Machine learning approaches for early detection of Cognitive decline: A comparative study using the OASIS-3 dataset

Sofiia A. MIKHAILOVA [1]
E-mail: *sofikat392@gmail.com*
ORCID: *https://orcid.org/0009-0006-4795-1198*

Olena A. MIKHAILOVA [2]
E-mail: *olena.mikhailova@estudiante.uam.es*
ORCID: *https://orcid.org/0009-0002-1881-3512*

[1)] *Ivan Franko National University of Lviv, Ukraine*

[2)] *Universidad Autónoma de Madrid, Spain*

Received March 31, 2026, in the final form July 14, 2026
https://doi.org/10.3842/kau.2026.math.04

Abstract. Cognitive decline associated with Alzheimer's disease (AD) and mild cognitive impairment (MCI) represents one of the most critical public health challenges of the twenty-first century, affecting tens of millions of people worldwide. This paper presents a comparative study of machine learning algorithms applied to early detection of cognitive decline using the open access series of imaging studies (OASIS-3) dataset, a longitudinal multimodal neuroimaging and clinical cohort of 1378 participants aged 42–95 years. To avoid data leakage, all features directly encoding a clinical diagnosis, including clinical dementia rating (CDR) scores, diagnostic (DX) codes, and mini-mental state examination (MMSE) results, are excluded; instead, fifteen structural neuroimaging features describing brain morphology (e.g., hippocampal and entorhinal measurements) and neuropsychological sub-test features are selected via a two-stage recursive feature elimination and greedy forward-search procedure. We evaluate six supervised models: logistic regression, support vector machine (SVM), random forest, gradient boosting, K-nearest neighbours (KNN), and CatBoost. Gradient boosting achieves the highest weighted F1-score (0.856) on the held-out test set, followed by CatBoost (0.8482); CatBoost offers approximately 2× lower inference latency. A neurosymbolic post-processing stage using the Scallop probabilistic logic engine identifies 166 participants with measurable structural brain atrophy (reduction in hippocampal volume) prior to a formal clinical diagnosis, demonstrating the added value of logic-augmented classifiers for early risk stratification. Feature-importance analysis identifies white-matter hyperintensity volume (MRI-visible lesions associated with vascular and neurodegenerative damage) and entorhinal/hippocampal measurements as the dominant predictors, consistent with established neurobiological evidence.

Keywords: cognitive decline; Alzheimer's disease; machine learning; OASIS-3 dataset; early detection; preprocessing optimisation; random forest; scikit-learn.

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterised by the gradual deterioration of memory, language, executive function, and the ability to perform daily activities [15]. According to the World Health Organisation, approximately 55 million people worldwide live with dementia, a number projected to rise to 139 million by 2050 as global populations age [15]. AD accounts for 60–70% of all dementia cases, making it the most prevalent form of dementia and the seventh leading cause of death globally [4].

A central challenge in managing AD is the absence of curative treatments; current pharmacological therapies can only slow disease progression if administered at the earliest stages. This reinforces the critical importance of *early detection*: identifying individuals in the pre-dementia phase, known as mild cognitive impairment (MCI), before irreversible neuronal damage has occurred. MCI is defined as a state of cognitive performance below what is expected for a person's age and education level, yet without significant impact on daily functioning, and is considered a prodromal stage of AD [11].

Traditional diagnostic pathways rely on neuropsychological testing and costly neuroimaging; they are resource-intensive and inaccessible in many primary-care settings [8]. Machine learning (ML) offers a scalable alternative: by learning non-linear patterns from large clinical and neuroimaging datasets, ML models integrate heterogeneous modalities to produce probabilistic predictions of cognitive status [2, 14], with several studies reporting accuracy approaching that of clinical experts [6, 7]. Computational efficiency is an additional consideration: inefficient preprocessing pipelines can become a bottleneck in population-scale screening, and optimising imputation, scaling, and resampling steps can substantially reduce wall-clock time without loss of model quality [9].

