VISTA: Velocity-Informed Sequential Transition Analysis

Prajvi Saxena[1], Lina Charlotte Jeran[2], Sabine Janzen[1], Iris Blotenberg[2], Jochen René Thyrian[2], Wolfgang Maass[1]
[1]German Research Center for Artificial Intelligence (DFKI), Germany
Emails: prajvi0123@gmail.com, sabine.janzen@dfki.de, wolfgang.maass@dfki.de
[2]German Center for Neurodegenerative Diseases (DZNE), Germany
Emails: linacharlotte.jeran@dzne.de, Iris.Blotenberg@dzne.de, Rene.Thyrian@dzne.de

***Abstract*—Accurate prediction of dementia progression is crucial for timely intervention and care planning, yet remains challenging due to highly variable patient trajectories. Recent machine learning approaches rely on neuroimaging and genetic biomarkers. While effective, these methods are costly, invasive, and difficult to deploy at scale in primary care settings. In contrast, routinely collected longitudinal clinical data, such as questionnaire based assessments and care intervention records remain underutilized for prognosis. We propose VISTA (Velocity-Informed Sequential Transition Analysis), a framework for dementia prognosis using real-world clinical data. It explicitly encodes temporal velocity - the rate of change in clinical features between consecutive assessments together with the absolute clinical state. It addresses the limitation of existing methods that treat assessments as independent snapshots. Our framework converts longitudinal prediction as transition-level forecasting via sliding-window decomposition, which increases the number of training instances while preserving patient-specific temporal structure. We implement it using gradient boosted decision trees to capture non-linear state-velocity interactions while handling missing data natively. We evaluate the framework on DelpHi dataset an 8-year real-world primary care longitudinal dataset comprising of 459 patients (768 annual transitions). VISTA achieves $R^2 = 0.74$ and MAE = 2.63 MMSE points, substantially outperforming linear mixed models ($R^2 = 0.25$) and standard baselines in cognitive prognosis. Ablation studies and SHAP analysis also revealed several crucial patterns. Qualitative analysis shows VISTA accurately captures systematic decline patterns while still maintaining predictions on volatile trajectories. These results demonstrate that velocity-aware modeling enables accurate dementia prognosis.**

I. Introduction

Predicting the course of cognitive decline in individuals with dementia, commonly referred to as dementia prognosis, is a critical challenge for both clinical care and computational modelling [1], [2]. Accurate forecasts of future cognitive function, such as scores on the Mini-Mental State Examination (MMSE), can guide early interventions, care planning, and treatment decisions tailored to individual patients [30].

This task is complicated by the highly heterogeneous nature of dementia progression. Patients with similar MMSE scores at a given time may follow very different cognitive trajectories in the months or years that follow [2]. This challenge is further reinforced in primary care, i.e., in non-specialist clinical settings, where dementia is often first detected. In these environments, assessments are performed intermittently, clinical records are incomplete, and data are sparsely sampled [3], [11], [15]. These characteristics mean that many traditional modelling approaches remain ineffective.

Previous work has attempted to address dementia prognosis using several modelling paradigms. Longitudinal statistical methods such as linear mixed models (LMMs) are standard in clinical research, but they typically assume smooth population-level trends and struggle to capture individual-level variability, particularly in the presence of non-linear progression patterns [5]. In parallel, deep learning models trained on neuroimaging or cerebrospinal fluid (CSF) biomarkers, e.g., positron emission tomography (PET) or amyloid/tau imaging, have demonstrated strong predictive performance [6]–[8]. However, these approaches depend on costly and invasive diagnostics that are impractical for widespread use in primary care settings. More recently, machine learning models have been applied to electronic health records and structured survey data [9], [10], offering a more scalable alternative. Yet many such models are limited by a stationary bias [12], [13]; they treat each patient record as an independent snapshot and overlook the temporal dynamics that characterize disease progression. This is a critical limitation, as the rate of change in clinical indicators, referred to as *velocity*, has been shown to carry important prognostic value in dementia and aging [3], [4].

