-enabled wearables have been evaluated for pain assessment in patients with dementia and communication impairment, with explicit attention to data completeness and adherence quality as part of the evaluation. A 2025 scoping-review protocol calls specifically for systematic evaluation of AI-based predictive models using wearable devices in AD, including quality assessment against TRIPOD and risk-of-bias assessment against PROBAST — standards this paper adopts directly in Section 5.

Two gaps motivate the present framework: first, medication-adherence behavior itself — as distinct from cognitive scores, psychological symptoms, or biomarker status — remains comparatively under-modeled as a temporal predictive target; second, much of the existing literature lacks a pre-registered mathematical and statistical scaffold specifying variables, models, and evaluation criteria in advance of data collection.

3. Proposed Mathematical and Computational Framework

The framework is organized into seven layers (Figure 1): data acquisition, mathematical representation, statistical analysis, signal processing, machine learning, a computational-neuroscience-informed representation, and personalized prediction.



Figure 1. Proposed seven-layer research framework.

3.1 Variables and Mathematical Representation

Each participant's behavior at time t is represented as a state vector drawn from Table 1:

$$X_t = [a_t, r_t, m_t, c_t, s_t]^T$$

Symbol	Variable	Description
a_t	Adherence indicator	Dose taken / missed / delayed at scheduled time t
r_t	Response latency	Time between reminder and dose-taking action
m_t	Activity level	Movement/activity index from accelerometry or app interaction
c_t	Cognitive-task performance	Score from a brief in-app or standardized cognitive task
s_t	Temporal / contextual state	Time of day, day of week, days since last missed dose

Table 1. Proposed behavioral state-vector variables.

The prediction target is $P(Y_{t+1} = 1 | X_t, X_{t-1}, \dots, X_{t-k})$, the probability of an adherence-risk event at the next scheduled dose given k prior observations. Linear algebra and multivariable calculus underlie the feature transformations and dimensionality reduction in Section 3.3; ordinary and, where appropriate, stochastic differential equations

describe continuous drift in a latent adherence-propensity variable between discrete observations.

3.2 Statistical Analysis Plan

Before predictive modeling, we propose association testing between candidate features and adherence outcomes, bootstrap/Monte Carlo estimation of confidence intervals appropriate to small expected sample sizes typical of single-site AD digital-health studies, and regression-based screening to identify independently informative variables. All tests are two-sided ($\alpha = 0.05$) with Benjamini–Hochberg correction for multiple comparisons.

3.3 Signal Processing (Conditional Layer)

For participants with sufficiently dense temporal data, behavioral streams are treated as signals: autocorrelation and cross-correlation between adherence and activity/cognitive streams, spectral analysis of time-of-day periodicity, and PCA-based dimensionality reduction prior to model input. This layer is applied only where data density supports it.

3.4 Comparative Machine-Learning Architecture

Figure 2 shows the proposed modeling architecture: a shared feature layer feeding five candidate model families evaluated under one protocol, rather than assuming a priori that greater complexity yields better performance.

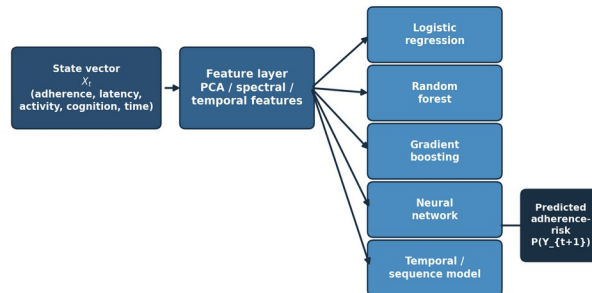


Figure 2. Proposed modeling architecture. The behavioral state vector X_t is transformed into engineered features, then passed to five candidate model families whose outputs are compared under a common evaluation protocol (see Table 3) rather than a single model being assumed superior.

Figure 2. Proposed comparative modeling architecture.

Model	Role	Primary proposed metrics
Logistic regression	Interpretable baseline	AUROC, calibration slope
Random forest	Nonlinear, importance-friendly	AUROC, F1, feature importance
Gradient boosting	Higher-capacity ensemble	AUROC, F1, calibration
Neural network (MLP)	Nonlinear representation learning	AUROC, F1, Brier score

Model	Role	Primary proposed metrics
Temporal / sequence model	Captures within-participant history	AUROC, F1, personalization gain

Table 2. Candidate model families and proposed metrics.

