

ORIGINAL RESEARCH ARTICLE

An interpretable decision tree model for prioritizing functional and cognitive assessments in early diagnosis of Alzheimer's disease

Wollner Materko^{1,2*} ¹Department in Health Sciences, Federal University of Amapá, Macapá, Brazil²Department of Physical Education, Federal University of Amapá, Macapá, Brazil

Abstract

The extensive clinical implementation of machine learning models for diagnosing Alzheimer's disease (AD) is often hindered by the fact that they are "black box" technologies. This lack of transparency undermines clinician trust. This study addressed this issue by developing and validating a high-performing, interpretable decision tree classifier. This model prioritizes features that are both statistically significant and clinically relevant for the early diagnosis of AD. The study used a publicly available dataset containing demographic, clinical, and cognitive data from 2,149 patients. A strategic feature selection process was implemented. The final model was meticulously trained and thoroughly evaluated using a separate test set. The evaluation revealed the model's exceptional diagnostic capability, demonstrating an overall accuracy of 91.8% and an area under the curve of 0.93. The model also demonstrated high sensitivity (92.1%) and specificity (95.7%). Feature importance analysis revealed that predictions were primarily influenced by direct measures of cognitive and functional status. Specifically, the most significant factors were the mini-mental state examination score, the activities of daily living score, and the Functional Assessment score. These core features overshadowed the predictive influence of established risk factors such as family history. Despite their strong statistical association with the disease, the model assigned negligible importance to these risk factors in its final classification. This work demonstrates that an interpretable decision tree can accurately diagnose AD by transparently prioritizing functional and cognitive assessments. The model's clarity and reliance on routine clinical data highlight its potential as a decision-support tool. It is ready to improve diagnostic pathways and promote further research in this area.

***Corresponding author:**Wollner Materko
(wollner.materko@gmail.com)

Citation: Materko W. An interpretable decision tree model for prioritizing functional and cognitive assessments in early diagnosis of Alzheimer's disease. *Artif Intell Health*. 2026;3(3):025330066. doi: 10.36922/AIH025330066

Received: August 14, 2025**Revised:** September 11, 2025**Accepted:** September 24, 2025**Published online:** October 10, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Alzheimer's disease; Machine learning; Decision tree; Cognitive assessment

1. Introduction

Alzheimer's disease (AD) is a relentless neurodegenerative disorder and the leading cause of dementia worldwide. It progressively erodes cognitive function and memory. Given the demographic shift toward an aging global population, the increasing prevalence of AD and its profound societal impact underscore the urgent need to deepen our understanding of its complex biology, clinical evolution, and formidable diagnostic

challenges.¹⁻³ Conventionally, the clinical diagnosis of AD has relied on a combination of medical history, cognitive assessments, and neuroimaging. However, a definitive diagnosis is only possible through post-mortem neuropathological examination.⁴⁻⁶ Due to the gradual onset of the disease and progressive cognitive decline, there is a critical need for accessible, accurate tools to facilitate early diagnosis. This is essential for timely intervention, care planning, and patient enrollment in clinical trials.^{4,7}

1.1. Fundamentals of AD

At its pathological core, AD is characterized by the insidious accumulation of insoluble beta-amyloid (A β) plaques and hyperphosphorylated tau proteins, which aggregate into neurofibrillary tangles within the brain parenchyma. This dual proteopathic burden results in widespread synaptic dysfunction, neuronal loss, and subsequent cerebral atrophy, which are the hallmarks of neurodegeneration.⁸⁻¹¹ While sporadic late-onset AD constitutes the vast majority of cases, specific genetic predispositions are recognized. Early-onset familial AD is unequivocally linked to autosomal dominant mutations in the amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*), and presenilin 2 (*PSEN2*) genes. For late-onset AD, the *APOE4* allele is the most significant genetic risk factor, and a growing number of additional susceptibility loci have been identified through large-scale genome-wide association studies.^{12,13} Beyond genetics, a variety of modifiable and non-modifiable risk factors contribute to AD pathogenesis, including advancing age, cardiovascular commodities, diabetes, and various lifestyle factors.^{1,12,14}

1.2. Evolution and diagnostic challenges of AD

The etiological complexity of AD, which is rooted in various pathological pathways such as the aggregation of amyloid-beta and tau proteins,^{15,16} poses a significant diagnostic challenge. This complexity is further compounded by a wide array of contributing risk factors ranging from genetic predispositions to lifestyle choices.¹⁷ Notably, there is a significant overlap between the risk factors for AD and atherosclerotic cardiovascular disease, suggesting a strong link between vascular health and neurodegeneration.^{18,19} The multifactorial nature of AD, involving numerous interacting variables in non-linear ways, makes it an ideal domain for machine learning applications. By analyzing these heterogeneous data sources simultaneously, machine learning models can help determine the relative importance of different risk profiles, identify patterns to inform personalized prevention strategies,²⁰ and ultimately advance our understanding of and ability to manage dementia in its various forms.²¹⁻²³

1.3. Machine learning and AD

Machine learning has emerged as a powerful methodology for analyzing the high-dimensional data characteristic of AD, offering the potential to uncover subtle patterns that can inform clinical decision-making in health.²⁴⁻²⁷ While complex models such as deep neural networks often achieve high predictive accuracy, their “black box” nature poses a significant obstacle to their adoption in clinical practice, where transparency and interpretability are paramount. The decision tree algorithm (DTA) offers a compelling alternative. By constructing a model based on a hierarchical series of data-driven rules, the DTA creates an inherently transparent framework that mirrors the logical reasoning process of a clinician and allows for expert validation of its predictions.^{29,29}

Developing an accurate and interpretable machine learning model for AD classification is of profound clinical relevance. Such a tool could function as a decision support system to help clinicians triage patients, focus diagnostic workups on the most important predictive factors, and minimize critical diagnostic errors, such as failing to identify a patient with AD. Reducing these errors is crucial for improving patient outcomes by enabling earlier access to therapeutic and support services.^{3,7}

The aim of this study was to develop, train, and evaluate an interpretable decision tree model for classifying AD, based on comprehensive patient demographic, clinical, and cognitive assessment data. Furthermore, the study aimed to identify the most significant predictive features for an AD diagnosis and elucidate the diagnostic logic learned by the model. This would provide a transparent, clinically relevant predictive tool.