In this work, we investigate whether explicitly modelling clinical velocity can improve dementia prognosis from real-world primary care data. We introduce *VISTA (Velocity-Informed Sequential Transition Analysis)*, a machine learning framework that models patient trajectories as sequences of clinical transitions, capturing both the absolute clinical state and its temporal velocity defined as the per-year rate of change in cognitive, physical, and medical indicators. By decomposing patient histories into transition-level segments, VISTA reframes dementia prognosis as a sequence modelling task and increases the effective number of training examples in sparse, real-world datasets. VISTA employs gradient-boosted decision trees (GBDT) [25]–[27] to learn non-linear relationships between state and velocity features while handling missing data natively; a critical requirement given the high missingness in primary care records [14], [15]. In addition, we incorporate SHAP [17], [18] to provide interpretable, patient-specific predictions, enabling clinicians to understand whether

the model's forecast is driven more by a patient's current condition or by recent trends of decline.

We evaluate VISTA on the DelpHi dataset [28], [29], an eight-year longitudinal cohort comprising 459 patients aged 70 and older [5]. Using a sliding-window decomposition, we extract 768 annual transitions and predict future MMSE scores. Our model achieves R^2 of 0.74 and a mean absolute error (MAE) of 2.63 MMSE points, substantially outperforming baseline models including LMM, Gaussian Process Regression (GPR), and linear regression.

The contributions of this paper are as follows:

- We propose a transition-level learning framework that improves modelling resolution and sample efficiency.
- We introduce a velocity-aware representation for dementia prognosis from routine clinical data.
- We implement a GBDT-based model that captures nonlinear progression patterns and handles missing data without imputation.
- We integrate SHAP-based interpretability to support clinical understanding and decision-making.

II. Velocity-Informed Sequential Transition Analysis (VISTA)

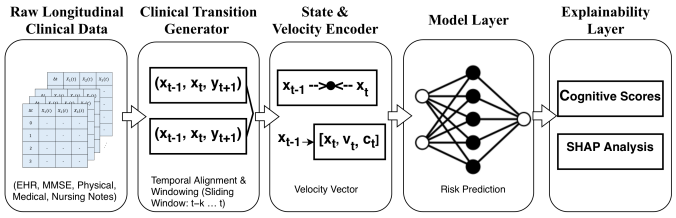

Fig. 1. Overview of the VISTA framework. Longitudinal primary-care data are transformed into sliding-window clinical transitions and velocity features, which are jointly modeled by the model to predict future cognitive outcomes, with integrated explainability for patient-specific prognosis.

The VISTA framework composed of four primary modules that transform raw longitudinal clinical records to dementia prognosis forecasts: (1) the Clinical Transition Generator, which transforms sparse longitudinal patient records into state transitions using a sliding-window approach; (2) the State-Velocity Encoder, which jointly represents the absolute clinical state and its temporal velocity across cognitive, physical, and medical data; (3) the Model Layer - which models the transitions to predict future cognitive scores (Forecasting next year cognitive score prediction); and (4) the Explainability Layer, which provides patient-specific interpretability via feature attribution (SHAP) to support clinical decision-making. An architectural overview is provided in Figure 1.

A. Clinical Transition Generator

A fundamental challenge in dementia prognosis using primary care data is the irregular, sparse, and patient-specific sampling of longitudinal records. Let $\mathcal{D} = \{P_1, P_2, \ldots, P_N\}$ denote a cohort of N patients. For each patient P_i, we define their longitudinal clinical history as an ordered sequence of observations

$$\mathcal{H}_i = \{(\mathbf{x}_i^t, w_i^t)\}_{t=1}^{T_i} \tag{1}$$

where $\mathbf{x}_i^t \in \mathbb{R}^D$ represents a D-dimensional clinical feature vector at the t-th observed visit, and $w_i^t \in \{0, 1, \ldots, W\}$ denotes the corresponding study wave (the absolute temporal index of the assessment, e.g., baseline, year-1, year-2). The observation index t captures the ordering of visits for an individual patient, while the study wave w_i^t represents the actual time of assessment, allowing the model to account for irregular follow-up intervals.

To model temporal dynamics of cognitive decline, the clinical transition generator frames longitudinal prediction as a transition-level learning problem via a sliding-window decomposition of $\mathcal{H}_i$. For each valid index $t \in \{2, \ldots, T_i - 1\}$, the generator constructs a clinical transition $\mathcal{T}_{i,t}$ defined as

$$\mathcal{T}_{i,t} = \langle (\mathbf{x}_i^{t-1}, w_i^{t-1}), (\mathbf{x}_i^t, w_i^t), y_i^{t+1} \rangle \tag{2}$$

where $y_i^{t+1} \in \{0, \ldots, 30\}$ denotes the cognitive function score (MMSE) observed at the subsequent follow-up wave w_i^{t+1}. This transforms patient trajectories into overlapping transitions, expanding the learning set from N patients to a pooled transition set

$$\mathcal{P} = \bigcup_{i=1}^{N} \bigcup_{t=2}^{T_i - 1} \mathcal{T}_{i,t}. \tag{3}$$

This transition-level formulation increases training examples while preserving temporal dependencies within individual patient histories, enabling the model to learn progression patterns including inflection points where decline accelerates.

