Exploration of Digital Health Technologies

Open Exploration

Open Access Original Article

Digital handwriting metrics and temporal behaviors for predicting early Alzheimer's onset

Charity Nyamuchengwa*

Department of Applied Data Science, Applied Data Science, Clarkson University, Potsdam, NY 136376, USA

***Correspondence:** Charity Nyamuchengwa, Department of Applied Data Science, Applied Data Science, Clarkson University, Potsdam, NY 136376, USA. nyamucc@gmail.com

Academic Editor: Hans-Leo Teulings, Neuroscript, LLC, USA

Received: November 18, 2025 **Accepted:** March 26, 2026 **Published:** July 26, 2026

Cite this article: Nyamuchengwa C. Digital handwriting metrics and temporal behaviors for predicting early Alzheimer's onset. Explor Digit Health Technol. 2026;4:1011100. https://doi.org/10.37349/edht.2026.1011100

Abstract

Aim: The aim of this paper was to explore stroke-level handwriting dynamics as early behavioral biomarkers for Alzheimer's disease (AD).

Methods: Stroke-level handwriting data were collected from 174 participants (89 probable AD or mild cognitive impairment; 85 cognitive healthy controls). Temporal, kinematic, and pressure features were extracted and aggregated. Classification performance was evaluated using Logistic Regression, Support Vector Machine (SVM), and Random Forest (RF) under five-fold stratified cross-validation. Random forest with SHapley Additive exPlanations (SHAP) was used to interpret feature contributions and temporal trends.

Results: Participants with AD exhibited longer and more variable in-air times, slower stroke speed, and higher-pressure variability. SVM achieved the highest ROC-AUC (0.923), while Random Forest demonstrated robust and balanced performance (accuracy = 0.844) and identified key predictive features (mean and variability of in-air time and pressure). Temporal analysis revealed progressive motor hesitation across strokes.

Conclusions: Stroke-level handwriting dynamics provide sensitive and interpretable biomarkers for early Alzheimer's disease detection. Variability in in-air time and related temporal features effectively distinguish Alzheimer's disease patients from cognitively healthy controls, reflecting underlying motor–cognitive coupling deficits. These findings highlight digital handwriting analysis as a scalable, non-invasive approach for early screening and monitoring, with potential to detect subtle impairments that may precede clinical symptoms.

Keywords

handwriting analysis, comparative analysis, temporal modeling, Alzheimer's neurodegeneration

Introduction

Handwriting is a routine behavior, writing one's name, drafting a quick note, or signing a check, but it reflects a complex interplay of motor control, perception, cognition, planning, and memory. Subtle changes in handwriting fluidity can provide early, unobtrusive indicators of neurological decline. Therein lies the promise of viewing handwriting as a digital biomarker, an objective, quantifiable indicator embedded in everyday behavior that can help detect disease early [1, 2]. When properly harnessed, these changes may alert clinicians to emerging cognitive impairment before overt symptoms appear, highlighting handwriting's potential as a digital biomarker, an objective, quantifiable measure embedded in everyday behavior [3].

The ideal situation is one of early, easily accessible detection of Alzheimer's risk. In such a scenario, we would have widely deployable tools ideally embedded [4] in everyday devices that could flag individuals who are entering the trajectory of cognitive decline [5], thus enabling timely intervention, tracking, and perhaps altering disease course. However, despite decades of work in neuropsychology, neuroimaging, and fluid biomarkers, we are still short of that goal [6]. The reality is that many cases of Alzheimer's are only detected once functional decline is clear, when intervention options are more limited and less effective [7] current diagnostic tools such as the Mini Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) [8] depend on clinical settings, trained personnel and are influenced by education and effort. While neuroimaging and cerebrospinal fluid (CSF) biomarkers provide valuable diagnostic information, their cost, limited accessibility, and invasiveness highlight the need for non-invasive, scalable approaches for early Alzheimer's disease detection [9–11].

Recent research has explored digitally captured handwriting and drawing as promising alternatives. Online handwriting features, including speed, hesitation, pen pressure, and trajectory irregularities can distinguish AD patients from healthy controls [1, 2, 5] with some studies achieving up to 0.918 AUC for differentiating mild cognitive impairment (MCI) due to AD from healthy individuals [6]. However, these approaches often rely on static summary metrics (e.g., mean speed, average pressure), small sample sizes, limited tasks, or controlled laboratory.

Why don't they solve the problem? First, many studies summarize handwriting behavior using static metrics such as mean speed and average pressure, which removes the temporal dynamics inherent in the writing process [12]. Second, experiments are often limited by small sample sizes, limited device types, controlled laboratory settings rather than real-world conditions, and typically focus on a restricted set of tasks (e.g., drawing a clock or writing a sentence), which limits generalizability and scalability [7]. Third, although some research investigates online or digital handwriting, comprehensive multi-feature temporal modelling, capturing how writing dynamics evolve over time, how features interact and how trajectories might predict cognitive decline rather than merely classify a single timepoint remains largely unexplored [1, 2]. Finally, although advanced machine-learning models can achieve high diagnostic accuracy, their increasing complexity may create challenges for interpretation and clinical implementation [3].

The consequences of these gaps are substantial. In the absence of sensitive, scalable, and non-invasive screening tools, individuals at the early stages of cognitive decline may remain undetected until irreversible neuropathological damage occurred [4]. Such delays in detection hinder timely intervention, exacerbate healthcare system burdens and associated costs [9, 10], and highlight the need for accessible, non-invasive approaches capable of identifying Alzheimer's disease at earlier stages [11, 13, 14].

This study addresses the gaps by systematically modeling multi-feature temporal dynamics of handwriting, capturing stroke-by-stroke, and pause-by-pause evolution of kinematic, spatial, temporal, and pressure features. Our approach differs in three keyways: (1) fine-grained temporal segmentation of handwriting (individual strokes, in-air vs. on-surface movements, pause durations); (2) integration of multiple features into a rich behavioral fingerprint; and (3) interpretation focused on early detection of AD, including pre-clinical and MCI populations [2, 7]. Conceptually, our framework emphasizes motor-cognitive coupling, recognizing that distributed brain systems (motor, visuo-spatial, executive, memory) contribute to handwriting, and subtle impairments manifest in temporal dynamics.

Unlike previous studies that rely on aggregated handwriting metrics, we examine stroke-by-stroke temporal dynamics, enabling finer resolution of motor-cognitive deficits. Integrating interpretable machine learning with these temporal features allows us to identify specific behavioral markers associated with early AD.