2. State of the art

2.1. DTA

The DTA is a fundamental supervised learning method that is widely used in healthcare to model and elucidate complex decision-making processes.^{28,29} Structurally, a DTA resembles a flowchart: Each internal node represents a test on a feature, each branch signifies an outcome, and each leaf node provides a final classification or prediction.³⁰ This versatility makes DTAs a powerful tool for applications such as disease diagnosis, treatment outcome prediction, and patient risk stratification. Ultimately, they aim to enhance the accuracy of medical assessments and the quality of care.³¹⁻³³

One of the key advantages of DTAs in medicine is their high interpretability. Unlike “black-box” models, their transparent structure enables clinicians to trace and comprehend the reasoning behind a prediction, which is

crucial for ensuring patient safety and fostering trust in clinical settings.³⁴ This clarity is also crucial for achieving responsible artificial intelligence (AI), as it facilitates the auditing of algorithmic fairness. Examining the decision pathways enables practitioners to identify and mitigate potential biases related to patient variables, which is a critical step toward reducing health disparities.^{33,35}

Despite these benefits, DTAs have notable limitations. For example, they are susceptible to overfitting, whereby the model learns the training data too closely and fails to generalize to new, unseen data. Furthermore, they can be unstable, meaning that small variations in the input data can result in significantly different tree structures.^{36,37} These technical vulnerabilities raise significant ethical concerns. For example, if a model is trained on biased data, its transparency can legitimize inequitable decision-making, with severe consequences for patient outcomes.³⁵ To mitigate issues such as overfitting and instability, ensemble methods such as decision forests, which combine the predictions of multiple trees, are often employed to improve predictive performance and robustness.³⁸

Addressing these challenges is a central focus of ongoing research, which explores techniques such as end-to-end learning and methodologies for handling uncertain data, with the aim of building more robust and reliable models.^{38,39} As machine learning continues to be integrated into medicine, the role of DTAs is evolving from that of a simple prediction tool to that of a cornerstone of responsible AI.³⁵ Continued refinement of DTAs has the potential to transform healthcare by translating vast and complex datasets into actionable, interpretable, and equitable insights for clinicians and health systems.³⁸⁻⁴⁰

3. Materials and methods

This section outlines the methodology employed in the development and evaluation of a predictive model for AD. It covers the description of the dataset, feature engineering, the model training protocol, and the metrics used to assess performance. The following schematic diagram illustrates the methodology in question (Figure 1). It provides a clear overview of the key steps, from data acquisition and preprocessing to model training, evaluation, and interpretability analysis.

3.1. Study population

The study utilized the “Alzheimer’s Disease Dataset” publicly available on Kaggle (<https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset>). The original dataset contains extensive health information for 2,149 patients, including demographic details, lifestyle factors, medical history, clinical measurements, and

cognitive assessments. For this analysis, a preprocessed and cleaned subset of all patients was used for model development and evaluation. The patient cohort had an age range of 60–90 years. The target variable was the diagnosis status, a binary feature indicating the presence (1) or absence (0) of AD.

3.2. Statistical analysis

Statistical analyses were conducted to characterize the study cohort and identify significant differences between patients with AD and the control group. Baseline characteristics were summarized using descriptive statistics: the mean $\pm$ standard deviation (SD) for continuous variables and absolute frequencies (n) with relative proportions (%) for categorical variables.

Inferential comparisons were made after first assessing the normality of continuous data within each group using the Shapiro–Wilk test. Group means were compared using an independent samples *t*-test for normally distributed data, while a Mann–Whitney *U* test was applied to data that violated the normality assumption ($p \leq 0.05$). Associations between categorical variables and the AD diagnosis were evaluated using Pearson’s Chi-squared test.

A two-tailed $p < 0.05$ was considered the threshold for statistical significance. All analyses were performed using Python, with the pandas library for data manipulation and the scipy.stats library for statistical testing.

3.3. Feature selection and preprocessing

A total of 15 features were selected for inclusion in the model, using a strategy that balanced statistical significance with clinical relevance in order to create a robust yet parsimonious model. This process was guided by the results of the initial statistical analysis (Table 1) and existing knowledge from the AD literature. These features included: (i) Cognitive and functional assessments: Functional Assessment, mini-mental state examination (MMSE), activities of daily living (ADL), and memory complaints. These were included due to their high statistical significance ($p < 0.001$) and their role as core metrics for diagnosing cognitive and functional decline; (ii) Symptoms: behavioral problems, and difficulty completing tasks. These symptoms were chosen because they are significant predictors of the disease’s impact on daily life. Other symptomatic variables, such as confusion and disorientation, were deliberately excluded to avoid data leakage, as they are almost synonymous with the diagnosis itself; (iii) Demographic and lifestyle factors: Age was included as it is the most significant non-modifiable risk factor ($p < 0.001$). Ethnicity and sleep quality were retained for their clinical relevance in dementia research,

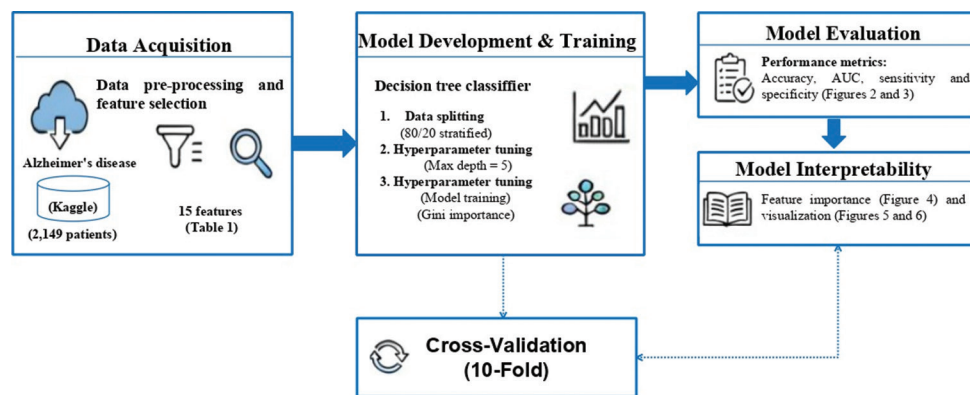


Figure 1. Flow chart depicting methodology of the study
Abbreviation: AUC: Area under the curve.

despite not reaching statistical significance in this specific cohort; (iv) Clinical measurements: Systolic blood pressure (SBP), low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and cholesterol/triglycerides. These were included as they represent key vascular risk factors that are widely implicated in the pathophysiology of AD; (v) Medical history: Family history of AD was included due to its extremely high significance ($p < 0.001$) as a genetic risk factor. Hypertension was retained due to its clinical relevance as a major vascular comorbidity.

Variables that showed no statistical significance or clinical relevance in the context of this study were excluded to reduce model complexity and prevent the introduction of noise. The dataset, with its categorical variables already encoded numerically, was then prepared for input into the machine learning model. As a tree-based model was used, no further feature scaling or normalization was necessary.