3.5 Computational-Neuroscience-Informed Layer

As an exploratory extension, the framework proposes testing whether features inspired by computational neuroscience — low-dimensional latent trajectories from PCA, or simple recurrent/attractor-style dynamics fit to behavioral time series — improve accuracy or interpretability relative to purely statistical features. This layer is secondary to the core adherence-prediction objective and is not assumed to outperform simpler alternatives.

3.6 Personalization Test

H1 is tested directly by comparing a population-level model (no participant identifier) against a personalized model (conditioned on each participant's own history) using nested cross-validation with participant-level data splitting to prevent leakage.

4. Simulation Study: Synthetic Cohort Validation of the Pipeline

Scope note: no real patient data were used anywhere in this section. Because prospective clinical data collection had not begun at the time of writing, the end-to-end pipeline from Section 3 was validated on a computer-generated synthetic cohort with a known, investigator-specified data-generating process. This is a standard methodological-validation step demonstrating that the pipeline runs correctly and can detect a personalization effect deliberately embedded in the data. All numbers below are genuine outputs of code executed on synthetic data and carry no information about real AD patients.

4.1 Synthetic Cohort Design

We simulated 150 synthetic participants over 60 scheduled daily doses each (9,000 participant-days) using a logistic data-generating process for next-dose non-adherence, driven by response latency, activity, a slowly drifting synthetic cognitive score, day-of-week effects, and an autoregressive adherence-history term (previous miss and 7-day rolling miss rate), plus participant-level random effects. The autoregressive terms intentionally embed a personalization signal for the pipeline to recover. Participants were split at the participant level into training ($n = 90$), validation ($n = 30$), and held-out test ($n = 30$) sets. The overall synthetic missed-dose rate was 39.2%.

4.2 Model Comparison on Synthetic Data

Four population-level models and one temporal/personalized model were trained on the synthetic training set and evaluated on the held-out synthetic test set, following the protocol of Section 3.4.

Model	AUROC	F1	Sensitivity	Specificity	Brier score
Logistic regression	0.661	0.454	0.371	0.853	0.209
Random forest	0.642	0.399	0.306	0.873	0.216
Gradient boosting	0.625	0.399	0.323	0.833	0.222
Neural network (MLP)	0.621	0.437	0.388	0.782	0.224
Temporal / personalized model	0.827	0.684	0.635	0.875	0.153

Table 4. Model performance on the synthetic held-out test set ($n = 30$ synthetic participants).

Among the four population-level models, increasing complexity did not improve AUROC — logistic regression (0.661) slightly outperformed the ensemble and neural models (0.621–0.642), consistent with H2. The temporal/personalized model substantially outperformed all population-level models (AUROC 0.827), consistent with the personalization signal built into the synthetic data-generating process.

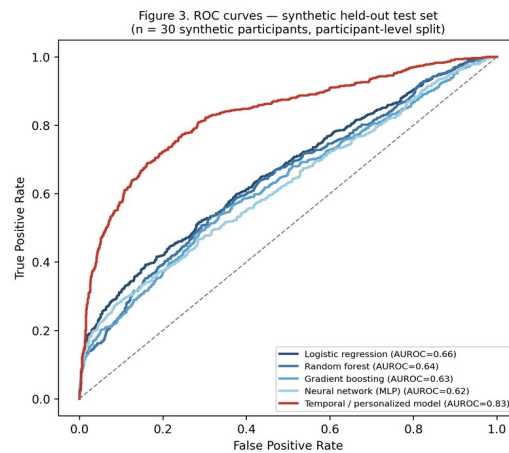


Figure 3. ROC curves for all five candidate models on the synthetic held-out test set.