Specifically, this study aims to: (a) extract and quantify temporally resolved handwriting features during a standardized task; (b) construct and validate machine-learning models predicting AD risk; (c) identify features and temporal patterns most predictive of early cognitive advance understanding of handwriting as a digital biomarker, provide methodological innovation in temporal modeling, and support scalable, community-based screening for earlier therapeutic intervention.

Materials and methods

Data source

The handwriting data used in this study were obtained from the DARWIN (Diagnosis Alzheimer with Handwriting) dataset, which comprises handwriting recordings from 174 participants, including individuals with Alzheimer's disease (AD) and cognitively healthy controls, as described by Cilia et al. [15, 16]. Participants were recruited according to clinical and cognitive assessment criteria, including standard neuropsychological tests such as the Mini Mental State Examination, Frontal Assessment Battery (FAB), and Montreal Cognitive Assessment. Healthy controls were selected to match demographic and educational characteristics and were screened to exclude medications that could affect cognitive performance.

The experimental protocol includes 25 handwriting tasks performed on A4 sheets placed over a digitizing graphic tablet (Wacom Co., Ltd., Japan) which recorded pen position (x–y coordinates), pen pressure, timestamps, and in-air movements. The tasks were designed to evaluate different cognitive and motor functions and were arranged in increasing order of difficulty. The tasks include graphic drawing tasks (lines, circles, shapes), copy and reverse-copy tasks (letters, words, numbers), memory-based writing tasks, dictation tasks, functional writing tasks (e.g., postal order copying), the Clock Drawing Test, and paragraph copying. These tasks were designed to capture motor control, spatial organization, working memory, semantic processing, and motor planning abilities.

From the recorded tablet signals, stroke-level handwriting features were extracted. A stroke was defined as a continuous pen trajectory between a pen-down event and the subsequent pen-up event. From each stroke, temporal, kinematic, and pressure features were computed, including stroke duration, in-air time, velocity, acceleration, jerk, and pressure variability. These stroke-level features were subsequently aggregated and engineered to generate participant-level feature representations used for machine learning modeling.

While the original DARWIN dataset includes 25 tasks, not all tasks are equally discriminative for early Alzheimer's detection. In this study, we focused on task-derived stroke-level features that were shown to be informative in prior work and aggregated these features from each task for subsequent analysis. The DARWIN dataset used in this study is publicly accessible via the UCI Machine Learning Repository [17].

The stroke-level data were then preprocessed, aggregated, and engineered to extract metrics such as in-air time, stroke speed, acceleration, jerk, and pressure variability. All experiments involving human subjects were conducted in accordance with the Declaration of Helsinki (2013). The original data collection received approval from the relevant institutional ethics committees, and all participants provided informed consent. Any limitations or unavailable details are explicitly acknowledged to support transparency and reproducibility. The dataset provides detailed stroke-level handwriting recordings.

Data preprocessing

For this study, sequential stroke-level features were analyzed and aggregated to capture temporal handwriting dynamics across writing sequences rather than task-specific performance alone. Stroke-level features provide higher temporal resolution and allow analysis of motor planning and execution dynamics

that may not be captured by aggregated task-level handwriting measures. To ensure data quality, we conducted the following steps:

Outlier handling

Stroke-level features were inspected for extreme values. Observations identified as extreme, but informative were retained, while others were winsorized to the 1st and 99th percentiles to mitigate undue influence. Negative values of in-air time reflect standardized (z-score normalized) feature values rather than raw time measurements. These values indicate deviations from the population mean rather than absolute time durations.

Long-format construction

Stroke-level features were extracted from all recorded strokes across the 25 handwriting tasks and reshaped into a long-format Data Frame to enable analysis of temporal dynamics across strokes for each participant. For temporal alignment and modeling consistency, a subset of sequential strokes (e.g., the first 25 strokes per participant) was analyzed.

Feature aggregation

Baseline features were aggregated per participant using standard descriptive statistics. Aggregated features include in-air time, mean writing speed, pressure, and total pen lifts.

Feature engineering

Temporal features capture stroke-level variability, including in-air duration, stroke speed trends, and pressure dynamics, enabling high-resolution analysis of motor-cognitive behavior.

We derived two sets of features:

- Enhanced aggregated features: Capturing stroke-level variability and dynamics, e.g., standard deviation of in-air time, trend of stroke speed, and jerk-pressure interactions.
- Baseline Features: Conventional aggregated metrics (mean in-air time, mean speed, mean pressure, total pen lifts).

All continuous features were standardized using z-score normalization prior to analysis to ensure comparability across variables. Consequently, values presented in the analysis (e.g., in-air time) are expressed in standardized units, where negative values reflect observations below the population mean rather than physically negative time durations. Data processing was performed using Python (version 3.11) with the pandas (v2.1.0) and NumPy (v1.26.0) libraries.

Model development

Three classifiers were evaluated:

1. Logistic Regression (LR): Linear model for baseline comparison.
2. Support Vector Machine: Evaluated for non-linear decision boundaries.
3. Random Forest: Ensemble tree-based models used for feature importance analysis and comparison between baseline and temporal feature sets.

To ensure robust and unbiased evaluation, all models were assessed using five-fold stratified cross-validation. The dataset was split into five equally sized folds while preserving the proportion of AD and control participants in each fold. In each iteration, four folds were used for training and the remaining fold for testing. Performance metrics (accuracy, ROC-AUC, sensitivity, and specificity) were averaged across all folds and reported as mean values.

Evaluation metrics

Performance was quantified using standard classification metrics:

Accuracy: Is the most used metric to evaluate the quality of a classification model. It measures the proportion of correctly classified cases (true positives (TP) and true negatives (TN)) relative to the total number of cases. However, this metric does not capture all aspects of classification performance. When the classes are imbalanced, accuracy may not provide a reliable assessment of the model's effectiveness.

Sensitivity (recall for AD):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Specificity:

$$Sensitivity = \frac{TP}{TP + FN}$$

The (ROC-AUC) summarizes classifier performance across all thresholds, capturing how accurately AD cases can be distinguished from controls.

SHapley Additive exPlanations (SHAP) feature importance analysis

To interpret the model, SHAP was used to quantify each feature's contribution to the RF predictions. SHAP values provide local and global interpretability, allowing identification of the most predictive handwriting features (e.g., in-air time variability, mean in-air time, pressure metrics) that drive the model toward AD classification.