3.4. Model development and training

A decision tree classifier was implemented using the scikit-learn Python library to predict AD diagnoses. To prepare the data for modeling, the dataset was partitioned into training and testing subsets, comprising 80% and 20% of the data, respectively. Stratified sampling was used to ensure that the proportion of patients with and without AD remained consistent in both subsets. To create a robust and highly interpretable model, hyperparameter tuning was performed with the key constraint of setting the maximum depth of the tree to 5. This prevents the model from becoming overly complex and ensures that the decision-making process can be easily visualized and understood. Gini importance, also known as Mean Decrease in Impurity, was calculated for each feature. This metric quantifies each feature's contribution to the classification task within the decision tree structure, enabling predictors to be ranked from most to least influential.

3.5. Performance evaluation

The predictive capability of the trained model was rigorously evaluated using unseen test data. This evaluation centered on a confusion matrix, which provides a detailed breakdown of correct and incorrect classifications. In the context of AD diagnosis, the matrix components were defined as follows: (i) true positives (TP) or the number of patients with AD who were correctly diagnosed by the model; (ii) true negatives (TN) or the number of healthy patients who were correctly identified as not having the disease; (iii) false positives (FP) or the number of healthy patients who were incorrectly diagnosed with AD; and (iv) false negatives (FN) or the number of patients with AD who were incorrectly identified as healthy.

The overall proportion of correct predictions (Equation I):

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (I)$$

The proportion of positive predictions that were actually correct (Equation II):

$$Precision = \frac{TP}{(TP + FP)} \quad (II)$$

The proportion of actual positive cases that were correctly identified is known as recall (Equation III):

$$Recall = \frac{TP}{(TP + FN)} \quad (III)$$

The harmonic mean of precision and recall, providing a balanced measure of performance (Equation IV):

$$F1\text{-Score} = 2 * (Precision * Recall) / (Precision + Recall) \quad (IV)$$

Table 1. Baseline characteristics of the study cohort

Characteristic	Overall cohort (<i>n</i> =2149) (%)	Control group (<i>n</i> =1389) (%)	AD group (<i>n</i> =760) (%)	<i>p</i> -value
Demographics				
Age, years	74.9±9.0	70.2±8.3	83.4±5.1	<0.001*
Gender, <i>n</i> (%)				0.845
Female	1088 (50.6)	700 (50.4)	388 (51.1)	
Male	1061 (49.4)	689 (49.6)	372 (48.9)	
Ethnicity, <i>n</i> (%)				0.446
Caucasian	1278 (59.5)	821 (59.1)	457 (60.1)	
African American	454 (21.1)	298 (21.5)	156 (20.5)	
Asian	206 (9.6)	130 (9.4)	76 (10.0)	
Others	211 (9.8)	140 (10.1)	71 (9.3)	
Education level, <i>n</i> (%)				0.513
None	446 (20.8)	285 (20.5)	161 (21.2)	
High school	854 (39.7)	555 (40.0)	299 (39.3)	
Bachelor's	636 (29.6)	406 (29.2)	230 (30.3)	
Higher	213 (9.9)	143 (10.3)	70 (9.2)	
Lifestyle and clinical				
BMI, kg/m ²	27.7±7.2	27.7±7.2	27.6±7.3	0.941
Smoker, <i>n</i> (%)	620 (28.9)	399 (28.7)	221 (29.1)	0.968
Systolic BP, mmHg	134.3±25.9	134.4±26.0	134.1±25.9	0.778
Medical history, <i>n</i> (%)				
Family history of AD	542 (25.2)	9 (0.7)	533 (70.1)	<0.001*
Hypertension	320 (14.9)	207 (14.9)	113 (14.9)	0.863
Diabetes	324 (15.1)	210 (15.1)	114 (15.0)	0.927
Depression	431 (20.1)	278 (20.0)	153 (20.1)	0.950
Cognitive and functional				
MMSE score	14.8±8.6	21.7±4.8	2.7±1.6	<0.001*
Functional assessment score	5.1±2.9	7.0±1.7	1.6±1.2	<0.001*
ADL score	5.0±2.9	6.9±1.8	1.5±1.2	<0.001*
Symptoms, <i>n</i> (%)				
Forgetfulness	648 (30.2)	0 (0.0)	648 (85.3)	<0.001*
Confusion	441 (20.5)	0 (0.0)	441 (58.0)	<0.001*

Notes: Data are presented as mean±SD for continuous variables and *n* (%) for categorical variables. *P* values were calculated using the Mann–Whitney *U* test for continuous variables and the Chi-squared (χ^2) test for categorical variables to compare the control and AD groups. Statistically significant differences are indicated by an asterisk ($p<0.05$).

Abbreviations: AD: Alzheimer's disease; ADL: Activities of Daily Living; BMI: Body mass index; BP: Blood pressure; MMSE: Mini-Mental State Examination.

3.6. Cross-validation

To assess the stability and robustness of the model further and mitigate potential bias from a single training and testing split, a 10-fold cross-validation procedure was performed. The entire dataset was divided into ten equal-sized subsets, known as folds. The model was then trained and validated ten times, with a different fold being used as the test set each time and the remaining nine folds being used for training. The final performance was reported as the mean and SD of the accuracy scores across all ten folds. This approach provides a more reliable estimate of the model's performance when faced with new, unseen data.

4. Results

The final study cohort comprised 2,149 patients. Of these, 760 (35.4%) were diagnosed with AD, and the remaining 1,389 patients (64.6%) formed the non-AD control group. The baseline demographic, clinical, and cognitive characteristics of both groups are detailed in [Table 1](#).

Comparative analysis revealed significant differences in key risk factors between the AD and control groups. Patients with AD were substantially older than their control counterparts (mean age: 83.4 ± 5.1 vs. 70.2 ± 8.3 years; $p<0.001$). Furthermore, a strong association was observed with genetic predisposition: Over 70% of patients with AD had a family history of the disease, compared to < 1% of controls ($p<0.001$). In contrast, a range of other demographic, lifestyle, and clinical factors showed no significant association with an AD diagnosis. These included gender ($p=0.845$), ethnicity ($p=0.446$), and educational attainment ($p=0.513$). Similarly, no significant differences were observed in body mass index, smoking status, SBP, or the prevalence of comorbidities such as hypertension, diabetes, and depression ($p>0.05$).

