

A Hybrid-AI Framework For Automated Gait And Speech Analysis For Early Screening Of Cognitive Impairment

P1-001

Yining Liao#, Xiang Fan#, Chenglong Zhang, Lin Hu, Keyan Yu, Ying Cai, Guanxun Cheng*, Zhenglong Sun*

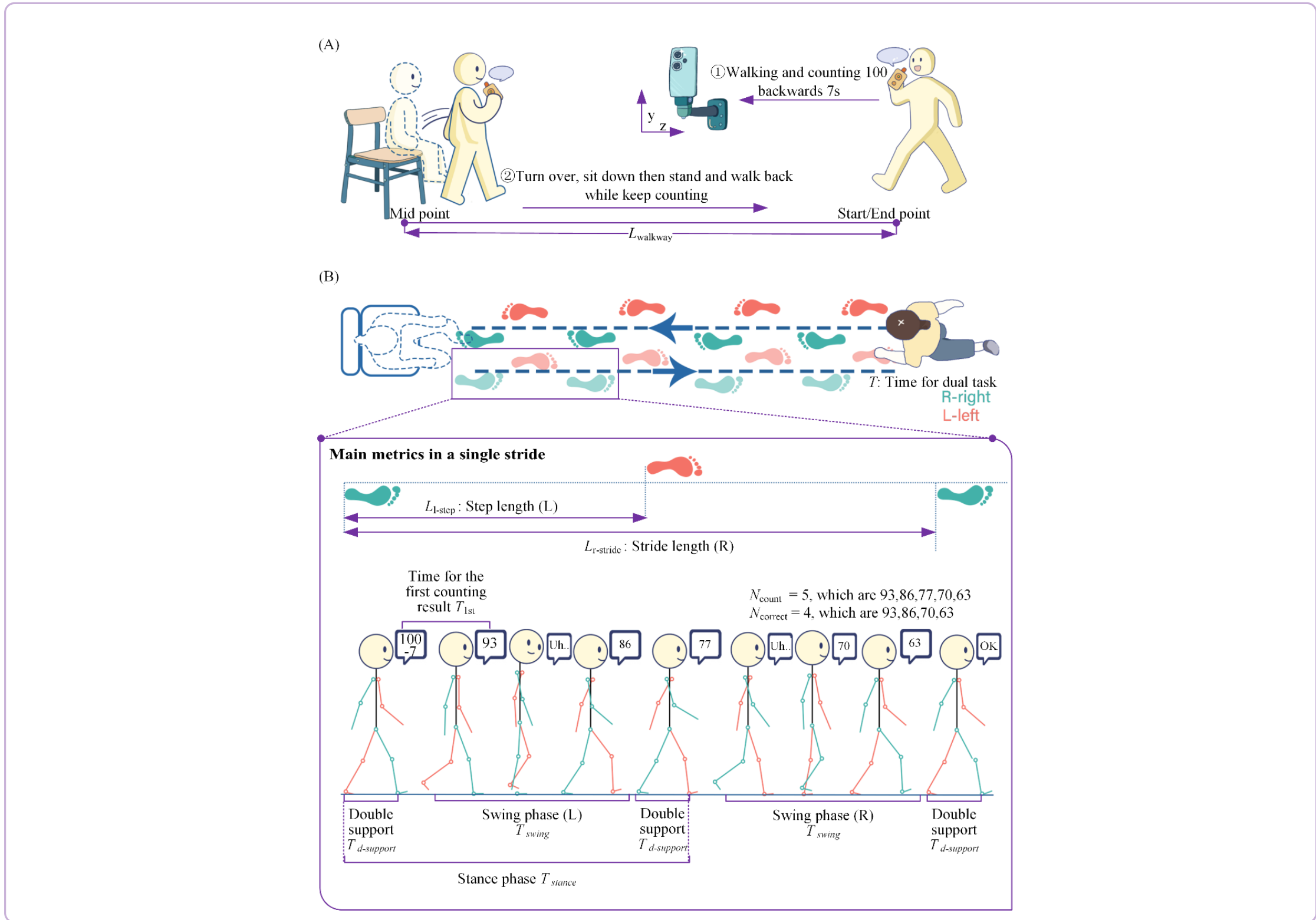
The Chinese University of Hong Kong, Shenzhen; Peking University Shenzhen Hospital; University College London

KEY TAKEAWAY: ADualNet integrates synchronized TUG-DT video and speech to screen cognitive impairment, with AdaBoost AUROC 0.893 and interpretable gait, speech-task and demographic features.

