NPG Neurologie - Psychiatrie - Gériatrie xxx (xxxx) xxx—xxx

Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

FUNDAMENTALS

# Predictive models for Alzheimer's disease diagnosis and MCI identification: The use of cognitive scores and artificial intelligence algorithms

## *Modèles prédictifs pour le diagnostic de la maladie d'Alzheimer et l'identification de la détérioration cognitive légère : exploitation des scores cognitifs et des algorithmes d'intelligence artificielle*

**S.-A. Sadegh-Zadeh [a,*], M.-J. Nazari [b], M. Aljamaeen [a], F.S. Yazdani [c], S.Y. Mousavi [d], Z. Vahabi [e]**

[a] *Department of Computing, Staffordshire University, Stoke-on-Trent, ST4 2DE, UK*
[b] *Computer Science and Mathematics Department, Amirkabir University of Technology, 424 Hafez Ave, Tehran, Iran*
[c] *Faculty of Education Sciences and Psychology, Ferdowsi University of Mashhad, Mashhad, Iran*
[d] *Department of Psychiatry, Tehran University of Medical Sciences, Tehran, Iran*
[e] *Department of Geriatrics Medicine and Cognitive Neurology and Neuropsychiatry Department, Tehran University of Medical Sciences, Tehran, Iran*

**KEYWORDS**
Alzheimer's disease ;
Mild cognitive impairment (MCI) ;
Predictive models ;
Artificial intelligence ;
Machine-learning classifiers ;
Cognitive scores

**Summary** The paper presents a comprehensive study on predictive models for Alzheimer's disease (AD) and mild cognitive impairment (MCI) diagnosis, implementing a combination of cognitive scores and artificial intelligence algorithms. The research includes detailed analyses of clinical and demographic variables such as age, education, and various cognitive and functional scores, investigating their distributions and correlations with AD and MCI. The study utilizes several machine-learning classifiers, comparing their performance through metrics like accuracy, precision, and area under the ROC curve (AUC). Key findings include the influence of gender on AD prevalence, the potential protective effect of education, and the significance of functional decline and cognitive performance scores in the models. The results demonstrate

* Corresponding author.
*Adresses e-mail :* ali.sadegh-zadeh@staffs.ac.uk (S.-A. Sadegh-Zadeh), mohammadjnazari@aut.ac.ir (M.-J. Nazari), murouj.aljamaeen1@staffs.ac.uk (M. Aljamaeen), fa.yazdani@mail.um.ac.ir (F.S. Yazdani), drsymousavi@gmail.com (S.Y. Mousavi), vahab1357@gmail.com (Z. Vahabi).

https://doi.org/10.1016/j.npg.2024.04.004


Pour citer cet article : S.-A. Sadegh-Zadeh, M.-J. Nazari, M. Aljamaeen et al., Predictive models for Alzheimer's disease diagnosis and MCI identification: The use of cognitive scores and artificial intelligence algorithms, Neurol psychiatr gériatr, https://doi.org/10.1016/j.npg.2024.04.004

the effectiveness of ensemble methods and the robustness of the models across different data subsets, highlighting the potential of artificial intelligence in enhancing diagnostic accuracy for Alzheimer's disease and mild cognitive impairment.


**MOTS-CLÉS**
Maladie d'Alzheimer (MA) ;
Détérioration cognitive légère (DCL) ;
Scores cognitifs ;
Apprentissage automatique ;
Intelligence artificielle

**Résumé** Cette étude explore l'application des algorithmes d'apprentissage automatique pour le diagnostic de la maladie d'Alzheimer (MA) et l'identification de la détérioration cognitive légère (DCL), en utilisant des scores cognitifs parmi d'autres variables cliniques et démographiques. Nous décrivons notre méthodologie, incluant la collecte de données, le prétraitement, la sélection des caractéristiques, et l'utilisation de divers classificateurs d'apprentissage machine. Les résultats mettent en évidence l'efficacité des méthodes d'ensemble dans la prédiction de la MA et de la DCL, discutent des implications de ces résultats pour le diagnostic précoce et l'intervention, et suggèrent des directions pour les recherches futures.


Alzheimer's disease (AD) is a leading neurodegenerative disorder that predominantly affects the elderly, causing irreversible damage to cognitive functions, especially memory [1]. This deterioration severely impacts the quality of life. In the same context, mild cognitive impairment (MCI) is another condition characterized by notable cognitive deficiencies [2]. While these deficiencies do not severely interfere with daily activities, they are significant enough to be a focus of concern. It is crucial to note that MCI often acts as a precursor to AD, underlining the importance of its early diagnosis [3].