B. State–Velocity Encoder

The State–Velocity Encoder transforms clinical transitions into a feature representation that captures both current impairment and trajectory. Traditional longitudinal models treat each observation as an independent snapshot, losing information about whether a patient's condition is stable, improving, or declining [12], [13]. VISTA addresses this by explicitly encoding velocity alongside state.

Given a clinical transition $\mathcal{T}_{i,t}$, the encoder computes the velocity vector $\mathbf{v}_i^t$ as the yearly rate of change:

$$\mathbf{v}_i^t = \frac{\mathbf{x}_i^t - \mathbf{x}_i^{t-1}}{w_i^t - w_i^{t-1}} \tag{4}$$

where $w_i^t - w_i^{t-1}$ is the time interval in years between consecutive assessments. This normalization is critical for handling irregular follow-up: a 3-point MMSE decline over 6 months ($v = -6$ points/year) carries different clinical meaning than the same decline over 2 years ($v = -1.5$ points/year). Each element $v_{i,d}^t$ quantifies the yearly rate of change for a specific feature (cognitive score, physical function, mood).

The velocity vector enables the model to distinguish between clinically distinct scenarios that appear identical in data. Consider two patients both scoring MMSE 24: one stable over the past year ($\|\mathbf{v}_i^t\| \approx 0$), another declining from 28

($\|\mathbf{v}_i^t\| \gg 0$). Static models treat these identically. However, velocity-aware modeling captures the declining patient as higher risk.

The encoder constructs the final representation $\mathbf{z}_i^t$ by concatenating state, velocity, and static characteristics:

$$\mathbf{z}_i^t = [\mathbf{x}_i^t \oplus \mathbf{v}_i^t \oplus \mathbf{c}_i] \tag{5}$$

where $\mathbf{x}_i^t$ is the current state, $\mathbf{v}_i^t$ is velocity, and $\mathbf{c}_i$ contains time-invariant features (age, gender, education). The prediction y_i^{t+1} thus depends on both current clinical state and velocity. This enables the model to capture whether the decline is accelerating, stabilizing, or decelerating that cannot be detected from absolute scores alone.

C. Model Layer

The State-Velocity Encoder produces representations $\mathbf{z}_i^t$ that combine current state and rate of change (velocity). The Model Layer learns how these features predict future outcomes. The key challenge is that velocity features have different predictive effects depending on absolute state. Rapid physical decline indicates high risk in cognitively stable patients but may be less informative in already severe cases. This requires modeling non-linear interactions between state and velocity components.

We implement the transition function $\Phi : \mathcal{Z} \to \mathcal{Y}$ as a gradient boosted decision tree (GBDT) ensemble, which maps $\mathbf{z}_i^t$ to predicted outcomes:

$$\hat{y}_i^{t+1} = \Phi(\mathbf{z}_i^t). \tag{6}$$

The function is trained to minimize the mean absolute error (MAE) over the pooled transition set $\mathcal{P}$:

$$\mathcal{L}_{\text{MAE}} = \frac{1}{|\mathcal{P}|} \sum_{\mathcal{T}_{i,t} \in \mathcal{P}} \left| y_i^{t+1} - \hat{y}_i^{t+1} \right|. \tag{7}$$

GBDT is well-suited for this task for two reasons. First, tree-based models naturally capture feature interactions through hierarchical splits, enabling the model to learn that velocity features have different impacts depending on absolute state values [16]. Second, GBDT handles missing values natively by learning optimal split directions for observed versus missing features at each node. This preserves the informative signal of missingness (e.g., inability to complete assessments) without requiring imputation [14], [15].

The ensemble constructs the final predictions as a sum of K sequential trees:

$$\hat{y}_i^{t+1} = \sum_{k=1}^{K} f_k(\mathbf{z}_i^t) \tag{8}$$

where each tree f_k is trained to fit the residual from previous trees. By combining the current clinical state $\mathbf{x}_i^t$ and velocity $\mathbf{v}_i^t$ within this ensemble architecture, the model generates 12-month cognitive forecasts (MMSE scores).