The present study focuses on the OASIS-3 dataset, a large longitudinal collection released by the Knight Alzheimer Disease Research Centre at Washington University in St Louis, providing rich multimodal data for 1378 participants [8]. We aim to: (1) benchmark classical ML models on a standardised clinical and volumetric feature set derived from OASIS-3; (2) investigate the effect of preprocessing pipeline design on both speed and predictive performance; and (3) identify the most discriminative predictors of cognitive decline through feature-importance analysis.

2 Literature review

Research on automated detection of cognitive decline and Alzheimer's disease has grown substantially over the past decade, driven by the availability of open-access neuroimaging datasets and advances in machine learning methodology.

Classical machine learning on clinical data

A significant body of work demonstrates the utility of classical machine learning methods on structured clinical data. Kavitha et al. [7] reported 84% accuracy with logistic regression on the OASIS dataset using demographic variables such as age, education, socioeconomic status (SES), and MMSE scores. MMSE is a brief clinician-administered cognitive screening test that evaluates memory, attention, language, and orientation. Their results highlighted the importance of feature selection and class balancing for improving predictive performance.

Dhakal et al. [3] applied nine machine learning classifiers together with LASSO and Chi-squared feature-selection techniques, achieving 96.77% accuracy with a Support Vector Machine (SVM). Bari Antor et al. [1] similarly found SVM to be the best-performing model, reaching 92% accuracy. These studies suggest that carefully tuned classical machine learning methods remain highly competitive for cognitive-decline prediction tasks.

Machine learning on multimodal OASIS-3 data

Several recent studies have combined clinical information with neuroimaging biomarkers. Islam et al. [6] evaluated nine classifiers on OASIS-3 clinical data together with Freesurfer-derived neuroimaging measurements. Freesurfer is an automated software framework that extracts volumetric and cortical-thickness biomarkers from structural magnetic resonance imaging (MRI) scans. Their random forest model achieved 98.81% accuracy in 10-fold cross-validation. Using SHAP

(shapley additive explanations) analysis, the authors identified CDR scores, a clinician-rated measure of cognitive impairment severity, and hippocampal volume as the strongest predictors of cognitive decline.

Tam et al. [14] focused on prognosis rather than diagnosis, training models on 628 cognitively unimpaired individuals and validating them on an independent cohort of 128 OASIS-3 participants. Their best model achieved an area under the ROC curve (AUC) of 79% for predicting cognitive decline within 48 months, with hippocampal and caudate measurements contributing most strongly to model performance.

The results reported by Islam et al. [6] are particularly relevant to the present study because they were obtained on the same OASIS-3 dataset. However, their model used variables such as CDR scores that directly encode diagnostic information. In contrast, the present work explicitly removes all diagnosis-related variables to prevent data leakage and provide a more realistic assessment of predictive performance.

Deep learning and preprocessing

Deep-learning approaches have also shown strong performance on neuroimaging data. Convolutional neural network (CNN)-based models achieved accuracies of up to 99.43% on Alzheimer's disease classification tasks [5], while three-dimensional CNN architectures have demonstrated promising results on OASIS-2 [13]. However, such approaches typically require large datasets, specialised graphics processing unit (GPU) hardware, and often provide limited interpretability compared with classical machine learning models [10].

Preprocessing choices remain an important but sometimes overlooked factor in predictive modelling. Pipeline components such as missing-value imputation, feature scaling, and synthetic minority oversampling technique (SMOTE) balancing can differ by an order of magnitude in runtime while producing only minor differences in predictive accuracy [9]. Across multiple studies, CDR, MMSE, hippocampal volume, and the APOE genotype have consistently emerged as some of the most informative predictors of cognitive decline [6, 12, 14].