The most substantial differences between the groups were observed in all measures of cognitive and functional performance. Patients with AD displayed severe cognitive impairment, as evidenced by a mean MMSE score of 2.7 (±1.6), which was drastically lower than the score of 21.7 (±4.8) observed in the control group ($p<0.001$). This functional decline was also evident in daily activities, with the AD group scoring significantly lower on both the Functional Assessment scale (1.6 ± 1.2 vs. 7.0 ± 1.7; $p<0.001$) and the ADL scale (1.5 ± 1.2 vs. 6.9 ± 1.8; $p<0.001$). Analysis of cardinal symptoms revealed near-perfect separation between the groups. Key indicators such as forgetfulness (85.3%), confusion (58.0%), and memory complaints (59.1%) were highly prevalent in the AD group, but absent in the control group.

The decision tree classifier developed in this study performed exceptionally well in the differential diagnosis of AD on a held-out test set. The model's efficacy was rigorously evaluated through a multifaceted assessment, including global performance metrics, error analysis through the confusion matrix, quantification of feature importance, and analysis of its transparent decision-making architecture.

The classifier achieved a high degree of diagnostic accuracy, with an overall accuracy of 91.8%. The model's exceptional discriminatory capability was further evidenced by its performance in the receiver operating characteristic curve analysis, yielding an area under the curve (AUC) value of 0.93 (Figure 2). This value indicates an excellent ability to distinguish between individuals with and without AD.

Analysis of the confusion matrix (Figure 3) showed that the classifier achieved a high degree of diagnostic accuracy on the 430-sample test cohort. The model correctly identified 266 individuals without AD (TN) and 140 AD patients (TP), and misclassified only 24 individuals. This performance corresponds to a specificity of 95.7%, demonstrating the model's robust capacity to correctly identify healthy individuals and minimize unnecessary follow-up investigations. Importantly, the model achieved a sensitivity or recall of 92.1%, indicating its strong ability to detect true cases of AD. The classifier's precision was also high, at 92.1%, which underscores the reliability of a positive diagnosis. This exceptional, balanced performance, characterized by a low number of both FP and FN, highlights the model's significant potential as a reliable tool to support clinical decision-making.

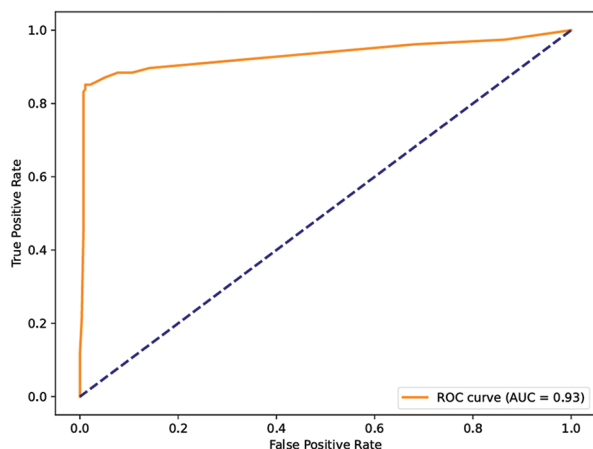


Figure 2. Discriminatory ability of the decision tree classifier on the test set
Abbreviations: AUC: Area under the curve; ROC: Receiver operating characteristic.

To elucidate the hierarchical importance of the predictors driving the model's diagnostic decisions, the Gini importance of each input variable was quantified (Figure 4). The analysis revealed that the model's predictive logic is overwhelmingly governed by a small subset of features that can be directly used to determine cognitive and functional status.

The four most influential predictors were, in descending order: MMSE score, ADL score, Functional Assessment score, and memory complaints. These four variables alone constituted the vast majority of the model's predictive power, establishing them as the core determinants for classification.

A sharp demarcation in importance was observed following this top tier. Behavioral problems represent the only other feature with a discernible, albeit minor,

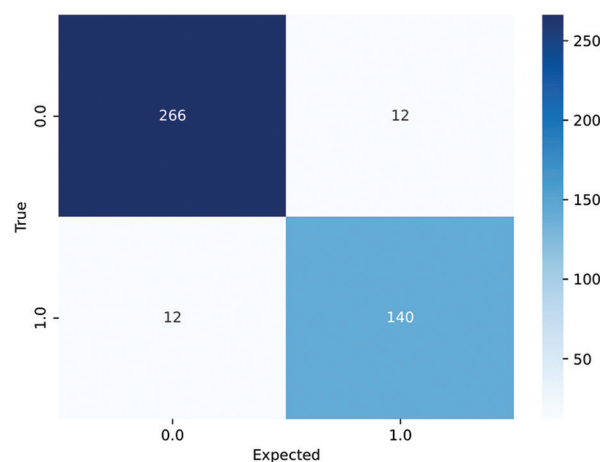


Figure 3. Diagnostic performance of the decision tree classifier on the test set

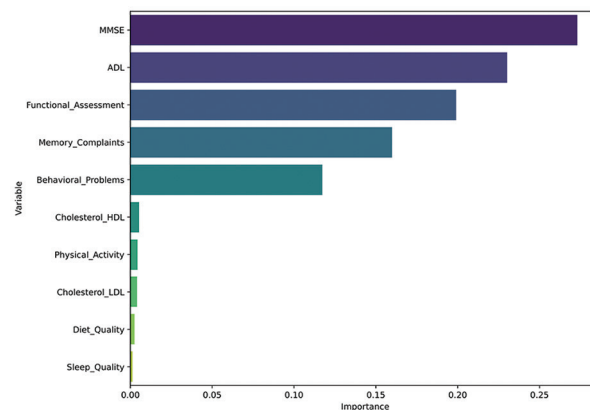


Figure 4. Relative importance of predictor variables in the diagnosis of Alzheimer's disease
Abbreviations: ADL: Activities of Daily Living; BMI: Body mass index; HDL: High-density lipoprotein; LDL: Low-density lipoprotein; MMSE: Mini-Mental State Examination.

contribution. Strikingly, several variables considered established clinical or genetic risk factors, including family history of AD, Hypertension, and all cholesterol metrics, registered a Gini importance of zero. This indicates that, in the presence of the powerful predictive signals from the cognitive and functional assessments, the model did not utilize these other features to achieve its final classification accuracy. This clear hierarchy underscores the primacy of direct, real-time assessments of a patient's cognitive performance and functional capacity in this data-driven diagnostic framework, suggesting that these measures can eclipse the contribution of more traditional, static risk factors in a machine learning context.

A key advantage of the decision tree architecture is its inherent transparency, which allows for a full deconstruction of its diagnostic logic (Figure 5). The model's classification pathways are initiated from a single root node based on Functional Assessment, the most informative feature from the initial statistical analysis.