As the world's population continues to age, there is a corresponding rise in the prevalence of neurodegenerative conditions, including AD [4]. Current estimates from the World Health Organization indicate that AD, along with other forms of dementia, ranks as the seventh primary cause of death worldwide [5]. On an individual level, Alzheimer's disease imposes considerable emotional and financial strain on the patients and their families. Additionally, from a broader perspective, the healthcare sector faces the challenge of managing the rising costs of care, pointing to the need for specialized caregivers. MCI, occurring in the initial stages of cognitive decline, also presents its unique challenges. However, an early detection of MCI can pave the way for proactive measures, potentially halting or at least slowing down its evolution into full-blown Alzheimer's [6].

This research primarily aims to delve into the screening potential of specific cognitive tests and functional scales, namely FAST, MoCA, MMSE, GDS, ADL, and IADL, for Alzheimer's disease and MCI. Furthermore, the study seeks to harness the power of advanced artificial intelligence techniques to augment the diagnostic precision for these conditions. By calling on various machine-learning algorithms, this research will also focus on understanding their comparative effectiveness in this specialized domain and on extracting valuable insights from their results.

The pivotal questions guiding this research include: how proficient are the aforementioned cognitive scores in predicting Alzheimer's disease and mild cognitive impairment? Which among the machine-learning algorithms offers the utmost accuracy and precision in diagnosing AD and identifying MCI using these cognitive scores? And lastly, are there specific cognitive scores that stand out in terms of their predictive ability for AD and MCI?

The structure of this research paper is methodically laid out for clarity and coherence. Following this introduction, the first section offers an exhaustive literature review, encompassing Alzheimer's disease, MCI, the various cognitive assessment tools, and the indispensable role of machine-learning in detecting AD. The second section ventures into the intricate processes of data collection and pre-processing, illustrating the origin of the data and how it is managed. The third section, meanwhile, sheds light on feature engineering—a pivotal step to amplify the predictive efficacy of machine-learning models. Subsequently, the fourth section elaborates on the research methodology, detailing the machine-learning algorithms selected, how the models are trained, and the techniques used for evaluation. The penultimate section, fifth section, details the experimental results, coupled with a thorough interpretation. The sixth section delves deeper into these findings, discussing their broader implications, the limitations of this study, and potential directions for future research. The concluding section sums up the principal discoveries and their substantial contribution to the broader scientific community.

## Literature review

### Alzheimer's disease and mild cognitive impairment (MCI)

AD is a progressive neurodegenerative disorder that affects memory, cognition, and daily living activities. It is the most common form of dementia and affects millions of individuals worldwide [1,7]. Pathologically, AD is characterized

by the accumulation of amyloid-beta plaques and tau protein tangles in the brain [8]. On the other hand, MCI is a transitional stage between normal aging and Alzheimer's. People with MCI have memory problems more severe than expected for their age but not severe enough to interfere with daily life. It is worth noting that not everyone with MCI will develop AD, but they are at a higher risk [9,10].

## Cognitive assessment tools

### FAST scale

The Functional Assessment Staging Test (FAST) is a tool that measures the progression of Alzheimer's disease across seven stages. It focuses on an individual's functional abilities and helps to determine the level of dementia severity [11].

### MoCA score

The Montreal Cognitive Assessment (MoCA) is designed to detect early stages of cognitive impairments. It evaluates different cognitive domains: attention, memory, language, and spatial skills, among others. Scores below 26 out of 30 are considered indicative of mild cognitive impairment [12].

### MMSE score

The Mini-Mental State Examination (MMSE) is one of the most commonly used cognitive tests for dementia. It assesses various cognitive functions including arithmetic, memory, and orientation. A score below 24 out of 30 is typically considered abnormal [13].

### ADL score

Activities of Daily Living (ADL) focus on an individual's ability to perform basic daily tasks such as bathing, dressing, and feeding. A decline in ADL scores can be indicative of cognitive impairments [14].

### IADL score

The Instrumental Activities of Daily Living (IADL) measures more complex activities like handling finances, managing medications, and transportation. Like ADL, deteriorating IADL scores can signal cognitive decline [15].

### GDS scale

The Geriatric Depression Scale (GDS) provides an overview of the six stages of cognitive decline, from no cognitive decline to very severe cognitive decline [16].

## Machine-learning in the detection of Alzheimer's disease

Over the past decade, machine-learning has emerged as a powerful tool for diagnosing and predicting the onset of AD [17]. As the complexity and volume of data related to patient health records and neuroimaging increases, traditional statistical methods have proven insufficient in detecting the subtle patterns associated with AD [18]. Machine-learning algorithms can process vast amounts of data and extract patterns that might not be immediately apparent to human researchers [17,18].

