

Factors Affecting the Diagnostic Performance of Machine Learning Models in Detecting Cognitive Decline from Handwriting Analysis: A Systematic Review and Meta-Regression Analysis

P4-08

Ching-Wei Ye¹, Chien-Liang Liu², Fu-Ling Yang³, Nga-Weng leong⁴, Shu-Wei Chang⁵, Tien-Ni Wang¹, Hao-Ling Chen^{1,6}



¹School of Occupational Therapy, National Taiwan University (NTU); ²Dept. of Neurology, Taipei City Hospital; ³School of Nursing, NTU; ⁴Institute of Epidemiology and Preventive Medicine, NTU; ⁵Dept. of Civil Engineering, NTU; ⁶Health Science and Wellness Center, NTU — Taipei, Taiwan

Handwriting Kinematics & Machine Learning detects Alzheimer’s disease with high accuracy (AUC 0.95) and MCI with moderate accuracy (AUC 0.83). Combining writing and drawing tasks with ensemble or deep learning models yields the best performance — a non-invasive, low-cost screening tool for cognitive decline.

INTRODUCTION	METHODS	RESULTS	CONCLUSIONS	REFERENCES
Cognitive decline—from mild cognitive impairment (MCI) to Alzheimer’s disease (AD)—is a growing global health burden, yet neuroimaging and cerebrospinal fluid biomarkers are costly, invasive, and inaccessible in primary care. Handwriting integrates fine motor control, visuospatial attention, and cognitive processing; its kinematics (velocity, in-air time, fluency) are a promising non-invasive digital biomarker. Machine learning (ML) enables automated HC / MCI / AD classification, but reported accuracy varies widely. We synthesized the evidence and identified the methodological factors that drive diagnostic performance.	PRISMA-guided systematic review; PubMed & Cochrane Library (2004–2024), English only; 18 studies included. Population: adults ≥60 with AD, MCI, or healthy controls (HC); Index test: ML on digitized handwriting or drawing; Reference: established clinical or neuropsychological criteria. The risk of bias was assessed using the QUADAS-2. The 2×2 confusion matrices were extracted or reconstructed for each study (TP/FP/TN/FN). Bivariate random-effects meta-analysis; HSROC curves; meta-regression on handwriting task type, ML algorithm category.	Common handwriting features: mean/peak velocity, motor quality index (velocity peaks, jerks), pressure variability, pen-in-air time. Pooled Diagnostic Accuracy: Fig A. HC vs. AD (16 studies): sensitivity 88.1% (83.9–91.3), specificity 90.9% (86.8–93.7), AUC 0.947. Fig B. HC vs. MCI (11 studies): sensitivity 77.8% (69.2–84.6), specificity 76.8% (71.1–81.1), AUC 0.83. Meta-regression: task type and ML algorithm significantly drove heterogeneity. Combining multiple tasks with ensemble or deep learning models achieved the highest accuracy.	Handwriting kinematics are promising behavioral markers for screening AD and MCI. Combined writing-and-drawing protocols with ensemble or deep-learning models give the best diagnostic performance. Trade-off: multi-task batteries raise patient burden; community screening may favor optimized single tasks. Suggestion: explainable AI (XAI), early pre-symptomatic MCI detection, standardized multi-center datasets. Future Work: include 2025-2026’s studies, identify the optimal single handwriting task, and examine the smallest feature amount required to achieve high accuracy through further meta-regression.	1. Fernandes CP, et al. Handwriting changes in Alzheimer’s disease: a systematic review. J Alzheimers Dis. 2023. 2. De Stefano C, et al. Handwriting analysis to support neurodegenerative diseases diagnosis: a review. Pattern Recognit Lett. 2019. 3. Vessio G. Dynamic handwriting analysis for neurodegenerative disease assessment: a review. Appl Sci. 2019. 4. Qi W, et al. Alzheimer’s disease digital biomarkers and AI model scoping review. npj Digit Med. 2025. 5. Livingston G, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020.

CONTACT



Presenter:

Ching-Wei Ye

Corresponding:

Hao-Ling Chen

hlchen@ntu.edu.tw