The tree's primary bifurcation at Functional Assessment score ≤ 4.967 effectively partitions the patient population into distinct risk strata. For instance, a clear diagnostic pathway for a low-risk individual involves having a functional assessment score above this threshold, followed by the absence of memory complaints, which consistently leads to a non-AD classification. Conversely, a patient presenting with impaired functional assessment and compromised ADL (ADL score ≤ 5.04) is channeled through pathways that strongly predict an AD diagnosis.

Figure 6 shows the learning curve, which illustrates the model's performance as the size of the training

set increases. Training accuracy (blue line) remains consistently high, indicating that the model is effectively learning from the data. Validation accuracy (red line), which represents performance on unseen data, generally increases with training set size, suggesting that more data leads to better generalization. However, there is a notable gap between the training and validation accuracy curves initially, which narrows as the training set size increases. This initial discrepancy is normal and suggests that, while the model learns well from the training data, there is some degree of variance or slight overfitting in smaller datasets. However, once the training set size reaches approximately 800–1,000 samples, both curves tend to stabilize. The validation accuracy reaches a plateau of around 0.94–0.95, while the training accuracy stabilizes at around 0.98–0.99. The consistently high validation accuracy, particularly with larger training set sizes, indicates good generalization without severe overfitting. This is further supported by the robust results from the 10-fold cross-validation.

This explicit, rule-based framework not only provides high accuracy but also offers a white-box model whose predictions can be clinically validated and understood, fostering trust and facilitating its potential integration into clinical workflows.

5. Discussion

This study successfully developed and validated a decision tree classifier with exceptional accuracy (91.8%) and discriminatory power (AUC = 0.93) for diagnosing AD. The model's architecture provides a transparent, rule-based framework for its predictions. The following discussion

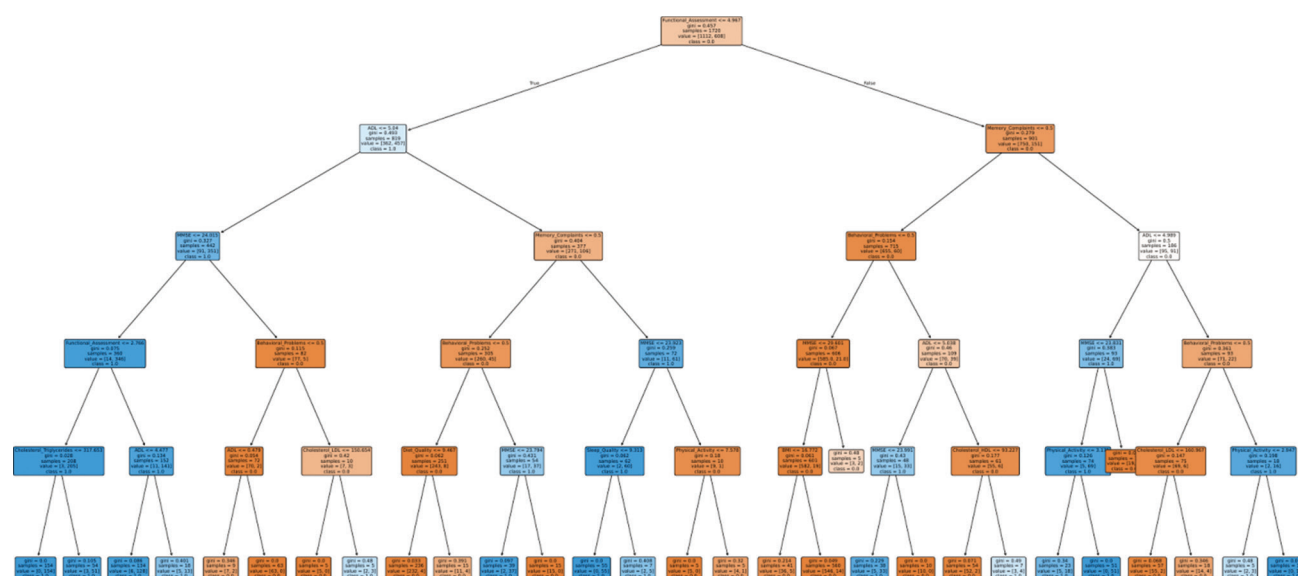


Figure 5. Visualization of the complete rule-based architecture of the trained decision tree

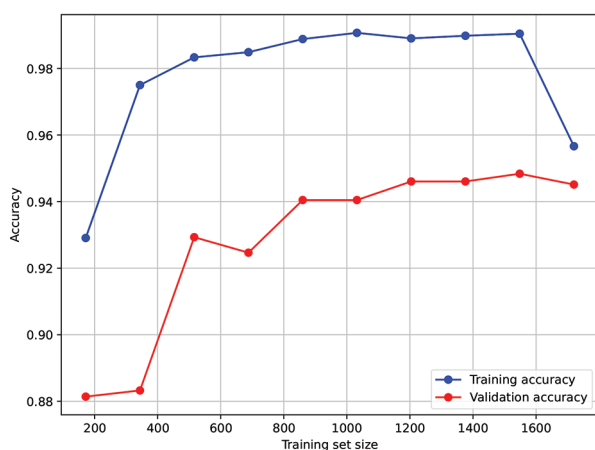


Figure 6. Learning curve for the decision tree classifier

interprets these findings, contextualizes them within the established landscape of AD diagnostics, and considers the broader implications and limitations of our data-driven approach. The model itself offers a transparent, rule-based architecture. This discussion will interpret these findings, contextualize them within the current AD diagnostic landscape, and consider the implications and limitations of this approach.

The most significant finding of this study is the clear hierarchy of predictive features. The analysis confirms that the model's diagnostic logic is predominantly governed by three features: Cognitive status (MMSE), functional ability (ADL and functional assessment), and patient-reported symptoms (memory complaints). This data-driven outcome strikingly mirrors established clinical wisdom. While the literature confirms that no single measure is paramount for diagnosis, it strongly advocates a composite evaluation combining these exact elements.⁴¹⁻⁴⁵ By autonomously identifying this combination as the most potent set of predictors, the model provides powerful validation of current multi-domain assessment strategies.

A particularly insightful result was the negligible or zero importance assigned to several well-established risk factors for AD, most notably family history of AD. Although the statistical analysis confirmed a strong association between family history and an AD diagnosis ($p < 0.001$), this feature was not utilized in the final classification architecture of the model. This suggests an inherent masking effect in the model's logic: Cognitive and functional declines are direct, measurable consequences of the underlying pathology, and family history is an upstream risk factor. The model has learned that, when it comes to quantifying this manifest decline, a low MMSE score is a more immediate and powerful predictor than the probabilistic risk conferred

by genetics. This effectively renders the historical factor redundant in terms of the model's decision-making.