3 Dataset and feature engineering

OASIS-3 dataset

The open access series of imaging studies (OASIS-3) is a retrospective compilation of longitudinal, multimodal data released by the Knight Alzheimer Disease Research Centre at Washington University in St Louis [8]. The dataset encompasses 1378 participants (755 cognitively normal; 622 at various stages of cognitive decline) aged 42–95 years, collected across multiple ongoing studies over a period of 30 years. For each participant, the dataset includes:

- **Clinical data:** longitudinal assessments using the uniform data set (UDS), a standardised protocol widely used in dementia research. These assessments include the CDR, the MMSE, as well as functional activity scores describing everyday functioning.
- **Neuroimaging:** 2842 MRI sessions comprising T1- and T2-weighted structural magnetic resonance imaging (MRI), which provide detailed anatomical information about brain structures. Additional modalities include fluid-attenuated inversion recovery (FLAIR) sequences used to identify white-matter lesions, arterial spin labeling (ASL) for estimation of cerebral blood flow, blood oxygen level dependent functional MRI (BOLD fMRI) for assessment of brain activity, and diffusion tensor imaging (DTI), which measures water diffusion in neural tissue and is commonly used to evaluate white-matter integrity. Many sessions are accompanied by Freesurfer-derived volumetric segmentations that quantify the size and thickness of specific brain regions.

- **PET imaging:** over 2157 positron emission tomography (PET) scans acquired using Pittsburgh Compound B (PIB) and Florbetapir (AV45) tracers for assessment of amyloid-beta accumulation, a hallmark of Alzheimer's disease pathology, as well as Fluorodeoxyglucose (FDG) PET for measuring cerebral glucose metabolism, an indicator of neuronal activity.
- **Biomarker and genetic data:** cerebrospinal fluid (CSF) biomarkers associated with neurodegeneration and Apolipoprotein E (APOE) genotyping, where the APOE-ε4 allele is a well-established genetic risk factor for Alzheimer's disease, are available for a subset of participants.

All participant identifiers were replaced with random identifiers, and all dates were normalised to reflect days from study entry, ensuring full de-identification in compliance with data protection standards.

Feature selection

For the present comparative ML study, we use structured neuroimaging and cognitive features derived from the Freesurfer volumetric segmentations and the ADRC neuropsychological battery. All features directly encoding a clinical diagnosis (CDR scores, DX codes, MMSE, and CDR sub-domain ratings) are explicitly excluded to prevent *data leakage*: the classification target is itself derived from CDR, so including CDR-based predictors would inflate performance estimates artificially.

The initial candidate pool is ranked by random forest mean decrease in impurity on the full cleaned cohort. Twelve features are then selected via recursive feature elimination (RFE) using each model's internal weights; three additional features are found by a greedy forward-search that maximises 5-fold weighted F1. The final feature set comprises fifteen variables:

(1) left entorhinal cortex thickness (`lh_entorhinal_thickness`),
(2) right entorhinal cortex volume (`rh_entorhinal_volume`),
(3) white-matter hyperintensity volume (`wm-hypointensities_volume`),
(4) left hippocampal volume (`left-hippocampus_volume`),
(5) right hippocampal volume (`right-hippocampus_volume`),
(6) left amygdala volume (`left-amygdala_volume`),
(7) right amygdala volume (`right-amygdala_volume`),
(8) cerebrospinal fluid volume (`csf_volume`),
(9) semantic fluency – animal naming (`animals`),
(10) right insular cortex thickness (`rh_insula_thickness`),
(11) left inferior temporal cortex thickness (`lh_inferiortemporal_thickness`),
(12) right entorhinal cortex thickness (`rh_entorhinal_thickness`),
(13) vegetable naming score (`veg`),
(14) associative memory score (`asscmem`),
(15) digit symbol substitution score (`digsym`).

Most neuroimaging variables represent volumetric or cortical-thickness measurements extracted from structural MRI scans. In particular, hippocampal, entorhinal, and amygdala measurements describe brain regions known to undergo atrophy during the progression of Alzheimer's disease. White-matter hyperintensity volume reflects structural abnormalities visible on MRI and is commonly associated with vascular and neurodegenerative pathology.