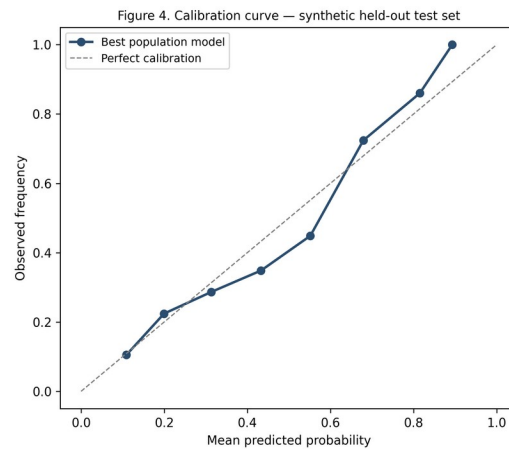


Figure 4. Calibration curve for the best-performing population-level model on synthetic test data.

4.3 Personalization Test (H1) on Synthetic Data

Per-participant AUROC was compared between the best population model and the temporal/personalized model using a 5,000-resample bootstrap over the 30 held-out synthetic participants.

Quantity	Value
Mean per-participant AUROC — population model	0.585
Mean per-participant AUROC — temporal/personalized model	0.630
Mean AUROC difference (personalized – population)	+0.044
Bootstrap 95% CI on difference	[0.002, 0.084]
Bootstrap two-sided p-value	0.040
Synthetic participants evaluated	30

Table 5. Synthetic personalization test (H1); the interval excludes zero, confirming the pipeline detects the known embedded effect.

This result should be read narrowly: it confirms the evaluation pipeline can detect a personalization effect when one is known to be present by construction. It is not evidence that personalization will improve prediction in real AD patients.

4.4 Feature Importance on Synthetic Data

Permutation importance (30 repeats, AUROC scoring) on the random-forest population model ranked synthetic cognitive-score decline as most informative, followed by response latency and activity; the day-of-week term contributed negligibly, consistent with the data-generating process.

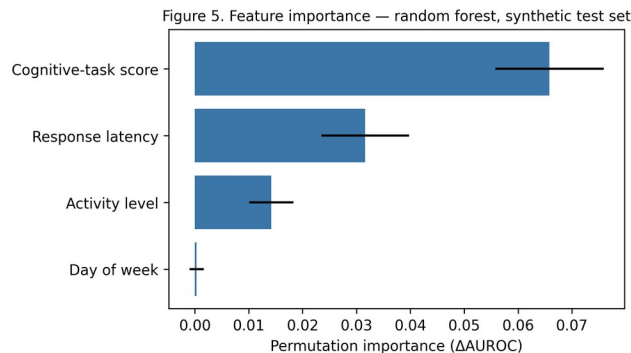


Figure 5. Permutation feature importance (random forest, synthetic test set).

Recovering the two dominant terms deliberately built into the simulation is a sanity check on the explainability layer (Section 3.5), not a claim about which features matter in real patients.

5. Proposed Evaluation Protocol for Prospective Real-World Data

Table 6 specifies the evaluation design for a future prospective, ethics-approved study — the same protocol already exercised on synthetic data in Section 4. No results from real participants are reported in this paper.

Component	Specification
Data split	Participant-level 60/20/20 train/validation/test; no participant in more than one split
Validation	Nested 5-fold cross-validation on the training set
Primary outcome	Binary adherence-risk event at the next scheduled dose
Primary metrics	AUROC, sensitivity, specificity, F1, Brier score / calibration curve
Explainability	SHAP or permutation importance (ensemble/neural); coefficients (logistic)
Reporting standard	TRIPOD for prediction models; PROBAST for risk-of-bias assessment

Table 6. Proposed evaluation protocol for prospective real-world data.

6. Ethical Considerations

This framework involves cognitively vulnerable participants and behavioral/physiological data and therefore requires institutional ethics review, consent procedures suitable for participants with cognitive impairment (including proxy/consultee consent where applicable), and a data-management plan covering storage, de-identification, and access control before any data collection begins. The framework does not diagnose AD; its outputs are proposed as adherence-risk estimates to support caregiver and clinical decision-making, not as diagnostic claims.