Model comparison and validation

To evaluate the contribution of stroke-level temporal features, we compared a baseline model using four simple aggregated features with an advanced model incorporating 14 features, including temporal, kinematic, and pressure metrics. All models were evaluated using five-fold stratified cross-validation, preserving the proportion of AD and control participants in each fold. Performance metrics (accuracy, ROC-AUC, sensitivity, and specificity) were averaged across folds and reported as mean ± standard deviation. An ablation comparison was performed to assess the added value of temporal features relative to baseline aggregated features.

Among the classifiers, SVM achieved the highest ROC-AUC and specificity and was designated as the primary model for reporting classification performance. Random Forest demonstrated balanced sensitivity and specificity and was used for feature importance and interpretability via SHAP analysis. Linear regression served as a baseline comparator.

Overview of the predictive pipeline

The overall study workflow is illustrated in Figure 1.

Results

This section presents the results from our investigation into handwriting-based digital biomarkers for early Alzheimer disease detection. We evaluated the performance of a Random Forest classifier trained on multi-feature temporal dynamics and compared it against traditional approaches.

Model performance and comparative analysis

Model performance is summarized in Figure 2. Figure 2 compares the performance of three machine learning models LR, SVM, and RF trained on temporal and kinematic handwriting features to classify Alzheimer disease patients and healthy controls. Model evaluation was conducted using five-fold stratified cross-validation**,** and performance metrics (accuracy, ROC-AUC, sensitivity, and specificity) are reported as mean ± standard deviation across folds.

Across models, the Support Vector Machine achieved the highest overall performance, with the largest ROC-AUC (0.923) and highest accuracy (0.856), indicating strong discriminatory ability between Alzheimer disease patients and cognitively healthy controls. SVM also demonstrated the highest specificity (0.884),

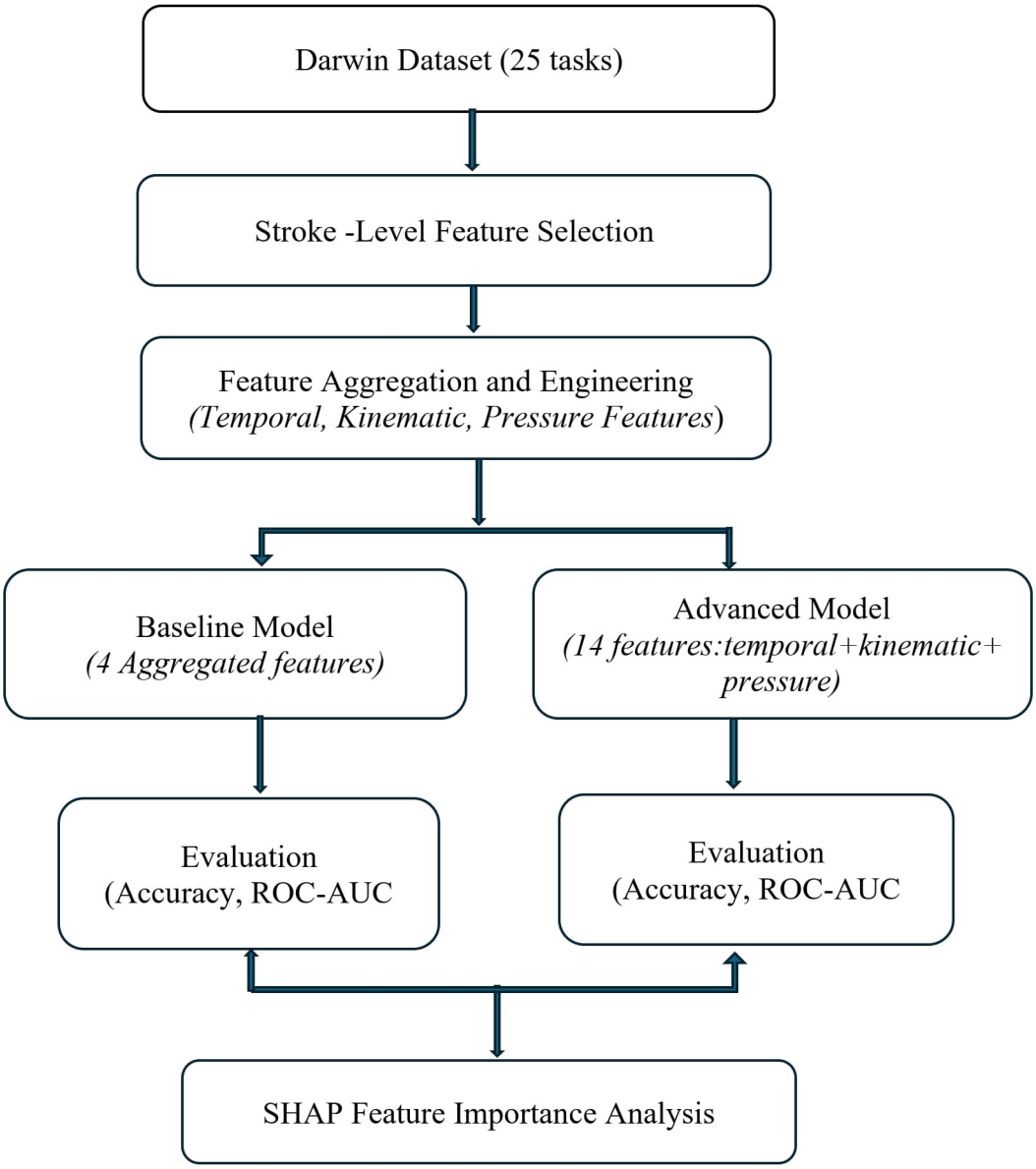

Figure 1. Overview of the predictive pipeline. Stroke-level features were extracted from the DARWIN dataset, aggregated, and used in baseline (4 features) and advanced (14 features) models. Model evaluation metrics were assessed, followed by SHAP analysis to interpret feature Contribution.

suggesting a strong ability to correctly identify healthy individuals, while maintaining competitive sensitivity (0.830).

Random Forest exhibited robust and well-balanced performance (accuracy = 0.827; ROC-AUC = 0.912; sensitivity = 0.801; specificity = 0.853) and was additionally leveraged for feature importance analysis due to its capacity to capture nonlinear relationships and feature interactions. Logistic regression showed stable but comparatively lower performance (accuracy = 0.798; ROC-AUC = 0.899; sensitivity = 0.788; specificity = 0.808), supporting its role as a transparent baseline classifier.