The cognitive variables (`animals`, `veg`, `asscmem`, and `digsym`) are neuropsychological test scores assessing semantic fluency, associative memory, and processing speed. These cognitive domains are frequently affected during the early stages of cognitive decline and Alzheimer's disease.

Two composite features are derived for use in the symbolic reasoning stage: `total_hippo` (sum of left and right hippocampal volumes) and `cog_score` (sum of `veg`, `digsym`, and `asscmem`).

The binary classification target is defined as `0 = cognitively normal` (CDR $= 0$) versus `1 = cognitive impairment` (CDR > 0), yielding a cohort of 1281 participants with a 73.6 / 26.4 class split.

Preprocessing pipeline

Raw OASIS-3 data is distributed across multiple CSV tables that are merged on participant identifier. For each participant, the single most recent visit containing both a CDR assessment and Freesurfer-derived volumetric segmentation is retained, yielding the 1281-participant cross-sectional cohort. The preprocessing pipeline applies the following steps in sequence:

1. **Leakage removal:** all columns whose names match the patterns *cdr*, *dx*, *mmse*, *moca*, *srt*, *orient*, *memory*, *judgment*, *psy*, and related clinical-interview fields are dropped.
2. **Missing value imputation:** median imputation is applied to all remaining continuous variables via `SimpleImputer`.
3. **Feature ranking:** a preliminary random forest ($n = 100$ trees) ranks candidates by mean decrease in impurity; the top 40 biomarker columns plus four demographic columns are retained as the high-quality candidate pool.
4. **RFE selection:** twelve features per model from the pool.
5. **Greedy forward search:** three additional features are identified by a greedy forward procedure that maximises 5-fold weighted F1, giving the final 15-feature set.
6. **Feature scaling:** `StandardScaler` (zero mean, unit variance) is fitted on the training split only and applied to both splits.
7. **Class balancing:** the Synthetic minority oversampling technique (SMOTE) is applied exclusively to the training split (74% normal, 26% impaired).
8. **Train/test split:** 80% training, 20% held-out test, stratified by class.

4 Machine learning methods

We evaluate six supervised learning algorithms spanning classical and modern gradient-boosted families. The five classical models use the `scikit-learn` library [9] in Python 3.10; CatBoost uses the `catboost` library. For each classical model, feature selection is performed with RFE (12 features); the final 15-feature set is used for CatBoost. Class balancing is applied via SMOTE on the training split. A neurosymbolic post-processing stage using the Scallop probabilistic logic engine is also evaluated. Inference latency per patient is measured with `time.perf_counter`.

Logistic regression (LR)

Logistic regression models the log-odds of the target class as a linear combination of input features, serving as a strong interpretable baseline. Regularisation parameter C is tuned over $\{0.01, 0.1, 1, 10\}$.

Support vector machine (SVM)

SVM seeks the maximum-margin separating hyperplane in feature space, aiming to maximise the distance between classes and improve generalisation. We evaluate a linear kernel with $C \in \{0.1, 1, 10\}$ and an RBF kernel with $\gamma \in \{\text{scale}, 0.01, 0.1\}$.

Random forest (RF)

Random forest is an ensemble of decision trees trained on bootstrap samples, with majority-vote predictions. We tune the number of trees $n \in \{100, 200, 300\}$ and maximum depth $d \in \{5, 10, \text{None}\}$. Random forest also provides feature-importance estimates that are used in the preliminary ranking stage of this study.

Gradient boosting (GB)

Gradient boosting constructs a sequential ensemble of shallow trees, each correcting residual errors of the current ensemble. `GradientBoostingClassifier` is used with default hyperparameters ($\eta = 0.1$, 100 estimators, max depth 3).