D. Explainability Layer

In clinical settings, predictive accuracy alone is insufficient, clinicians require transparent, patient-specific explanations to support decision-making and build trust in model recommendations. The Explainability Layer provides interpretable feature attributions for individual patient predictions. We employ SHAP (SHapley Additive exPlanations) [17], [18], a method that decompose each prediction into feature-level contributions. SHAP is well-suited for our framework because it provides consistent, patient-level feature attributions for non-linear tree-based models.

For a given clinical transition $\mathcal{T}_{i,t}$, the predicted cognitive score $\hat{y}_i^{t+1}$ is decomposed as:

$$\hat{y}_i^{t+1} = \phi_0 + \sum_{j=1}^{M} \phi_{i,j}^t \tag{9}$$

where ϕ_0 is the expected prediction over all transitions in $\mathcal{P}$, and $\phi_{i,j}^t$ is the shapley value quantifying the j-th feature's contribution to this prediction. Positive values ($\phi_{i,j}^t > 0$) indicate features that increase predicted scores (e.g., stable physical health), while negative values indicate features that decrease predictions (e.g., rapid cognitive decline velocity).

This layer is valuable for distinguishing static risk factors from dynamic trajectories. By separating attributions for current state $\mathbf{x}_i^t$ and velocity $\mathbf{v}_i^t$ components, clinicians can identify whether predicted decline is driven by current impairment level or rate of change, enabling targeted interventions.

III. Experiments

A. Dataset

We evaluate VISTA on the DelpHi dataset, an 8-year longitudinal primary care cohort comprising $N = 459$ older adults (age ≥ 70). Patients were assessed at baseline and up to eight follow-up waves ($w \in \{0, \ldots, 8\}$), with irregular and patient-specific intervals between visits. Participant retention decreases over time, with 411 patients completing baseline assessment, 359 at year 1, 131 at year 4, and 41 at year 8. This sparsity and dropouts poses significant challenges for longitudinal modeling. The detailed dataset descriptive analysis are reported in [28], [29].

Using the sliding-window transition generation described in Section II-A, we extract 768 clinical transitions from the original patient trajectories. Each clinical state aggregates routinely collected primary care variables such as cognitive status, physical and mental function, medical burden, clinical interventions, and demographic information. Time-invariant demographic variables are incorporated as contextual features, while velocity features are derived from time-varying variables as described in Section II-B.

The dataset exhibits substantial missingness, exceeding 30% for core cognitive measures and 45% for socio-demographic variables. Rather than performing imputation, we preserve this sparsity and rely on the native missing-value handling of tree-based models, allowing missingness itself to serve as an informative clinical signal.

B. Baseline Methods

To evaluate the predictive performance of VISTA, we compare against statistical and machine learning approaches commonly used in dementia prognosis.

- **Linear Mixed Model (LMM)** [5], [20]: A standard longitudinal statistical model widely used in clinical studies, accounting for population-level effects while modeling inter-patient variability.
- **Last Observation Carried Forward (LOCF)** [21]: A longitudinal baseline that predicts future cognitive status using the most recent observed MMSE score.
- **Gaussian Process Regression (GPR)** [22], [23]: A non-parametric bayesian regression model that captures smooth, non-linear disease trajectories over time.
- **Linear Regression (OLS)** [24]: A linear baseline modeling the relationship between static clinical snapshots and future cognitive outcomes.

To isolate the contribution of individual components, we additionally report ablation results for VISTA without the State–Velocity Encoder (No Velocity), as well as standard Gradient Boosted Decision Tree implementations, including XGBoost [26] and LightGBM [25].

The comparative performance of all methods is summarized in Table I.

C. Experimental Settings

We formulate the dementia prognosis prediction task as forecasting MMSE score at wave $t+1$ given clinical state at wave t. In the DelpHi dataset, consecutive waves are separated by approximately 12 months, making this an yearly prognosis task.

Train-Test Split: The transition pool $\mathcal{P}$ is partitioned into training (80%; 617 transitions) and test (20%; 151 transitions) sets using patient-level stratification. All transitions from a given patient are assigned to either training or test, preventing data leakage where the model could learn patient-specific patterns [19].

Model Configuration: We implement VISTA using XGBoost [27] with 1,000 trees, maximum depth 6, learning rate 0.1, and MAE loss. Hyperparameters were selected via 5-fold cross-validation on the training set.