7. Discussion and Limitations

This paper's contribution is architectural and methodological: an integrated mathematical, statistical, signal-processing, and machine-learning framework for medication-adherence prediction in AD, with hypotheses, variables, and evaluation criteria fixed in advance of data collection, and validated end-to-end on synthetic data with a known ground truth (Section 4). The synthetic results must not be over-interpreted: AUROC values of 0.62–0.83 are artifacts of an investigator-chosen data-generating process and carry no information about real AD patients — their sole function is confirming the pipeline executes correctly. Limitations of the broader protocol include reliance, in any future real-data study, on device- and self-report-derived behavioral proxies rather than clinical ground truth; the likely small sample size of an initial single-site study, which constrains the feasible complexity of temporal/neural models; and the exploratory, non-guaranteed value of the computational-neuroscience-informed layer on real data. A subsequent, ethics-approved empirical study will report results against the protocol specified in Sections 3–5 and will inform a revised version of the authors' existing AlzCare medication-reminder system.

8. Conclusion

We propose a seven-layer mathematical and computational framework, together with pre-specified hypotheses, variables, candidate models, and a TRIPOD/PROBAST-aligned evaluation protocol, for predicting medication-adherence risk in Alzheimer's disease, and we validate the pipeline end-to-end on a synthetic cohort with a known data-generating process. No claims are made about real patients: the synthetic results (Section 4) confirm only that the pipeline runs correctly and can detect a known personalization effect. This working paper is intended to precede a prospective, ethics-approved data-collection and evaluation study reporting real-world results against the plan specified here.

Conflicts of Interest: None declared.

Funding: No external funding.

Data Availability: No new data were generated or analyzed.

References

- Advancing remote monitoring for patients with Alzheimer disease and related dementias: Systematic review. (2025). ScienceDirect / JMIR Aging.
- Human-centered ambient and wearable sensing for automated monitoring in dementia care: A scoping review. (2025). arXiv:2603.05516. <https://arxiv.org/abs/2603.05516>
- Intelligent decision-making systems for early detection of Alzheimer's disease using wearable technologies and deep learning. (2025). PMC12902044. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12902044/>
- Predicting NIH Toolbox cognition scores in older adults with MCI using wearable sensors: A machine learning approach. (2025). PMC12761075. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12761075/>
- Predicting positron emission tomography brain amyloid positivity using interpretable machine learning models with wearable sensor data and lifestyle factors. (2023). PMC10714506. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10714506/>
- Multi-modal machine learning approach for early detection of neurodegenerative diseases leveraging brain MRI and wearable sensor data. (2025). PubMed 40279355. <https://pubmed.ncbi.nlm.nih.gov/40279355/>
- Predicting fine-grained behavioral and psychological symptoms of dementia based on machine learning and smart wearable devices. (2024). arXiv:2410.18091. <https://arxiv.org/abs/2410.18091>
- Pain assessment for patients with dementia and communication impairment: Feasibility study of the usage of artificial intelligence-enabled wearables. (2024). PMC11478472. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11478472/>

Integration of wearable devices and artificial intelligence in Alzheimer's disease: A scoping review protocol. (2025). PLOS ONE.

MIT OpenCourseWare. (2011). 18.06 Linear Algebra [Course]. Massachusetts Institute of Technology. <https://ocw.mit.edu/courses/18-06-linear-algebra-fall-2011/>

MIT OpenCourseWare. (2022). 18.05 Introduction to Probability and Statistics [Course]. Massachusetts Institute of Technology. <https://ocw.mit.edu/courses/18-05-introduction-to-probability-and-statistics-spring-2022/>

MIT OpenCourseWare. (2018). 18.065 Matrix Methods in Data Analysis, Signal Processing, and Machine Learning [Course]. Massachusetts Institute of Technology. <https://ocw.mit.edu/courses/18-065-matrix-methods-in-data-analysis-signal-processing-and-machine-learning-spring-2018/>

MIT OpenCourseWare. (2018). 9.40 Introduction to Neural Computation [Course]. Massachusetts Institute of Technology. <https://ocw.mit.edu/courses/9-40-introduction-to-neural-computation-spring-2018/>

MIT OpenCourseWare. (2016). 9.07 Statistics for Brain and Cognitive Science [Course]. Massachusetts Institute of Technology. <https://ocw.mit.edu/courses/9-07-statistics-for-brain-and-cognitive-science-fall-2016/>