K-nearest neighbours (KNN)

KNN classifies by majority vote among k nearest neighbours. The top 10 features from the global importance ranking are used (RFE is not applicable as KNN has no internal feature weights). We tune $k \in \{3, 5, 7, 9, 11\}$ with uniform voting.

CatBoost

CatBoost [16] builds symmetric (oblivious) trees with ordered boosting to reduce overfitting and natively handles categorical features. We train with 500 iterations, learning rate 0.01, depth 6, and early stopping (patience 50) on the validation set. CatBoost was selected because gradient-boosted tree ensembles have demonstrated strong performance on structured biomedical datasets while maintaining relatively low inference latency.

Neurosymbolic pipeline (CatBoost + Scallop)

To augment the ML-only decision with explicit biological reasoning, the CatBoost output probability is combined with a probabilistic logic programme written in Scallop [17], a Datalog-based neurosymbolic engine supporting differentiable provenance. Three binary evidence signals are derived from the feature set:

- **Hippocampal atrophy risk** – inverted normalised `total_hippo`;
- **Cognitive decline signal** – inverted normalised `cog_score`;
- **ML prediction** – CatBoost posterior probability.

The logic programme fires `final_diagnosis(i, "Decline")` when the ML probability and at least one structural signal exceed 0.5. This adds an interpretable symbolic layer that can identify participants at increased risk of cognitive decline based on structural brain changes, even before a formal diagnosis is assigned.

Evaluation metrics

Model performance is evaluated on the held-out 20% test set using: accuracy, weighted F1-score (to account for class imbalance), and area under the ROC curve area under the receiver operating characteristic curve (AUC). Inference latency (milliseconds per patient) is measured for CatBoost and the ensemble to quantify deployment efficiency.

5 Results

Model performance

Table 1 summarises weighted F1-score for all six models evaluated on the 257-sample held-out test set (1281-participant cohort, 80/20 stratified split, SMOTE applied to training only).

Table 1. Model performance on the OASIS-3 test set ($n = 257$; leakage-free 15-feature set; SMOTE on training split only).

Model	Wtd. F1	Accuracy	AUC-ROC	Latency (ms)
Gradient Boosting	**0.8564**	**0.86**	**0.879**	0.017
CatBoost	0.8482	0.85	—	**0.009**
SVM (Linear)	0.8240	—	—	—
Random Forest	0.8218	—	—	—
Logistic Regression	0.8184	—	—	—
KNN ($k = 5$)	0.7344	—	—	—
Neurosymbolic (CB + Scallop)	0.8104	0.81	—	0.016

Table 2 presents the full per-class classification report for the two best individual models on the held-out test set.

Table 2. Per-class precision, recall, and F1-score for gradient boosting (GB) and CatBoost (CB) on the 257-sample test set.

Model	Class	Precision	Recall	F1	Support
Gradient Boosting	Normal (0)	0.90	0.90	0.90	189
	Impaired (1)	0.72	0.74	0.73	68
	Weighted avg	0.86	0.86	0.86	257
CatBoost	Normal (0)	0.87	0.94	0.90	189
	Impaired (1)	0.77	0.60	0.68	68
	Weighted avg	0.85	0.85	0.85	257

Gradient boosting achieves the highest weighted F1 (0.8564) and AUC (0.879) among all six models. CatBoost attains comparable overall accuracy (0.85) with notably higher normal-class recall (0.94 vs. 0.90), but lower impaired-class recall (0.60 vs. 0.74), reflecting a more conservative classification boundary. Both models outperform SVM, random forest, and logistic regression by a margin of at least 0.016 weighted F1.

The performance obtained in this study is lower than the 98.81% accuracy reported by Islam et al. [6] on the same OASIS-3 dataset. However, their analysis included variables such as CDR scores, which directly encode diagnostic information and substantially simplify the classification task. In contrast, the present study explicitly removes all diagnosis-related variables to prevent

data leakage and provide a more realistic estimate of predictive performance. Despite this restriction, the proposed models achieve competitive results while relying only on neuroimaging and cognitive features.