All performance metrics are reported as mean values across five-fold stratified cross-validation, with error bars in Figure 2 representing standard deviation, reflecting variability across folds.

Temporal feature analysis using Random Forest

Figure 3 compare baseline (4 aggregated features) and advanced (14 features including temporal, kinematic, and pressure metrics) models using five-fold stratified cross-validation. Metrics are reported as mean ± standard deviation.

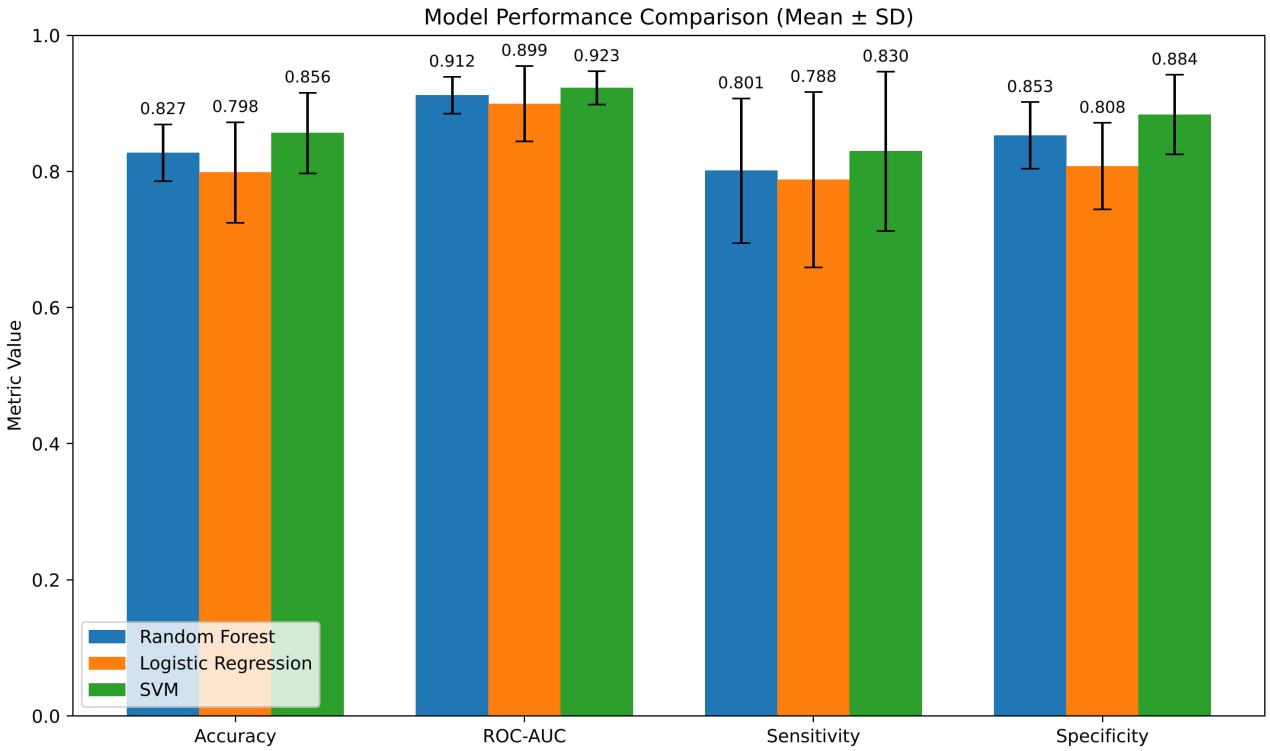

Figure 2. Cross-validated performance comparison of Logistic Regression, SVM, and Random Forest. Bars represent mean values, and error bars indicate standard deviation across five-fold stratified cross-validation.

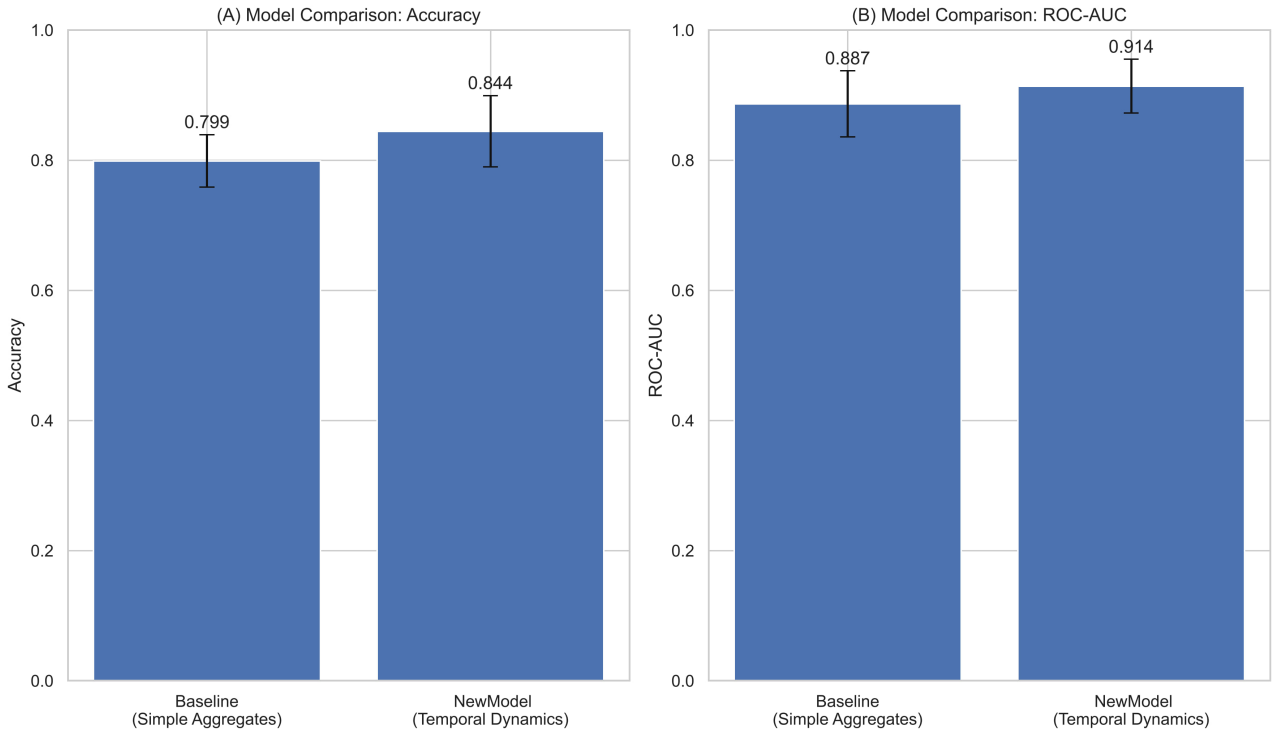

Figure 3. Cross-validated performance of Random Forest on baseline versus enhanced aggregate handwriting features. (**A**) Accuracy and (**B**) ROC-AUC are reported as mean ± standard deviation across five-fold stratified cross-validation. The enhanced feature set outperforms the baseline model, demonstrating the added discriminative value of richer stroke-level feature representations for early Alzheimer's disease detection.