The model's performance positions it as a compelling addition to the current diagnostic approach. The gold standard for confirming AD pathology often relies on expensive neuroimaging (positron emission tomography)⁴⁶ or invasive procedures involving blood-based biomarkers.⁴⁷⁻⁴⁹ This study shows that highly accurate classification can be achieved using data that is more routinely collected, less invasive, and more accessible. With a specificity of 95.7% and a sensitivity of 92.1%, the model shows significant potential as an effective screening or clinical triage tool. It could help clinicians stratify patients by identifying those at high risk who would benefit most from confirmatory biomarker testing. This would optimize resource allocation and reduce the diagnostic burden on patients.

A key strength of this work is its use of a fully transparent white-box model. Unlike more complex, opaque algorithms, every prediction made by our decision tree can be broken down into a series of clinically intuitive rules (Figure 5). Clinicians can trace the path from the root node, functional assessment ≤ 4.967 to a terminal leaf and understand precisely the evidence base for a given diagnostic suggestion. This inherent interpretability is essential for establishing clinical trust and is a prerequisite for the responsible adoption of machine learning tools in real-world healthcare settings.

This approach reflects the increasing demand for interpretable AI in medical diagnosis, as demonstrated by models such as the sparse high-order interaction model with rejection option, which was developed by Das *et al.*⁵⁰ Their work also emphasizes the importance of interpretability and reliability in AD diagnosis using proteomics data. It demonstrates how models can provide clear explanations and abstain from uncertain diagnoses, thereby complementing clinical judgment and supporting cost-effective, multistage diagnostic frameworks. Similarly, our decision tree contributes to designing more accessible and efficient diagnostic pathways for AD by transparently prioritizing readily available clinical assessments.

The learning curve analysis (Figure 6) provides further evidence of the model's robust performance and generalization capabilities. The consistently high training accuracy suggests that the decision tree effectively captures patterns in the training data, and the increasing and stabilizing validation accuracy shows that it can generalize to unseen data. The observed discrepancy between the training and validation accuracies is typical of a decision tree, reflecting its ability to fit the training data closely. However, the fact that both curves plateau with an increasing

training set size, particularly the stable validation accuracy, confirms that the model is not severely overfitting. This is crucial for clinical application, as it indicates reliability when faced with new patient data. Together, the findings from the learning curve and the high sensitivity, specificity, and AUC affirm the model's strong potential as a consistent and reliable diagnostic tool, particularly when sufficient training data is provided.

5.1. Ethical considerations and societal implications of AI in AD diagnosis

Deploying AI models in healthcare, especially for sensitive diagnoses such as AD, requires careful consideration of the ethical implications and societal impact.⁵¹ While our interpretable decision tree is transparent, it still operates within a broader context where issues such as patient autonomy, equitable access, the potential for bias, and accountability are paramount.

- (i) Patient autonomy and informed consent: Although our model provides a clear rationale for its predictions, ensuring patient autonomy requires individuals to understand how AI contributes to their diagnosis. Clinicians must clearly communicate that AI acts as a decision support tool rather than making definitive pronouncements, enabling patients to make informed decisions about subsequent diagnostic steps or treatment plans. The transparency of a decision tree facilitates this, as its logic can be explained more readily than that of a black box model.
- (ii) Equity of access and bias in sensitive populations: AI models trained on specific datasets inherently reflect the characteristics and potential biases of that data. While our dataset included diverse demographic information, real-world deployment must address potential biases related to ethnicity, socioeconomic status, or geographical location, as these could influence assessment scores or access to care. If a model performs differently across various sensitive populations, it could exacerbate existing health disparities. Future research and deployment strategies must therefore ensure equitable access to, and performance of, AI diagnostic tools. This could be achieved by implementing fairness-aware machine learning techniques and ensuring diverse representation in training datasets.
- (iii) Accountability for AI-driven decisions: The interpretable nature of our decision tree helps to establish clear accountability. When a diagnosis is supported by the model, clinicians can review the exact decision rules that led to the recommendation, enabling human oversight and validation. This is crucial for medical-legal considerations and for

building trust between patients, clinicians, and the AI system. The absence of an opaque black box judgment means accountability shifts toward the informed clinician using the tool rather than being placed solely on the algorithm.

- (iv) Societal implications: The ability of AI to facilitate earlier and more accessible AD diagnosis has significant societal implications. It could lead to earlier interventions, which could potentially slow disease progression or improve quality of life, as well as enabling better resource allocation. However, it also raises questions about the psychological impact of earlier diagnosis, the adequacy of support systems, and the ethical use of predictive information. To harness the full potential of AI for public health, policymakers and healthcare providers must proactively address these challenges.

5.2. Adherence to standardized validation frameworks

The clinical utility and reproducibility of AI tools in healthcare depend fundamentally on their development and validation within standardized frameworks. While our study is rigorous in its internal validation, it also acknowledges the broader landscape of regulatory guidance, such as that put forward by the US Food and Drug Administration for AI/machine learning-based medical devices.⁵² These frameworks emphasize the meticulous documentation of algorithmic choices, data provenance, pre-processing steps, dataset splitting protocols, comprehensive model evaluation, and critically, external validation strategies.

Many studies on AI model accuracy lack generalizability because they do not adhere to such rigorous documentation and validation standards, potentially conflating internal performance with true real-world utility. Our commitment to transparency, as demonstrated in the "Materials and methods" section and the interpretable nature of our decision tree architecture, aligns with these guidelines by offering a transparent record of our model's development and decision-making processes. Future work will prioritize aligning our model with these regulatory benchmarks, conducting prospective validation in diverse external cohorts, and documenting every stage to ensure its clinical utility and widespread adoption.

5.3. Limitations and future directions

Despite these promising results, several limitations must be acknowledged. First, as this study is based on a retrospective, cross-sectional dataset, it is not possible to make any claims about predicting the future onset of AD. Second, while the model's performance has been validated



Figure 7. Strengths, weaknesses, opportunities, threats analysis of the interpretable decision tree model for Alzheimer's disease diagnosis
Abbreviations: ADL: Activities of Daily Living; AI: Artificial intelligence; AUC: Area under the curve; EMRs: Electronic medical records; MMSE: Mini-Mental State Examination.

on an internal test set, its generalizability to external, more diverse populations must be established through further research. Third, the near-perfect separation of the “forgetfulness” symptom between the AD and control groups in this dataset may not accurately reflect the more nuanced and overlapping presentations of the symptom in a broader clinical population. This could lead to an overly optimistic estimation of the model's performance. Therefore, future work must prioritize prospective validation in a real-world clinical workflow.