Feature importance

The top predictors identified by the gradient boosting model after leakage removal are white-matter hyperintensity volume, entorhinal cortex measurements (bilateral thickness and right volume), and bilateral hippocampal and amygdala volumes. Semantic fluency sub-tests (vegetable naming, animal naming) and the digit-symbol substitution score also rank highly. These findings are consistent with established neurobiological evidence linking medial temporal lobe atrophy and white-matter damage to early cognitive decline [6, 14].

Inference speed

Table 3 benchmarks inference latency for the three deployment-relevant architectures measured on the 257-patient test batch with `time.perf_counter`.

Table 3. Inference latency benchmark on the 257-patient test set.

Architecture	Latency (ms/patient)	Speedup vs GB	Weighted F1
CatBoost (symmetric trees)	**0.0085**	2.0×	0.8482
Legacy GB (asymmetric trees)	0.0170	1.0× (baseline)	0.8564
Ensemble (CB + RF voting)	0.0411	0.4×	0.8167
Scallop reasoning (full batch)	0.006 s for 257 patients		—

CatBoost achieves a mean inference latency of 0.0085 ms per patient, approximately **2× faster** than the legacy `GradientBoostingClassifier` (0.017 ms). The neurosymbolic Scallop reasoning stage processes the full 257-patient test batch in 0.006 s, contributing negligible overhead.

Neurosymbolic early-risk detection

The Scallop-based pipeline identified 166 participants clinically labelled normal (CDR = 0) but exhibiting measurable hippocampal atrophy (normalised volume loss > 0.0002 per intracranial volume unit) across longitudinal sessions. These individuals may represent potentially high-risk cases whose structural brain changes appear before a formal clinical diagnosis of cognitive decline. Table 4 reports representative cases from both the pre-clinical atrophy scan and the full probabilistic pipeline.

6 Conclusions

This paper presented a comparative study of six machine learning classifiers for early detection of cognitive decline using the OASIS-3 dataset, with explicit elimination of leaky clinical features and a focus on structural neuroimaging biomarkers. The pipeline integrates 15 features selected via a two-stage procedure (RFE + greedy forward search) within a leakage-free preprocessing workflow applying median imputation, z-score scaling, and SMOTE.

Gradient boosting achieved the highest weighted F1 (0.8564) on the held-out test set, followed by CatBoost (0.8482), SVM (0.824), and Random Forest (0.822). CatBoost demonstrated approximately 2× lower inference latency than the legacy GB implementation. Feature-importance analysis confirmed white-matter hyperintensities and entorhinal/hippocampal volumes as the

Table 4. Representative Scallop-flagged participants. *Top*: clinically normal participants (CDR = 0) with the highest observed hippocampal atrophy rates across longitudinal visits. *Bottom*: highest probabilistic diagnoses from the full neurosymbolic pipeline. Follow-up diagnostic outcomes were not available for all identified participants and therefore are not reported.

Rank	ID	Hippo. atrophy rate	Confidence	Verdict
1	OAS30253	−0.001280	1.000	Pre-clinical Risk
2	OAS30083	−0.001229	1.000	Pre-clinical Risk
3	OAS31354	−0.000976	1.000	Pre-clinical Risk
4	OAS30112	−0.000734	1.000	Pre-clinical Risk
5	OAS30241	−0.000647	1.000	Pre-clinical Risk

Rank	ID	Confidence	Reasoning pathway
1	OAS30031	0.9900	Cognitive Decline
2	OAS30043	0.9870	Cognitive Decline; Structural Atrophy
3	OAS31214	0.9820	Cognitive Decline
4	OAS30682	0.9774	Cognitive Decline; Vascular (WM) Damage
5	OAS30472	0.9596	Cognitive Decline

dominant predictors, consistent with established neurobiological evidence [6, 14]. A neurosymbolic post-processing stage using Scallop identified 166 clinically normal participants with measurable structural atrophy, illustrating the potential of logic-augmented classifiers for early risk stratification.