INTRODUCTION

Need: Low-cost tools are needed to triage older adults for specialist cognitive or biomarker evaluation. Rationale: Dual-task gait exposes cognitive-motor interference; serial subtraction adds an in-task cognitive signal. Objective: Automate TUG-DT video-audio feature extraction for cognitive impairment screening.

TUG-DT TASK DESIGN



Timed Up and Go dual-task protocol and synchronized video-audio feature acquisition.

STUDY SAMPLE & VALIDATION

Cohort: 164 STAR participants; 78 cognitive impairment and 86 cognitively unimpaired controls. Signals: Smartphone RGB video plus close-range WAV audio; unusable or mismatched recordings excluded. Validation: All preprocessing and feature selection were confined to training folds.

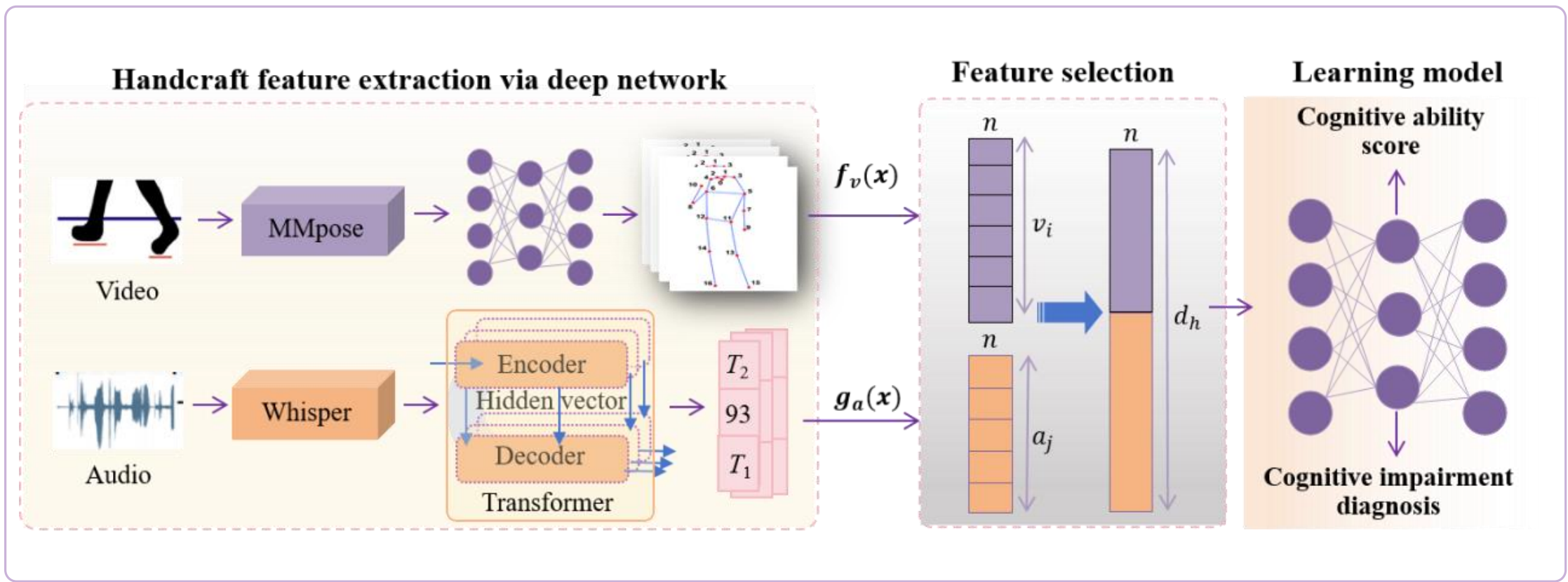
Characteristic	CI	CU	Total
n	78	86	164
Age, y	65.50	61.50	63.33
Education, y	8.71	12.88	11.27
MoCA	17.63	26.21	22.13

Repeated stratified 5-fold cross-validation with 10 repeats; top 14 features selected inside each training fold.