Evaluation Metrics. Performance is measured using:

- R^2: Proportion of variance in future MMSE scores explained by the model
- MAE: Mean absolute prediction error in MMSE points (scale 0-30)

IV. Results

Table I presents performance of VISTA and baselines on the held-out test set. VISTA achieves $R^2 = 0.74$ and MAE = 2.63 MMSE points, substantially outperforming all baselines.

Representation Gap: The performance difference between Linear Mixed Models (LMM, $R^2 = 0.25$) and VISTA ($R^2 = 0.74$) shows that temporal representation, not model complexity, is the primary challenge in dementia prognosis.

TABLE I
PERFORMANCE COMPARISON OF VISTA AND BASELINE MODELS ON THE DELPHI DATASET, EVALUATED USING THE COEFFICIENT OF DETERMINATION ($\mathbf{R^2}$) AND MEAN ABSOLUTE ERROR (MAE).

Baselines	**Performance Metrics**		
	Variant	$\mathbf{R^2} \uparrow$	**MAE** $\downarrow$
Statistical	Linear Mixed Model (LMM)	0.25	5.35
Statistical	Last Obs. Carried Forward (LOCF)	0.64	3.16
Statistical	Gaussian Process Regression (GPR)	0.697	2.91
Statistical	Linear Regression (OLS)	0.698	2.91
VISTA (Abl)	VISTA (No Velocity)	0.68	3.01
VISTA (Abl)	VISTA-XGBoost	0.70	2.89
VISTA (Abl)	VISTA-LightGBM	0.71	2.86
VISTA (Pro)	**VISTA-Ensemble (GBDT)**	**0.74**	**2.63**

LMM assumes smooth, population-averaged trajectories and cannot capture individual-level acceleration patterns. The improvement shows that explicitly encoding velocity is critical for individual-level prediction. When the same velocity-augmented features are provided to linear regression (OLS), performance jumps to $R^2 = 0.698$, approaching VISTA. This indicates velocity encoding linearizes the progression signal. VISTA's gradient boosting adds 4.2 points ($R^2 = 0.74$) by capturing remaining non-linear state-velocity interactions.

Ablation Study: Removing velocity features reduces performance to $R^2 = 0.68$ (MAE = 3.01), only 4 points above the naive LOCF baseline ($R^2 = 0.64$). The 6-point drop in R^2 and 14% increase in MAE demonstrate velocity encoding is crucial for VISTA's performance. Without rate-of-change information, even sophisticated models cannot distinguish stable impairment from accelerating decline. Among tree-based variants, performance is consistent: LightGBM ($R^2 = 0.71$), XGBoost ($R^2 = 0.70$), and VISTA-Ensemble ($R^2 = 0.74$). This robustness across implementations confirms that gains come from the velocity-augmented feature space rather than specific algorithmic choices.

Feature Importance Analysis: To assess model interpretability, we analyze feature attributions using SHAP on the held-out test set. The baseline MMSE is the dominant feature for future cognitive score prediction. Importantly, MMSE velocity ranks as the second most influential feature, demonstrating that the recent rate of cognitive change provides predictive information beyond absolute score alone. Negative velocity (rapid decline) produces strong negative SHAP values, while stable velocity contributes neutrally or positively, validating that the model distinguishes accelerating trajectories from stable impairment.

We also observed physical function (PCS velocity) and depressive symptoms (GDS velocity) contribute more to the prediction than their corresponding static measures (PCS and GDS). This suggests that recent decline in physical or mental health acts as an early warning signal for future cognitive decline. Intervention related variables, including antidementia medication use and caregiver support hours also has crucial influence on model predictions. In contrast, social variables such as living arrangement and social support exhibit less

importance, could be due to short-term prognosis of next year prediction. SHAP distributions reveal non-uniform patterns: MMSE velocity impact is strongest at moderate impairment (MMSE 15-25) and weaker at extreme values. This context-dependent relationship justifies the GBDT architecture, as linear models cannot capture such interactions.

Trajectory Analysis: Figure 2 shows VISTA predictions for four test-set patients exhibiting different dementia trajectory patterns. Patient 31 demonstrates rapid progressive decline, dropping from MMSE 22 to 2 over 8 years with a sharp acceleration after year 4. The model accurately tracks this trajectory, successfully predicting the inflection point when decline accelerates. At year 3, when the patient still scores MMSE 24, the model detects negative velocity in physical and cognitive features and correctly forecasts the upcoming decline. This demonstrates the ability to identify high-risk trajectories before substantial decline occurs.