Future work includes: (1) external validation on the Alzheimer's disease neuroimaging initiative (ADNI) and National Alzheimer's Coordinating Center (NACC) cohorts, which contain larger and more diverse participant populations and would allow assessment of model generalisability across independent datasets; (2) survival analysis formulations to predict time-to-conversion rather than binary cognitive status [12]; (3) end-to-end neurosymbolic training that jointly optimises the ML perceiver and the Scallop reasoner; and (4) systematic evaluation of model fairness across demographic subgroups.

Acknowledgements

The authors sincerely thank their supervisor Vitaly Vlasov (vitaly.vlasov@lnu.edu.ua) for invaluable guidance, insightful feedback, and constant encouragement throughout the research and writing process. The authors would like to thank the anonymous referee for careful reading and helpful comments.

References

[1] Bari Antor M., Jamil A.H.M.S., Mamtaz M., Monirujjaman Khan M., Aljahdali S., Kaur M., Singh P., Masud M., A comparative analysis of machine learning algorithms to predict Alzheimer's disease, *J. Healthcare Eng.* **2021** (2021), article no. 9917919, 12 pages.

[2] Battista P., Salvatore C., Berlingeri M., Cerasa A., Castiglioni I., Artificial intelligence and neuropsychological measures: the case of Alzheimer's disease, *Front. Aging Neurosci.* **12** (2020), article no. 134, 13 pages.

[3] Dhakal S., Azam S., Hasib K.M., Karim A., Jonkman M., Farhan Al Haque A.S.M., Dementia prediction using machine learning, *Procedia Computer Sci.* **219** (2023), 1297–1308.

[4] Ahmad F.B., Anderson R.N., The leading causes of death in the US for 2020, *MMWR Morbidity Mortality Weekly Rep.* **70** (2021), no. 17, 643–644.

[5] El-Assy A., Amer H.M., Ibrahim H., Mohamed M., A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data, *Sci. Rep.* **14** (2024), article no. 3463, 19 pages.

[6] Islam J., Zhang Y., Islam M.R., Islam M.N., Explainable AI-based Alzheimer's prediction and management using multimodal data, *Plos One* **18** (2023), no. 11, article no. e0290877, 23 pages.

[7] Kavitha C., Mani V., Srividhya S., Khalaf O.I., Tavera Romero C.A., Early-stage Alzheimer's disease prediction using machine learning models, *Frontiers in Public Health* **10** (2022), article no. 853294, 13 pages.

[8] LaMontagne P.J., Benzinger T.L.S., Morris J.C., Keefe S., Hornbeck R., Xiong C., Grant E., Hassenstab J., Moulder K., Vlassenko A.G., Raichle M.E., Cruchaga C., Marcus D., OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease, *medRxiv* (2019).

[9] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **12** (2011), 2825–2830, arXiv:1201.0490.

[10] Pellegrini E., Ballerini L., Hernandez M.C.V., Chappell F.M., González-Castro V., Anblagan D., Danso S., Muñoz-Maniega S., Job D., Pernet C., Mair G., MacGillivray T.J., Trucco E., Wardlaw J.M., Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review, *Front. Neurosci.* **12** (2018), article no. 786, 11 pages.

[11] Petersen R.C., Mild cognitive impairment as a diagnostic entity, *J. Internal Med.* **256** (2004), no. 3, 183–194.

[12] Saeed A., Waris A., Fuwad A., Iqbal J., Khan J., AlQahtani D., Abutaleb A., Random survival forest model for early prediction of Alzheimer's disease conversion in early and late mild cognitive impairment stages, *Plos One* **19** (2024), no. 12, article no. e0314725, 21 pages.