FEATURE DOMAINS

Speech-task Response count, accuracy, rate, latency, pauses and error pattern	Gait Speed, cadence, timing, length, stance/swing and asymmetry
Demographics Age, sex and education as interpretable context variables	Endpoint Clinical cognitive impairment status; MoCA analyzed continuously

ADUALNET WORKFLOW



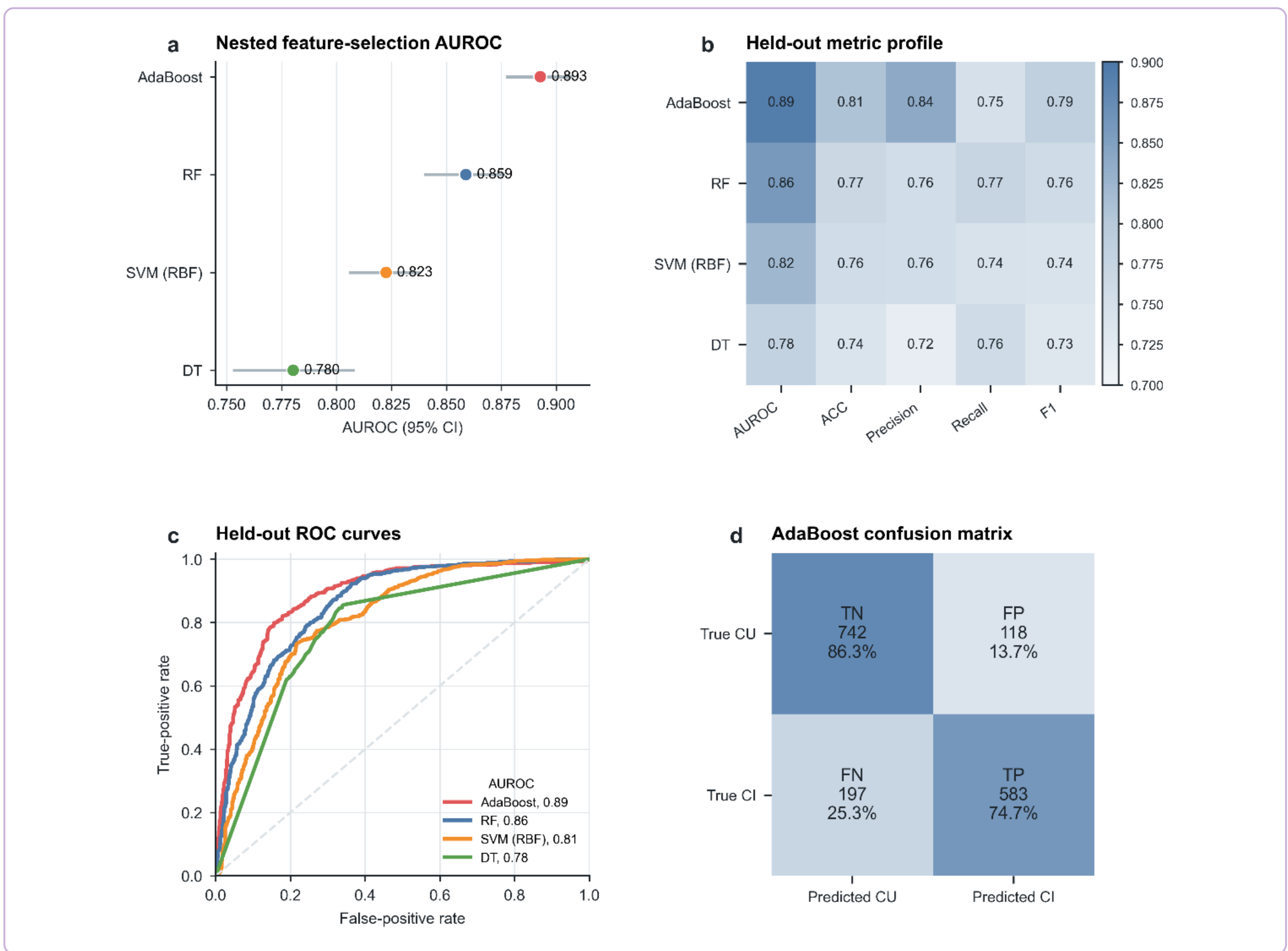
ADualNet workflow for automated gait-speech feature extraction and cognitive-impairment modelling.

Video: MMPose keypoints produce walking speed, cadence, timing, length and asymmetry features. Speech: WhisperX transcripts are parsed for response count, accuracy, rate, latency, pauses and errors. Fusion: Gait, speech-task, demographic and clinical variables are merged at participant level.