5.4. Strengths, weaknesses, opportunities, threats (SWOT) analysis

The following section presents a SWOT analysis, the purpose of which is to provide a comprehensive overview of the interpretable decision tree model for AD diagnosis. This structured analysis underscores the model's inherent advantages, its current limitations, potential avenues for future development, and external factors that could influence its clinical integration (Figure 7).

In summary, this study demonstrates that an interpretable decision tree model based on a combination of cognitive and functional data can diagnose AD with a high level of accuracy, highlighting its potential in a clinical setting. The findings reinforce the importance of functional metrics and suggest that transparent, data-driven models have significant potential as decision support tools to help clinicians make timely and accurate diagnoses of this devastating disease. Future research should validate this model externally and explore hybrid approaches that integrate these powerful functional features with accessible, emerging blood-based biomarkers.

6. Conclusion

This study demonstrates that a transparent, rule-based decision tree classifier can diagnose AD with exceptional accuracy (91.8%) and discriminatory power (AUC = 0.93). The study unequivocally establishes that direct measures of cognitive and functional decline, specifically MMSE, ADL, and Functional Assessment scores are the most powerful predictors, capable of eclipsing the influence of traditional risk factors within a machine learning framework. The model's combination of high sensitivity, specificity, and interpretability highlights its potential as a reliable clinical decision support tool. This work highlights a viable pathway toward augmenting current diagnostic practices with accessible, trustworthy machine learning models, thereby enhancing the timely and accurate identification of individuals with AD.

Acknowledgments

None.

Funding

None.

Conflict of interest

The author declares no conflict of interest.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The dataset supporting the findings of this study, "Alzheimer's Disease Dataset," is publicly available on Kaggle at <https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset>.

References

- Monfared A, Byrnes M, White L, Zhang Q. Alzheimer's disease: Epidemiology and clinical progression. *Neurol Ther*. 2022;11(1):553-569.
doi: 10.1007/s40120-022-00338-8
- Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E. Alzheimer's disease. *Lancet*. 2011;377(9770):1019-1031.
doi: 10.1016/S0140-6736(10)61349-9
- Clifford K, Moreno M, Kloske C. Navigating late-stage dementia: A perspective from the Alzheimer's Association. *Alzheimers Dement (Amst)*. 2024;16(1):e12530.
doi: 10.1002/dad2.12530
- Weller J, Budson A. Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Res*. 2018;7:14506.
doi: 10.12688/f1000research.14506.1
- Mehta R, Schneider J. What is 'Alzheimer's disease'? The neuropathological heterogeneity of clinically defined Alzheimer's dementia. *Curr Opin Neurol*. 2021;34(2):237-245.
doi: 10.1097/WCO.0000000000000912
- Karantzoulis S, Galvin J. Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Rev Neurother*. 2011;11(11):1579-1591.
doi: 10.1586/ern.11.155
- Bondi M, Edmonds E, Salmon D. Alzheimer's disease: Past, present, and future. *J Int Neuropsychol Soc*. 2017;23(9-10):818-831.
doi: 10.1017/S135561771700100X
- Nasb M, Tao W, Chen N. Alzheimer's disease puzzle: Delving into Pathogenesis hypotheses. *Aging Dis*. 2023;15(1):43-73.
doi: 10.14336/AD.2023.0608
- Kim J, Jeong M, Stiles W, Choi H. Neuroimaging modalities in Alzheimer's disease: Diagnosis and clinical features. *Int J Mol Sci*. 2022;23(11):6079.
doi: 10.3390/ijms23116079
- Knopman D, Amièva H, Petersen R, et al. Alzheimer disease. *Nat Rev Dis Primers*. 2021;7(1):33.
doi: 10.1038/s41572-021-00269-y
- Long J, Holtzman D. Alzheimer disease: An update on pathobiology and treatment strategies. *Cell*. 2019;179(2):312-339.
doi: 10.1016/j.cell.2019.09.001
- Krishnamurthy H, Jayaraman V, Krishna K, et al. An overview of the genes and biomarkers in Alzheimer's disease. *Ageing Res Rev*. 2024;104:102599.
doi: 10.1016/j.arr.2024.102599
- Varesi A, Carrara A, Pires V, et al. Blood-based biomarkers for Alzheimer's disease diagnosis and progression: An overview. *Cells*. 2022;11(8):1367.
doi: 10.3390/cells11081367
- Kamatham P, Shukla R, Khatri D, Vora L. Pathogenesis, diagnostics, and therapeutics for Alzheimer's disease: Breaking the memory barrier. *Ageing Res Rev*. 2024;101:102481.
doi: 10.1016/j.arr.2024.102481
- Du X, Wang X, Geng M. Alzheimer's disease hypothesis and related therapies. *Transl Neurodegener*. 2018;7(1):1-14.
doi: 10.1186/s40035-018-0107-y
- Ashrafian H, Zadeh E, Khan R. Review on Alzheimer's disease: Inhibition of amyloid beta and tau tangle formation. *Int J Biol Macromol*. 2020;164:1451-1463.
doi: 10.1016/j.ijbiomac.2020.11.192
- Silva MVE, Loures CMG, Alves LCV, De Souza LC, Borges KBG, Carvalho MG. Alzheimer's disease: Risk factors and potentially protective measures. *J Biomed Sci*. 2019;26(1):33.
doi: 10.1186/s12929-019-0524-y
- Cortes-Canteli M, Iadecola C. Alzheimer's disease and vascular aging: JACC Focus seminar. *J Am Coll Cardiol*. 2020;75(8):942-951.
doi: 10.1016/j.jacc.2019.10.062
- Nordestgaard L, Christoffersen M, Frikke-Schmidt R. Shared risk factors between dementia and atherosclerotic cardiovascular disease. *Int J Mol Sci*. 2022;23(17):9777.
doi: 10.3390/ijms23179777
- Masurkar A, Marsh K, Morgan B, Leitner D, Wisniewski T. Factors affecting resilience and prevention of Alzheimer's disease and related dementias. *Ann Neurol*. 2024;96(1):136-148.
doi: 10.1002/ana.27055
- Lane CA, Hardy J, Schott JM. Alzheimer's disease. *Eur J Neurol*. 2018;25(1):59-70.
doi: 10.1111/ene.13439