Patient 18 shows steady decline from MMSE 23 to 16, with predictions remaining within 2-3 points throughout most of the trajectory. However, the model underestimates a sudden drop at year 6 (actual: 16, predicted: 20) and does not anticipate the subsequent recovery to 23 at year 8. This indicates the model is more accurate for gradual, continuous decline than for abrupt reversals or recoveries. Patient 14 exhibits fluctuation between MMSE 15-24 while maintaining overall late-stage stability. Model correctly identifies the lack of progressive decline, with predictions smoothing over short-term volatility. The model prioritizes long-term trends over short-term fluctuations, which is clinically appropriate since sustained decline requires intervention while temporary variations do not.

Patient 24 shows the most challenging pattern: 5-7 point fluctuations between waves while maintaining MMSE 27. Predictions remain stable around 27, correctly identifying overall stability but missing high-frequency variation. Examination of this patient's features reveals no corresponding volatility in physical health or mood that would explain MMSE fluctuations, suggesting measurement noise or unmeasured confounders. The model appropriately maintains stable predictions rather than fitting noise.

Overall, VISTA achieves high accuracy for systematic trajectory patterns (smooth linear decline, rapid acceleration, late-stage stability, and high-frequency fluctuation) while showing reduced performance on volatile or irregular cases. Importantly, even in challenging cases, prediction errors remain bounded within clinically acceptable ranges, and the model produces no catastrophic mispredictions that would lead to inappropriate clinical decisions.

V. Conclusion

We presented VISTA, a framework for dementia prognosis from routinely collected primary care data. VISTA addresses a fundamental limitation of existing approaches by explicitly encoding temporal velocity—the rate of change in clinical features alongside absolute state. This enables distinguishing stable impairment from accelerating decline, a distinction critical for early interventions. Evaluated on 459 patients

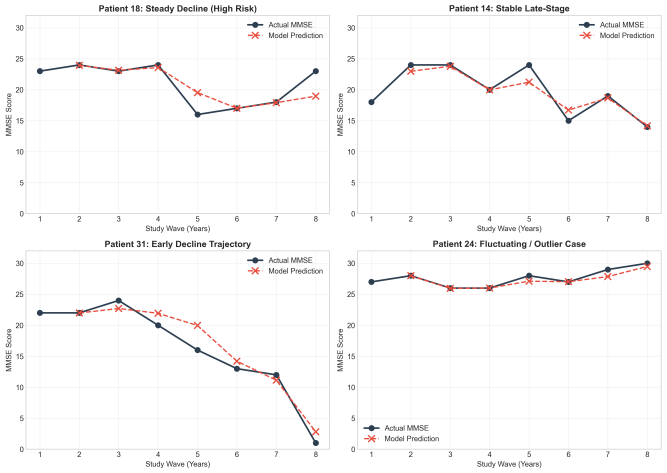

Fig. 2. Predicted versus actual Mini-mental state examination (MMSE) trajectories for four representative test patients. VISTA accurately tracks systematic progression patterns (Patients 18, 31, 14) including steady decline, rapid acceleration, and stable fluctuation. Patient 24 exhibits high volatility that the model conservatively smooths rather than fitting

over 8 years, VISTA achieves $R^2 = 0.74$ and MAE = 2.63 MMSE points, substantially outperforming standard baselines. Ablation studies confirm velocity features drive this improvement: removing them reduces performance to baseline levels. It also provides transparent explanations through SHAP analysis, revealing that recent changes in cognitive, physical, and mental health measures play a critical role in short-term prognosis. The framework shows that accurate dementia prognosis is achievable using low-cost clinical assessments without expensive neuroimaging or genetic biomarkers. This makes scalable deployment feasible in primary care settings where specialized diagnostic resources are limited.

VI. Limitation

VISTA is designed for short-term prognosis i.e. predicting cognitive scores at the next follow-up visit, and does not directly model long-horizon disease progression. In addition, evaluation is conducted on a single primary-care cohort, and further validation is required to assess generalizability across populations and care settings. The reliance on observed inter-visit changes to compute velocity features may limit performance for patients with extremely sparse or irregular follow-up. Finally, while SHAP provides transparent feature attributions, these explanations are predictive rather than causal and are intended to support clinical decision-making, not to replace clinical judgment.