As shown in Figure 3, the advanced model outperformed the baseline across all evaluation metrics:

- Baseline model: accuracy = 0.799 ± 0.040, ROC-AUC = 0.887 ± 0.051
- Advanced model: accuracy = 0.844 ± 0.055, ROC-AUC = 0.914 ± 0.041

This corresponds to an approximate 4.5% absolute improvement in accuracy and a 2.7% increase in ROC-AUC, demonstrating that incorporating richer stroke-level feature representations enhances predictive performance beyond simple aggregated measures.

Feature importance and key biomarkers

Analysis of feature importance using SHAP revealed that temporal dynamics of in-air movement dominated the predictive model. As shown in Figure 4, the mean and variability of in-air time, followed by pressure-related metrics**,** were among the most predictive features for distinguishing Alzheimer's disease patients from healthy controls. This indicates that early motor-cognitive deficits are primarily reflected in motor planning rather than execution.

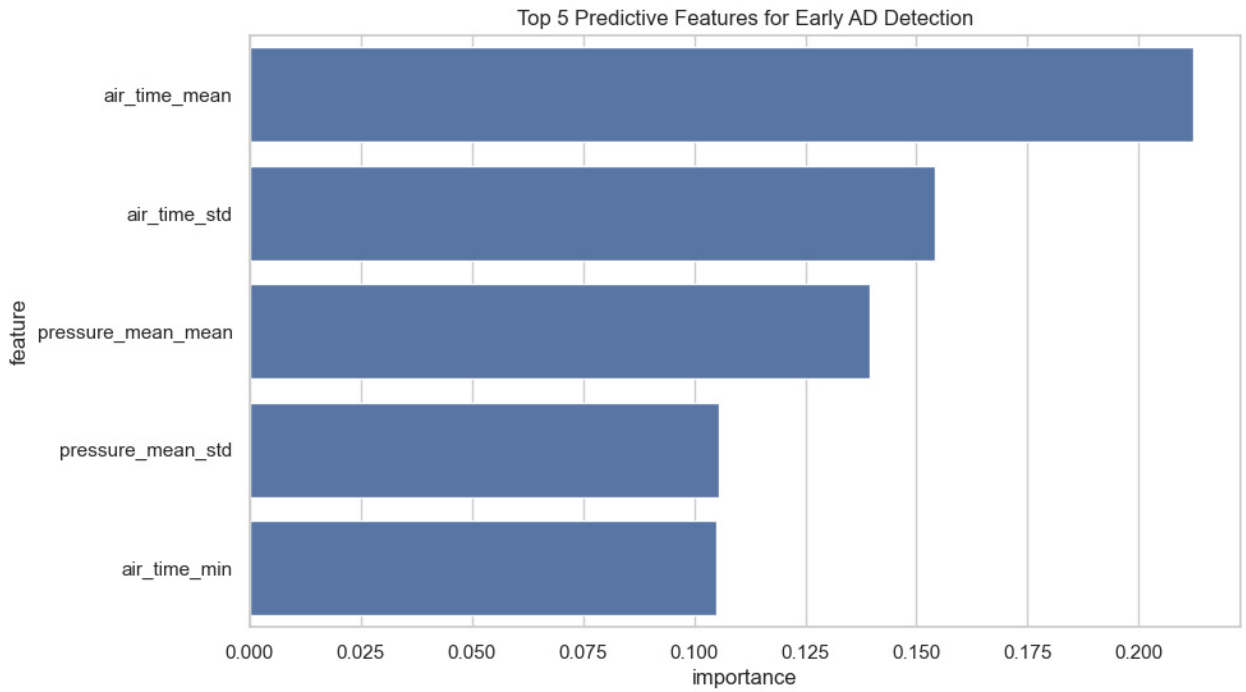

Figure 4. Top predictive handwriting features for early Alzheimer's detection based on Random Forest feature importance. In-air time variability and mean _in-air time and pressure metrics are among the most influential features.

The most important features were:

1. Mean in-air time (air_time_mean)
2. In-air time standard deviation (air_time_std)
3. Mean pressure (pressure_mean_mean)
4. Pressure variability (pressure_mean_std)
5. Minimum in-air time (air_time_min)

Temporal dynamics of handwriting features

Visualization of stroke-level temporal patterns (Figure 5) provides direct evidence of the behavioral manifestations underlying model performance. Participants with Alzheimer's disease exhibited systematically prolonged and more variable in-air times, along with less stable on-paper speed, across sequential strokes. These fine-grained temporal patterns are not captured by aggregated metrics, supporting the utility of temporally resolved handwriting features.