22. Błaszczyk JW. Pathogenesis of dementia. *Int J Mol Sci.* 2022;24(1):543.
doi: 10.3390/ijms24010543
23. Yulin Hswen, Jiangmei Xiong, Margaret Hurley, Thu T. Nguyen. Experiences of Alzheimer's disease and related dementia family caregivers on Reddit communities: A topic modeling and sentiment analysis. *Artif Intell Health.* 2024;1(3):127–135.
doi: 10.36922/aih.3075
24. Materko W. Stratifying autonomic nervous system regulation patterns in healthy men: A machine learning approach. *Artif Intell Health.* 2025;2(4):025050006.
doi: 10.36922/AIH025050006
25. Materko W, Nascimento DGB, da Silva AS. Factor analysis of electrocardiographic findings, anthropometric measures, and age in patients with chronic kidney disease. *Brain and Heart.* 2025;3(2):5548.
doi: 10.36922/bh.5548
26. Zafar S, Mahmood T, Hoodbhoy Z, Hasan B. Pediatric patient hospital length of stay prediction: A comparative analysis of Bayesian inference and machine learning approaches. *Artif Intell Health.* 2025;2(4):025160030.
doi: 10.36922/AIH025160030
27. Kourmpanis N, Liaskos J, Zoulias E, Mantas J. Predicting mortality outcomes in individual COVID-19 patients using machine learning algorithms. *Artif Intell Health.* 2024;1(3):31-52.
doi: 10.36922/aih.2591
28. Rakhmetulayeva SB, Kulbayeva AK. Symptomatic assessment of diseases using decision trees and analysis of electronic medical records. *Int J Inf Commun Technol.* 2022;3(1):66-73.
doi: 10.54309/ijict.2022.9.1.009
29. Mienye ID, Jere N. A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access.* 2024;12: 86716-86727.
doi: 10.1109/ACCESS.2024.3416838
30. Gupta B, Rawat A, Jain A, Arora A, Dhami N. Analysis of various decision tree algorithms for classification in data mining. *IJCA.* 2017;163(8):15-19.
doi: 10.5120/ijca2017913660
31. Arunakumari BN, Poornima B, Shobha G. Enhancing the Quality and Efficiency of Mental Health Care using Decision Trees. In: *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT).* 2023:1-5.
doi: 10.1109/ICCCNT56998.2023.10306470
32. Venkatasubramaniam A, Wolfson J, Mitchell N, Barnes T, JaKa M, French S. Decision trees in epidemiological research. *Emerg Themes Epidemiol.* 2017;14(1):11.
doi: 10.1186/s12982-017-0064-4
33. Costa VG, Pedreira CE. Recent advances in decision trees: An updated survey. *Artif Intell Rev.* 2022;56:4765-4800.
doi: 10.1007/s10462-022-10275-5
34. Blockeel H, Devos L, Frénay B, Nanfack G, Nijssen S. Decision trees: from efficient prediction to responsible AI. *Front Artif Intell.* 2023;6:1124553.
doi: 10.3389/frai.2023.1124553
35. Kotsiantis SB. Decision trees: a recent overview. *Artif Intell Rev.* 2011;39(4):261-283.
doi: 10.1007/s10462-011-9272-4
36. Quinlan JR. Induction of decision trees. *Mach Learn.* 2004;1(1):81-106.
doi: 10.1023/A:1022643204877
37. Rokach L. Decision forest: Twenty years of research. *Inf Fusion.* 2016;27:111-125.
doi: 10.1016/j.inffus.2015.06.005
38. Hehn TM, Mumbach M, Payer M. End-to-end learning of decision trees and forests. *Int J Comput Vis.* 2019;127(10):1464-1477.
doi: 10.1007/s11263-019-01237-6
39. Tsang S, Kao B, Yip KY, Ho WS, Lee SD. Decision Trees for Uncertain Data. *IEEE Trans Knowl Data Eng.* 2011;23(1):64-78.
doi: 10.1109/tkde.2009.175
40. Karalis G. Decision Trees and Applications. In: *Advances in Experimental Medicine and Biology.* Springer International Publishing; 2020:239-242.
doi: 10.1007/978-3-030-32622-7_21
41. Nikita E, Nikitas P. Data mining and decision trees. In: *Statistics and Probability in Forensic Anthropology.* Elsevier; 2020:87-105.
doi: 10.1016/b978-0-12-815764-0.00012-5
42. Wang G, Estrella A, Hakim O, et al. Mini-mental state examination and montreal cognitive assessment as tools for following cognitive changes in Alzheimer's disease neuroimaging initiative participants. *J Alzheimers Dis.* 2022;90(1):263-270.
doi: 10.3233/JAD-220397
43. Arévalo-Rodríguez I, Smailagic N, Roqué-Figuls M, et al. Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev.* 2021;7(7):CD010783.
doi: 10.1002/14651858.CD010783.pub3

44. Çebi M, Babacan G, Tanör Ö, Gürvit H. Discrimination ability of the Short Test of Mental Status (STMS) compared to the Mini Mental State Examination (MMSE) in the spectrum of normal cognition, mild cognitive impairment, and probable Alzheimer's disease dementia: The Turkish standardization study. *J Clin Exp Neuropsychol*. 2020;42(5):450-458.
doi: 10.1080/13803395.2020.1758633
45. Skorga P, Young C. Mini-mental state examination for the detection of Alzheimer disease and other dementias in people with mild cognitive impairment. *Clin Nurse Spec*. 2015;29(5):E1-E4.
doi: 10.1097/NUR.0000000000000150
46. Razani J, Wong JT, Dafaeeboini N, *et al*. Predicting everyday functional abilities of dementia patients with the mini-mental state examination. *J Geriatr Psychiatry Neurol*. 2009;22(1):62-70.
doi: 10.1177/0891988708328217
47. Márquez F, Yassa M. Neuroimaging biomarkers for Alzheimer's disease. *Mol Neurodegener*. 2019;14(1):21.
doi: 10.1186/s13024-019-0325-5
48. Hansson O, Blennow K, Zetterberg H, Dage J. Blood biomarkers for Alzheimer's disease in clinical practice and trials. *Nat Aging*. 2023;3(5):506-519.
doi: 10.1038/s43587-023-00403-3
49. Teunissen CE, Verberk IMW, Thijssen EH, *et al*. Blood-based biomarkers for Alzheimer's disease: Towards clinical implementation. *Lancet Neurol*. 2021;21(1):66-77.
doi: 10.1016/S1474-4422(21)00361-6
50. Das D, Ito J, Kadowaki T, Tsuda K. An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ*. 2019;7:e6543.
doi: 10.7717/peerj.6543
51. Ursin F, Timmermann C, Steger F. Ethical implications of Alzheimer's disease prediction in asymptomatic individuals through artificial intelligence. *Diagnostics (Basel)*. 2021;11(3):440.
doi: 10.3390/diagnostics11030440
52. Joshi G, Jain A, Araveeti SR, Adhikari S, Garg H, Bhandari M. FDA-Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. *Electronics*. 2024;13(3):498.
doi: 10.3390/electronics13030498