VII. Ethics Statement & Reproducibility

The DelpHi dataset used in this study was collected under ethical approval from the Ethics Committee of the Chamber of Physicians of Mecklenburg-Western Pomerania (registry number: BB 20/11). Also, the DelpHi trial was registered under clinicaltrials.gov (Identifier: NCT01401582). All participants provided written informed consent prior to study enrollment.

Data were de-identified and processed in accordance with applicable privacy regulations.

Implementation: VISTA is implemented using Python 3.9 with XGBoost 1.7.0, scikit-learn 1.3.0, and SHAP 0.42.0. Training time is approximately 15 minutes for the full dataset. Hyperparameters are specified in Section III.

Declaration on Generative AI: During the preparation of this work, the authors used GPT-5 and Grammarly in order to: Grammar and spelling check. After using these tools/services, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

REFERENCES

[1] Alzheimer's Disease International, "Dementia facts & figures," Alzheimer's Disease International, 2023.

[2] A. L. Murray, M. Vollmer, I. J. Deary, G. Muniz-Terrera, and T. Booth, "Assessing individual-level change in dementia research: A review of methodologies," Alzheimer's Research & Therapy, vol. 13, no. 1, p. 26, 2021, doi: 10.1186/s13195-021-00768-w.

[3] X. Wang, T. Ye, W. Zhou, J. Zhang, and Alzheimer's Disease Neuroimaging Initiative, "Uncovering heterogeneous cognitive trajectories in mild cognitive impairment: a data-driven approach," Alzheimers Res Ther, vol. 15, no. 1, p. 57, Mar. 2023, doi:10.1186/s13195-023-01205-w.

[4] S. Grueso and R. Viejo-Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review," Alzheimers Res Ther, vol. 13, no. 1, p. 162, Sep. 2021, doi: 10.1186/s13195-021-00900-w

[5] I. Blotenberg, F. Wittström, B. Michalowsky, M. Platen, D. Wucherer, S. Teipel, W. Hoffmann, and J. R. Thyrian, "Modifiable risk factors and symptom progression in dementia over up to 8 years—Results of the DelpHi-MV trial," Alzheimers Dement (Amst), vol. 17, no. 1, p. e70050, 2025, doi: 10.1002/dad2.70050.

[6] H. Choi and K. H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," Behavioral Brain Research, vol. 344, pp. 103–109, 2018, doi:10.1016/j.bbr.2018.02.017

[7] V. K. Tiwari, P. Indic, and S. Tabassum, "A study on machine learning models in detecting cognitive impairments in Alzheimer's patients using cerebrospinal fluid biomarkers," Am. J. Alzheimers Dis. Other Demen., vol. 39, 2024, Art. no. 15333175241308645, doi: 10.1177/15333175241308645.

[8] C. Groot, R. Smith, L. E. Collij, S. E. Mastenbroek, E. Stomrud, A. P. Binette, A. Leuzy, S. Palmqvist, N. Mattsson-Carlgren, O. Strandberg, H. Cho, C. H. Lyoo, G. B. Frisoni, D. E. Peretti, V. Garibotto, R. La Joie, D. N. Soleimani-Meigooni, G. Rabinovici, R. Ossenkoppele, and O. Hansson, "Tau positron emission tomography for predicting dementia in individuals with mild cognitive impairment," JAMA Neurol., vol. 81, no. 8, pp. 845–856, Aug. 2024, doi:10.1001/jamaneurol.2024.1612.

[9] S. Akter, Z. Liu, E. J. Simoes, and P. Rao, "Using machine learning and electronic health record (EHR) data for the early prediction of Alzheimer's disease and related dementias," J. Prev. Alzheimers Dis., vol. 12, no. 7, p. 100169, Aug. 2025, doi:10.1016/j.tjpad.2025.100169.

[10] H. Ren, Y. Zheng, C. Li, F. Jing, Q. Wang, Z. Luo, D. Li, D. Liang, W. Tang, L. Liu, and W. Cheng, "Using machine learning to predict cognitive decline in older adults from the Chinese Longitudinal Healthy Longevity Survey: model development and validation study," JMIR Aging, vol. 8, p. e67437, 2025, doi: 10.2196/67437. [Online]. Available: https://aging.jmir.org/2025/1/e67437.