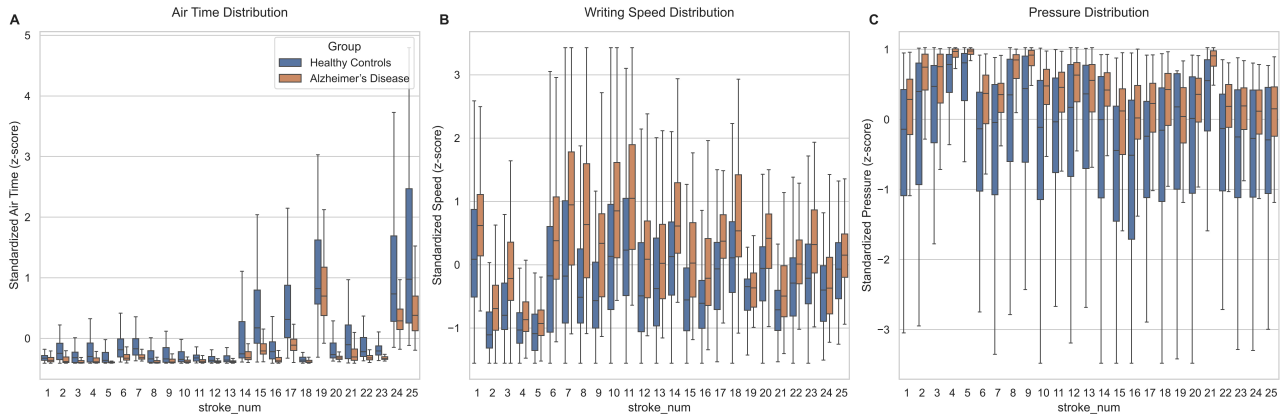

Figure 5. Distribution of stroke-level handwriting features across 25 sequential strokes for Alzheimer's disease patients and healthy controls. Boxplots represent the interquartile range (25th–75th percentile), with median values indicated by horizontal lines. All features are standardized (z-scores). (**A**) In-air time shows greater variability and higher median values in Alzheimer's disease, indicating motor planning deficits. (**B**) Writing speed distributions indicate reduced stability and increased variability in Alzheimer's disease. (**C**) Pressure distribution shows comparatively minor differences in pressure variability between groups.

According to Figure 5, the Air time distribution (Figure 5A) demonstrates greater variability and higher median standardized values in the Alzheimer's disease group compared to healthy, indicating consistent motor planning deficits. In the writing speed distribution (Figure 5B), Alzheimer's disease group exhibits reduced stability and greater dispersion in on-paper speed, reflecting impaired motor execution. Pressure distribution (Figure 5C) was comparatively subtle but still contributed to group differentiation.

Model interpretability via SHAP analysis

SHAP analysis confirmed the dominant role of in-air time dynamics in Random Forest model decision process as shown in Figure 6. Features, with higher SHAP values consistently correspond to stroke with prolonged in-air times durations and greater pressure variability, providing an interpretable link between handwriting behavior and the model's predictions for Alzheimer's disease.

These results provide both validation of our feature importance results and clinical interpretability for the model's decision-making process. The primacy of in-air time dynamics provides compelling evidence that motor planning deficits precede the execution of impairments in early Alzheimer's pathology.

Discussion

The present study demonstrated that temporally resolved handwriting features show promising discriminative power for early Alzheimer's detection compared with traditional aggregated metrics. SVM, used as the primary classifier, achieved high performance (ROC-AUC = 0.923; accuracy = 0.856), and provided interpretable feature importance rankings via SHAP analysis (AUC = 0.914), classified as excellent under standard diagnostic thresholds. Previous studies have also demonstrated the potential of digital handwriting biomarkers for community-based early Alzheimer's screening [4]. This indicates that temporal dynamics, particularly stroke-level timing, in-air transitions, and pressure variability, capture subtle motor-cognitive variations that may reflect early deficits not apparent in summary measures.

Consistent with the study objectives:

- Objective (a): Extraction and quantification of temporally resolved handwriting features

Stroke-level analyses (Figure 5) revealed systematically prolonged and more variable in-air times in participants with AD, suggesting potential differences in motor planning and execution. These temporally resolved features provide a richer behavioral fingerprint than aggregated metrics, but further validation in independent datasets is required.

- Objective (b): Construction and validation of machine-learning models

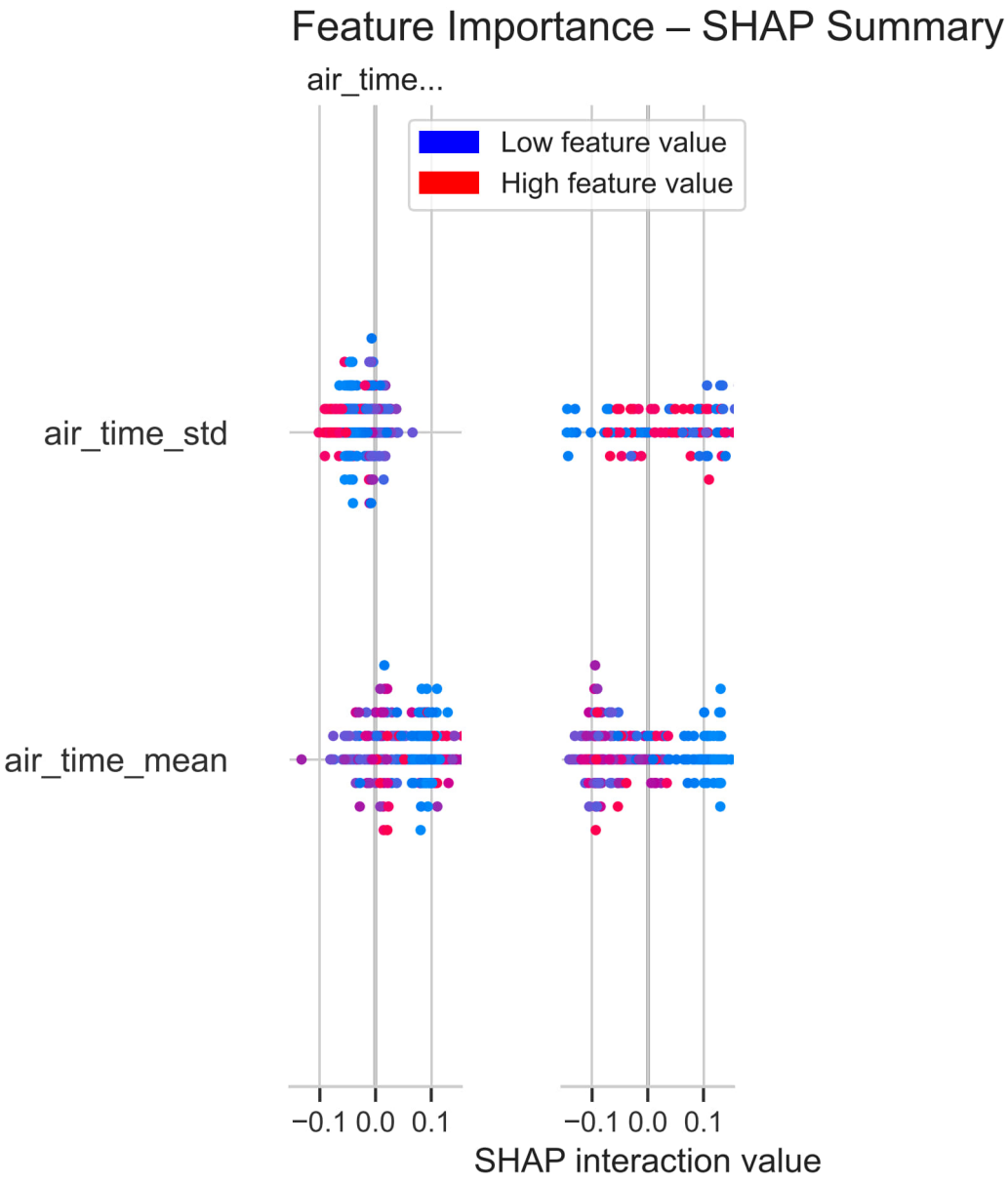

Figure 6. SHAP summary plot showing the contribution of handwriting features to the Random Forest model's classification of Alzheimer's disease. Each point represents an individual sample, with color indicating the magnitude of the standardized feature value (blue = low, red = high, purple= midrange standardized feature value). Inair time mean and inair time variability show the strongest SHAP contributions, highlighting their dominant role in distinguishing Alzheimer's disease from healthy controls.