RESULTS

0.893 AUROC	0.808 Accuracy	0.839 Precision	0.747 Recall	0.786 F1 score
-----------------------	--------------------------	---------------------------	------------------------	--------------------------

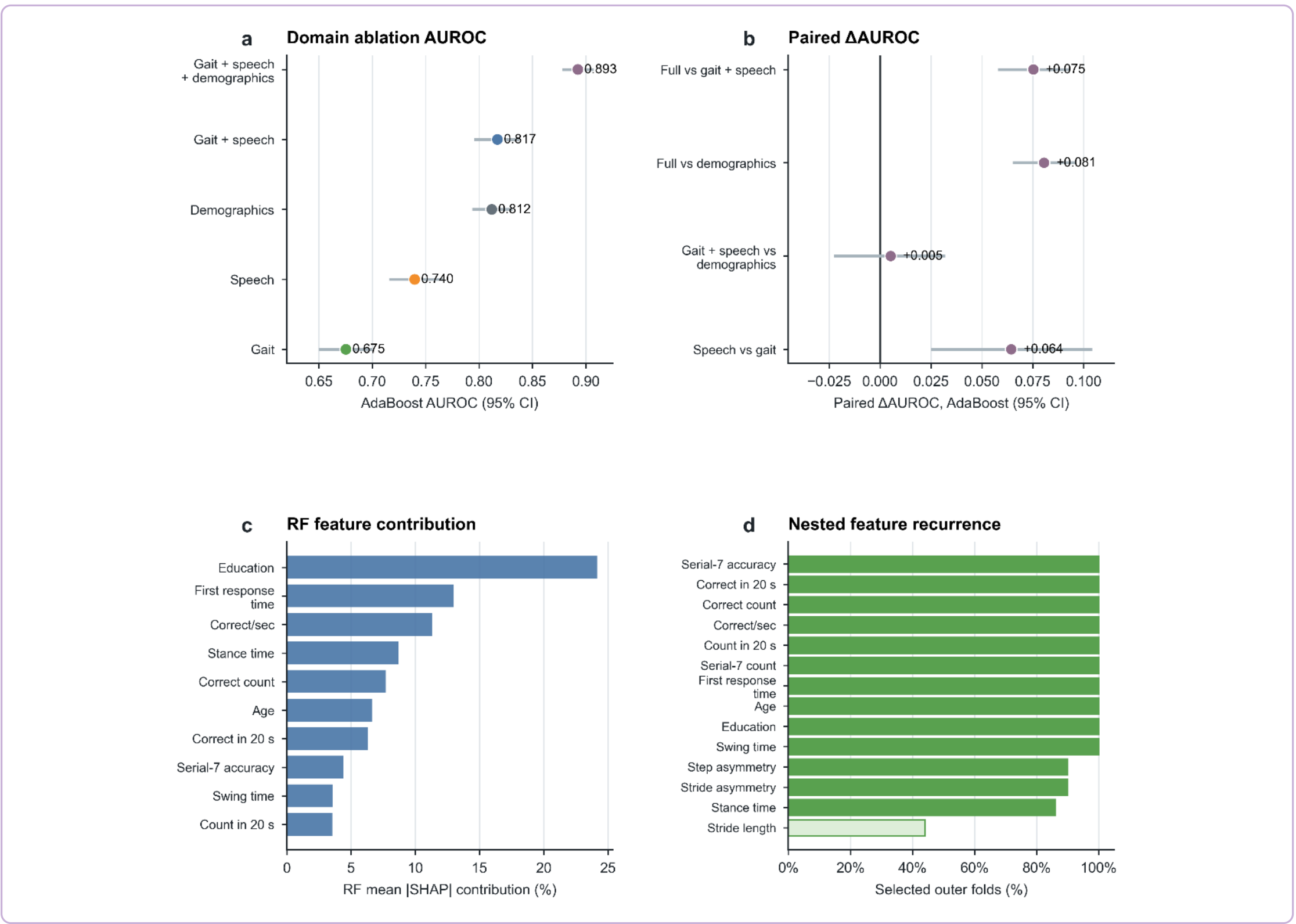
Best classifier: AdaBoost, AUROC 0.893 (95% CI 0.877-0.907). Ensemble methods showed the strongest held-out discrimination.



AdaBoost ranked first, followed by random forest, SVM (RBF) and decision tree. Held-out predictions favored high precision with moderate sensitivity. MoCA regression supported a convergent association with continuous cognitive scores.