[11] A. Y. Bayahya, H. Banjar, O. Talabay, et al., "Comparative analysis of multiple deep learning models with mitigation-driven approaches for enhanced Alzheimer's disease classification," Sci. Rep., vol. 15, p. 44098, 2025, doi: 10.1038/s41598-025-27914-w. [Online]. Available: https://doi.org/10.1038/s41598-025-27914-w

[12] T. Fukasawa, "The biases in applying static demand models under dynamic demand," Rev. Ind. Organ., vol. 65, pp. 561–594, 2024, doi: 10.1007/s11151-024-09946-0.

[13] K. Ren, J. Qin, L. Zheng, Z. Yang, W. Zhang, L. Qiu, and Y. Yu, "Deep recurrent survival analysis," in Proc. 33rd AAAI Conf. Artif. Intell. (AAAI), Honolulu, HI, USA, 2019, pp. 4798–4805, doi: 10.1609/aaai.v33i01.33014798.

[14] A. L. M. Tan et al., "Informative missingness: What can we learn from patterns in missing laboratory data in the electronic health record?," J. Biomed. Inform., vol. 139, Art. no. 104306, 2023, doi: 10.1016/j.jbi.2023.104306.

[15] M. Ehrig, G. S. Bullock, X. I. Leng, N. M. Pajewski, and J. L. Speiser, "Imputation and missing indicators for handling missing longitudinal data: Data simulation analysis based on electronic health record data," JMIR Med. Inform., vol. 13, Art. no. e64354, Mar. 2025, doi: 10.2196/64354.

[16] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," in Proc. 36th Int. Conf. Neural Inf. Process. Syst. (NeurIPS), New Orleans, LA, USA, 2022, Art. no. 37

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, 2017, pp. 4768–4777.

[18] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nat. Mach. Intell., vol. 2, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

[19] Y. Xu and R. Goodacre, "On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," J. Anal. Test., vol. 2, pp. 249–262, 2018, doi: 10.1007/s41664-018-0068-2.

[20] P. J. van der Veere et al., "Predicting cognitive decline in amyloid-positive patients with mild cognitive impairment or mild dementia," Neurology, vol. 103, no. 3, Art. no. e209605, Aug. 2024, doi: 10.1212/WNL.0000000000209605.

[21] F. J. Molnar, M. Man-Son-Hing, B. Hutton, and D. A. Fergusson, "Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review," Open Med., vol. 3, no. 2, pp. e31–e50, Mar. 2009.

[22] E. V. Bonilla, K. M. A. Chai, and C. K. I. Williams, "Multi-task Gaussian process prediction," in Adv. Neural Inf. Process. Syst. (NIPS), vol. 20, pp. 153–160, 2008.

[23] A. Nemali et al., "Gaussian process-based prediction of memory performance and biomarker status in ageing and Alzheimer's disease—A systematic model evaluation," Med. Image Anal., vol. 90, Art. no. 102913, 2023, doi: 10.1016/j.media.2023.102913.

[24] R. Yubero et al., "Predictive regression models for cognitive impairment, dementia, and Alzheimer's disease using real-world electronic health records," Front. Neurol., vol. 16, Art. no. 1522340, Oct. 2025, doi: 10.3389/fneur.2025.1522340.

[25] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Adv. Neural Inf. Process. Syst. (NIPS), vol. 30, 2017.

[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[27] J.H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[28] J. R. Thyrian, T. Fiß, A. Dreier, G. Böwing, A. Angelow, S. Lueke, S. Teipel, S. Fleßa, H. J. Grabe, H. J. Freyberger, and W. Hoffmann, "Life- and person-centred help in Mecklenburg-Western Pomerania, Germany (DelpHi): Study protocol for a randomised controlled trial," Trials, vol. 13, no. 1, p. 56, 2012, doi: 10.1186/1745-6215-13-56.

[29] J. R. Thyrian, J. Hertel, D. Wucherer, T. Eichler, B. Michalowsky, A. Dreier-Wolfgramm, I. Zwingmann, I. Kilimann, S. Teipel, and W. Hoffmann, "Effectiveness and safety of dementia care management in primary care: A randomized clinical trial," JAMA Psychiatry, vol. 74, no. 10, pp. 996–1004, 2017, doi: 10.1001/jamapsychiatry.2017.2124.

[30] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," J. Psychiatr. Res., vol. 12, no. 3, pp. 189–198, 1975, doi: 10.1016/0022-3956(75)90026-6