Evaluation of LR, SVM, and RF models demonstrated robust performance within the DARWIN dataset. SVM served as the primary model for reporting classification metrics due to its highest ROC-AUC, while RF provided interpretable feature importance rankings, highlighting key predictive features such as in-air time variability and pressure metrics (Figures 2–4). LR served as a baseline comparator.

- Objective (c): Identification of the most predictive features and temporal patterns

SHAP analysis indicated that variability and mean values of in-air time were the dominant contributors to model decisions, followed by pressure metrics. Temporal visualizations supported these findings, showing prolonged and unstable in-air intervals across the writing sequence in participants with AD.

- Objective (d): Assessment of longitudinal progression potential

Stroke-by-stroke temporal trends (Figure 5) suggest systematic motor hesitation, characterized by prolonged and variable in-air times along with less stable on-paper speed. While these findings indicate potential for longitudinal monitoring, this study's cross-sectional design prevents confirmation of

predictive utility for disease progression. Future longitudinal studies are needed to validate these temporal features as early indicators of cognitive decline.

Clinical interpretation of model-discriminative patterns

To further illustrate the clinical relevance of the SHAP findings, representative feature patterns were examined. Participants in the AD group consistently exhibited elevated in-air time variability (air_time_std), increased pressure variability (pressure_mean_std), and greater stroke duration variability, indicative of irregular motor execution. In contrast, non-AD participants showed lower in-air-time dispersion, more stable pressure profiles, and consistent stroke timing. These patterns align with disruptions in motor planning, visuomotor coordination, and fine motor control reported in early Alzheimer's disease. Mapping model predictions to observable handwriting behaviors enhances the interpretability of our framework and supports its potential utility as a non-invasive digital motor biomarker.

Comparative and literature context

To place our findings in context, we compared our approach to recent studies using the DARWIN dataset and other handwriting-based AD detection methods (Table 1). While deep learning approaches such as Convolutional Neural Network, Long ShortTerm Memory networks, and hybrid transformers have achieved high accuracy and F1-scores, our study emphasizes interpretable machine learning with stroke-level temporal features. Unlike black-box deep learning models, our method provides direct insights into which handwriting dynamics contribute most to early AD detection, supporting both predictive performance and explainability.

Table 1. Performance comparison of recent DARWIN-based studies on early Alzheimer's detection.

Study	Dataset	Features/Models	Validation	Performance (accuracy/AUC)	Notes
Cilia et al., 2018 [16]	Darwin	450 features, dimensionality reduction	Train/test split	Accuracy (N/A)/AUC/(N/A)	Identified in-air time as key
Demircioglu Diren, 2025 [1]	Darwin	Dimensionality reduction + explainability techniques	CV	Accuracy 0.962/AUC(N/A)	High performance using feature selection
Kang et al., 2024 [18]	Darwin	Self-attention	CV	Accuracy 0.943/AUC(N/A)	Outperformed CNNs
Gong et al., 2025 [3]	Darwin	Hybrid transformer	CV	Accuracy 0.909/AUC(N/A)	Multimodal 2D + 1D features
Bazarbekov et al., 2026 [14]	Sensor-based Smart Pen dataset	Hybrid CNN-BiLSTM + Sim-to-Real Domain Adaptation	Train/test split	Accuracy 0.91/AUC 0.96	Deep learning, physics-based augmentation
Current study	Darwin	14 temporals + kinematic + pressure features, SVM/RF	5-fold CV	Accuracy 0.844/AUC 0.923	Interpretable, stroke-level temporal analysis

Metrics were extracted from the original publications. Missing values are indicated as N/A. All numerical values are rounded to three decimal places for consistency. Bazarbekov et al. (2026) [14] used a proprietary Smart Pen dataset with sensor-based motion acquisition rather than the publicly available DARWIN dataset.

Neurocognitive interpretation of temporal writing dynamics

Prolonged and variable in-air times likely reflect disruptions in motor planning mediated by prefrontal cortex and associated prefrontal-parietal networks. Preservation of on-paper speed suggests that execution processes governed by primary motor areas remain relatively intact in early AD stages. This motor-cognitive coupling deficit provides a mechanistic explanation for the observed temporal dynamics and supports handwriting as a non-invasive behavioral biomarker.

Generalizability

While the present study utilized the DARWIN dataset under controlled writing tasks, the results may generalize other handwriting tasks and digital devices. Future work should evaluate performance in real-world settings, including home use on tablets or smartphones.

Potential confounders

Potential confounders such as age, education, and comorbidities may influence handwriting patterns. The DARWIN dataset primarily includes stroke-level handwriting features and class labels but does not provide detailed demographic or clinical information. As a result, these factors were not explicitly controlled in the current analysis. Future studies should incorporate such participant characteristics to ensure that observed differences in handwriting dynamics are attributable to Alzheimer's disease rather than demographic or health-related factors.

Limitations

Several limitations should be acknowledged. First, handwriting data were collected from the DARWIN dataset under standardized writing tasks, which may not fully reflect real-world writing behavior or generalize across different devices. Second, this study is cross-sectional; longitudinal follow-up is required to determine whether temporal handwriting features can reliably predict disease progression. Third, potential confounders such as age, education, and comorbidities were recorded but not explicitly controlled for. Finally, while the sample size is sufficient for exploratory analysis, generalizability across diverse populations remains to be established. Overall, the study should be considered an exploratory investigation highlighting methodological potential rather than providing definitive clinical conclusions. Overall, the study should be considered an exploratory investigation highlighting methodological potential rather than providing definitive clinical conclusions.