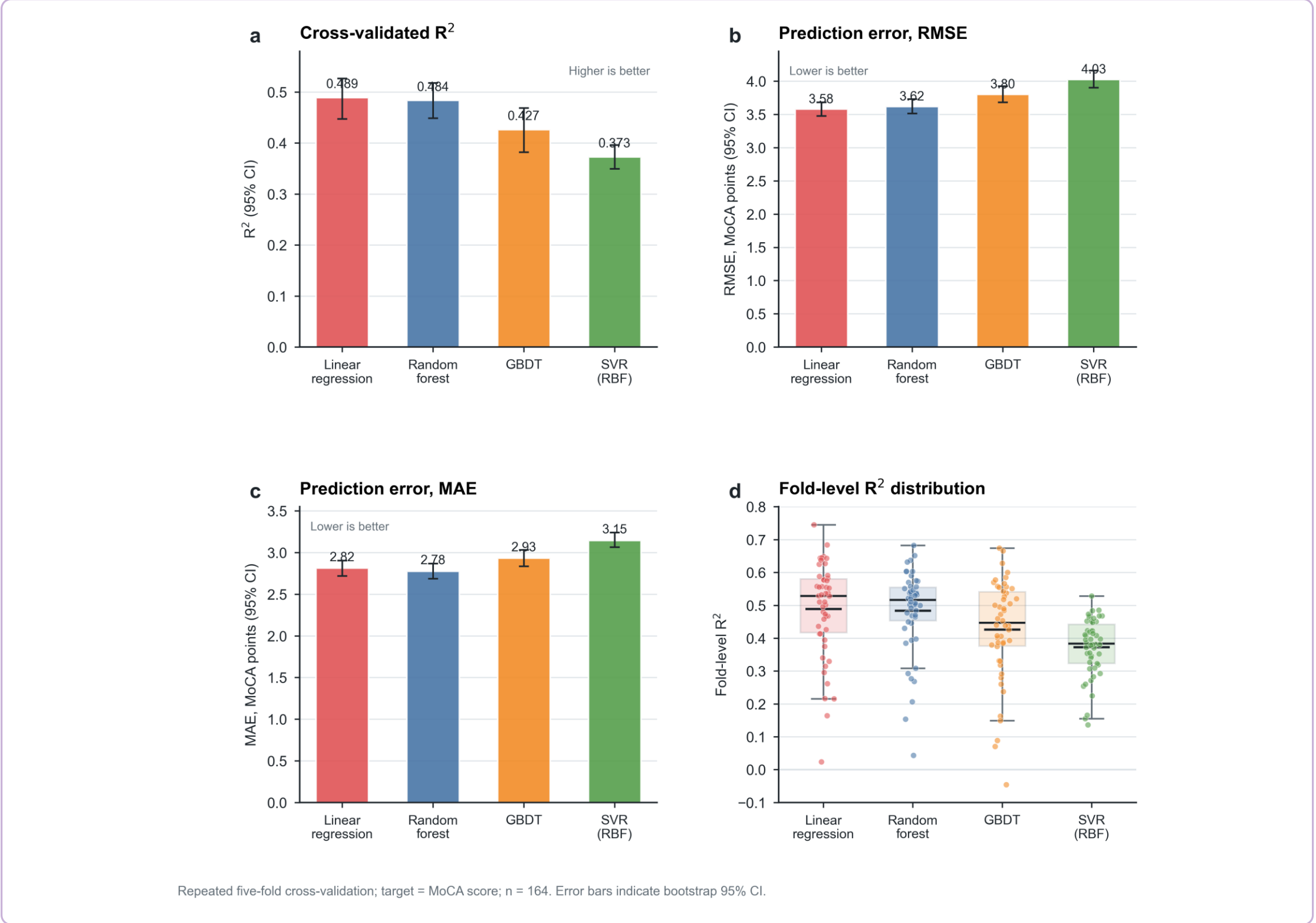
Model	AUROC	Acc.	F1
AdaBoost	0.893	0.808	0.786
Random forest	0.859	0.767	0.759
SVM (RBF)	0.823	0.757	0.742
Decision tree	0.780	0.740	0.732

DOMAIN CONTRIBUTIONS



Full model: Gait + speech-task + demographics achieved the highest discrimination. Speech-task signal: Serial-subtraction measures outperformed gait-only features in behavioral-domain comparison. Feature stability: Education, first-response time, correct responses/sec, stance time, response count, age and gait asymmetry contributed repeatedly. Interpretive boundary: ADualNet is a multimodal screening framework, not an education-independent digital biomarker.

CONTINUOUS COGNITION



Continuous MoCA prediction showed moderate out-of-fold association; linear regression and random forest had the lowest errors.

CONCLUSIONS

Accessible assessment: Ordinary video and audio recordings can support automated TUG-DT feature extraction. Clinical use: The framework may help triage individuals for further cognitive and biomarker evaluation. Next steps: External validation, biomarker comparison and task calibration are required.