[13] Saratxaga C.L., Moya I., Picón A., Acosta M., Moreno-Fernandez-de-Leceta A., Garrote E., Bereciartua-Pérez A., MRI deep learning-based solution for Alzheimer's disease prediction, *Diagnostics* **12** (2022), no. 4, article no. 928, 13 pages.

[14] Tam A., Dansereau C., Iturria-Medina Y., Urchs S., Orban P., Sharmarke H., Breitner J., Bellec P., Alzheimer's Disease Neuroimaging Initiative, Prediction of cognitive decline for enrichment of Alzheimer's disease clinical trials, *J. Prevention Alzheimer's Disease* **9** (2022), 400–409.

[15] World Health Organization, *Dementia: Key facts*, 2023, available at `https://www.who.int/news-room/fact-sheets/detail/dementia`.

[16] Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A., CatBoost: unbiased boosting with categorical features, *Adv. Neural Inf. Process. Syst.* **31** (2018), 11 pages, arXiv:1706.09516.

[17] Li Z., Huang X., Naik A., Zhai Z., Shi W., Raghothaman M., Baral C., Xu J., Scallop: A language for neurosymbolic programming, *Proc. ACM Program. Lang.* **7** (2023), PLDI, 1177–1201, arXiv:2304.09812.

Підходи машинного навчання для раннього виявлення когнітивного спаду: порівняльне дослідження на основі датасету OASIS-3

Софія А. МІХАЙЛОВА [1]
E-mail: *sofikat392@gmail.com*
ORCID: *https://orcid.org/0009-0006-4795-1198*

Олена А. МІХАЙЛОВА [2]
E-mail: *olena.mikhailova@estudiante.uam.es*
ORCID: *https://orcid.org/0009-0002-1881-3512*

[1)] *Львівський національний університет імені Івана Франка, Україна*

[2)] *Автономний університет Мадрида, Іспанія*

Отримано 31 березня 2026, у фінальній формі 14 липня 2026
https://doi.org/10.3842/kau.2026.math.04

Анотація. Когнітивний спад, пов'язаний із хворобою Альцгеймера та легким когнітивним порушенням, є одним із найважливіших викликів сучасної охорони здоров'я. Раннє виявлення осіб із підвищеним ризиком розвитку когнітивних порушень має вирішальне значення для своєчасного клінічного втручання та уповільнення прогресування захворювання. У цій статті представлено порівняльне дослідження алгоритмів машинного навчання для раннього виявлення когнітивного спаду на основі датасету OASIS-3 – поздовжньої мультимодальної нейровізуалізаційної та клінічної вибірки, що охоплює 1378 учасників віком від 42 до 95 років. Для запобігання витоку інформації з аналізу було виключено всі ознаки, що безпосередньо кодують клінічний діагноз, зокрема показники CDR, MMSE та діагностичні коди. Натомість використано п'ятнадцять структурних нейровізуалізаційних та нейропсихологічних ознак, відібраних за допомогою двоетапної процедури відбору ознак, що поєднує recursive feature elimination та жадібний пошук. Досліджено шість моделей класифікації: логістичну регресію, метод опорних векторів, випадковий ліс, градієнтний бустинг, метод k-найближчих сусідів та CatBoost. Найкращий результат на тестовій вибірці продемонстрував градієнтний бустинг із зваженим F1-показником 0,856, тоді як CatBoost забезпечив близько дворазове скорочення часу інференсу. Додатково запропоновано нейросимвольний етап післяобробки на основі логічного рушія Scallop, який дозволив виявити 166 учасників із ознаками структурної атрофії мозку до встановлення формального клінічного діагнозу. Аналіз важливості ознак підтвердив провідну роль об'єму гіперінтенсивностей білої речовини та енториальних і гіпокампальних структур як ключових предикторів когнітивного спаду.

Ключові слова: когнітивний спад; хвороба Альцгеймера; машинне навчання; OASIS-3; нейровізуалізація; раннє виявлення; CatBoost; нейросимвольний штучний інтелект.