Practical implications and future directions

The integration of temporal dynamics with machine learning offers both high interpretability and practical utility. Random Forest analysis revealed key predictive features (mean and variability of in-air time and pressure), while SVM, the primary classifier, achieved high accuracy (0.856) and ROC-AUC (0.923), supporting its use for early AD detection. Future research should evaluate performance in larger, diverse cohorts, assess longitudinal prediction of disease progression, and explore integration with complementary digital biomarkers (e.g., speech, gait) to enhance early-stage detection. Evaluation across varied tasks, real-world devices, and home-based settings will be essential to ensure clinical generalizability.

Abbreviations

AD: Alzheimer's disease

LR: Logistic Regression

MCI: mild cognitive impairment

MMSE: Mini Mental State Examination

MoCA: Montreal Cognitive Assessment

RF: Random Forest

SHAP: SHapley Additive exPlanations

SVM: Support Vector Machine

Declarations

Acknowledgments

The author thanks Michael Mutsa Munanairi for assistance with proofreading and language editing.

Author contributions

CN: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—original draft, Writing—review & editing. The author read and approved the final version of the manuscript.

Conflicts of interest

The author declares no conflicts of interest.

Ethical approval

This study uses the DARWIN dataset, an openly available public dataset released under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Ethical approval for the original data collection was obtained by the investigators who created and published the DARWIN dataset. Because this study conducts secondary analysis of publicly available, de-identified data, no additional ethical approval was required.

Consent to participate

Informed consent for participation was obtained by the original investigators during the initial DARWIN data collection. This study did not involve any new data collection; the dataset was only downloaded and used in accordance with the dataset's license and citation requirements. Therefore, no new participant consent was necessary.

Consent to publication

Consent for publication of the DARWIN dataset was obtained by the original investigators during the initial data collection process. The dataset is publicly distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits reuse, adaptation, and redistribution with appropriate attribution. Because this study involves only secondary analysis of the openly available, de-identified DARWIN dataset, no additional consent for publication was required.

Availability of data and materials

The datasets analyzed for this study can be found in the UCI Machine Learning Repository (link: https://archive.ics.uci.edu/datasets/?search=Darwin).

Funding

No funding was received for this study.

Publisher's note

Open Exploration maintains a neutral stance on jurisdictional claims in published institutional affiliations and maps. All opinions expressed in this article are the personal views of the author(s) and do not represent the stance of the editorial team or the publisher.

References

1. Demircioglu Diren D. Design and Validation of a Hybrid Machine Learning Model for Alzheimer's Detection Using Handwriting Data. Am J Alzheimer's Dis Other Dementias. 2025;40:15333175251374913. [DOI] [PubMed] [PMC]
2. Garre-Olmo J, Faúndez-Zanuy M, López-de-Ipiña K, Calvó-Perxas L, Turró-Garriga O. Kinematic and Pressure Features of Handwriting and Drawing: Preliminary Results Between Patients with Mild Cognitive Impairment, Alzheimer Disease and Healthy Controls. Curr Alzheimer Res. 2017;14:960–8. [DOI] [PubMed] [PMC]
3. Gong C, Qin H, El-Yacoubi MA. Hybrid Transformer for Early Alzheimer's Detection: Integration of Handwriting-Based 2D Images and 1D Signal Features. IEEE J Biomed Health Inform. 2025;29:8293–305. [DOI] [PubMed]

4. Li S, Li K, Liu J, Huang S, Wang C, Tu Y, et al. A new method for community-based intelligent screening of early Alzheimer's disease populations based on digital biomarkers of the writing process. Front Comput Neurosci. 2025;19:1564932. [DOI]
5. van den Berg RL, van der Landen SM, Keijzer MJ, van Gils AM, van Dam M, Ziesemer KA, et al. Smartphone- and Tablet-Based Tools to Assess Cognition in Individuals With Preclinical Alzheimer Disease and Mild Cognitive Impairment: Scoping Review. J Med Internet Res. 2025;27:e65297. [DOI] [PubMed] [PMC]
6. van den Berg RL, de Boer C, Zwan MD, Jutten RJ, van Liere M, van de Glind MABJ, et al. Digital remote assessment of speech acoustics in cognitively unimpaired adults: feasibility, reliability and associations with amyloid pathology. Alzheimers Res Ther. 2024;16:176. [DOI] [PubMed] [PMC]
7. Thebaud T, Favaro A, Chen C, Chavez G, Moro-Velazquez L, Butala A, et al. Explainable Metrics for the Assessment of Neurodegenerative Diseases through Handwriting Analysis. arXiv:2409.08303v2 [Preprint]. 2025 [cited 2025 Nov 6]. Available from: https://arxiv.org/html/2409.08303v2
8. Pinto TCC, Machado L, Bulgacov TM, Rodrigues-Júnior AL, Costa MLG, Ximenes RCC, et al. Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? Int Psychogeriatr. 2019;31:491–504. [DOI] [PubMed]
9. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. World Alzheimer report 2015. The global impact of dementia: an analysis of prevalence, incidence, costand trends. Doctoral dissertation. Alzheimer's Disease International; 2015.
10. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011; 7:280–92. [DOI] [PubMed] [PMC]
11. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14: 535–62. [DOI] [PubMed] [PMC]
12. Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken (NJ): John Wiley & Sons; 2013.
13. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396:413–46. [DOI] [PubMed] [PMC]
14. Bazarbekov I, Almisreb A, Ipalakova M, Bazarbekova M, Daineko Y. Sim-to-Real Domain Adaptation for Early Alzheimer's Detection from Handwriting Kinematics Using Hybrid Deep Learning. Sensors. 2026;26:298. [DOI] [PubMed] [PMC]
15. Cilia ND, De Gregorio G, De Stefano C, Fontanella F, Marcelli A, Parziale A. Diagnosing Alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. Eng Appl Artif Intell. 2022;111:104822. [DOI]
16. Cilia ND, Stefano CD, Fontanella F, Di Freca AS. An Experimental Protocol to Support Cognitive Impairment Diagnosis by using Handwriting Analysis. Procedia Comput Sci. 2018;141:466–71. [DOI]
17. DARWIN dataset [Internet]. UCI Machine Learning Repository; [cited 2025 Nov 5]. Available from: https://archive.ics.uci.edu/datasets/?search=DARWIN
18. Kang L, Zhang X, Guan J, Huang K, Wu R. Early Alzheimer's disease diagnosis via handwriting with self-attention mechanisms. J Alzheimers Dis. 2024;102:173–80. [DOI] [PubMed]