There are various studies that have utilized machine-learning methods to diagnose and predict AD. For example, Abed et al. [19] used deep learning, a subfield of machine-learning, to process neuroimaging data and achieved an accuracy of over 90% in detecting AD. Similarly, Sheng et al. [20] utilized support vector machines (SVM) on genetic data, showing the versatility of machine-learning applications in this domain.

Furthermore, machine-learning has not only been applied to neuroimaging and genetic data but also to other data types such as electronic health records [21], neuropsychological test results [22], and others. This multi-modal approach has the potential to provide a holistic view of patients and can improve the predictive accuracy of the models.

## Previous research on cognitive scores and Alzheimer's disease

Cognitive scores serve as key indicators of cognitive decline and potential onset of neurodegenerative disorders like Alzheimer's disease. Various cognitive assessment tools, such as the FAST, MoCA, MMSE, GDS, ADL, and IADL scores, have been employed in both clinical and research settings to evaluate cognitive functions [23—25]. Numerous studies have attempted to correlate the results from these cognitive assessments with AD onset. For instance, Abed et al. [19] found that the MoCA score was particularly effective in distinguishing individuals with MCI from those with AD. Abed et al. [19] highlighted that the MMSE score was effective in detecting early-stage Alzheimer's disease but had limitations in its ability to differentiate between MCI and normal.

In the context of machine-learning, cognitive scores have been used to train predictive models. Combining cognitive scores with other types of data, such as neuroimaging or genetic data, can potentially increase the accuracy of AD predictions [26].

## Identification of shortfalls

Despite the significant advances made in the detection of Alzheimer's disease using machine-learning algorithms and cognitive scores, certain shortfalls remain:

- holistic approach: while many studies have focused on either machine-learning or cognitive scores individually, there is a limited body of work that integrates the two for a holistic diagnostic approach;
- significant features: few studies have delved deep into understanding which specific features or combinations of features (among the cognitive scores) provide the highest predictive power for AD diagnosis;
- generalizability: many machine-learning models are trained and tested on limited datasets, often from specific populations or geographical regions, limiting their generalizability;
- early Detection: while there is a focus on differentiating between MCI and AD, there is a need for predictive models that can detect the shift from a normal cognitive state to MCI, enabling earlier interventions.

The current research aims to address these shortfalls, providing a comprehensive approach to AD diagnosis using both machine-learning algorithms and cognitive scores.

## Data collection and pre-processing

### Participant recruitment and description

The study recruited participants from the outpatient geriatric centre associated with the Tehran University of Medical Sciences (TUMS), Iran, over a six-month period from October 2021 to March 2022. This cohort comprised individuals initially assessed for memory problems or those previously diagnosed with AD or MCI. The recruitment strategy was designed to include a diverse sample, reflecting a range of cognitive impairments from mild to moderate stages.

Participants were included on the basis of a comprehensive evaluation that incorporated clinical history, neurological examination, and neuropsychological assessment. The diagnosis of AD or MCI was established according to the criteria set out by the National Institute on Aging and the Alzheimer's Association (NIA-AA). This included both individuals newly diagnosed during the study timeframe and those with prior diagnoses seeking ongoing care at the centre.

To ensure a holistic understanding of each participant's condition, the study meticulously recorded not only the cognitive status but also the treatment regimens for dementia and any psychiatric symptoms. This included documentation of prescribed medications for dementia (e.g. acetylcholinesterase inhibitors, memantine) and any psychiatric medications to manage associated symptoms such as depression, anxiety, or sleep disturbances.

The inclusion criteria were specifically designed to encompass:

- adults aged 45 to 90 years;
- a diagnosis of mild to moderate Alzheimer's disease or mild cognitive impairment;
- consent to participate in the study.

Exclusion criteria were applied to omit cases involving recent stroke, vascular dementia, Parkinson's disease, and major psychiatric disorders to minimize confounding variables that could impact cognitive assessments.

The study was conducted in strict adherence to the ethical guidelines of the Declaration of Helsinki and received approval from the Ethics Committee of the Tehran University of Medical Sciences under the number: IR.TUMS.MEDICINE.REC.1399.1183.

### Data collection procedure

AD identification was based on the NINDS-ADRDA criteria as set out by McKhann et al. [27] in 1984, in conjunction with the criteria established by the National Institute on Aging Alzheimer's Association, as presented by Sperling et al. [28]. To be considered for the study, participants were to be aged between 45 and 90 years, and their condition needed to be categorized as mild or moderate AD based on the Functional Assessment Staging Tool (FAST) score [11], ranging from 4—6B.

Cognitive function assessment: the preliminary evaluation of cognitive abilities was performed using two main tests:

- Mini-Mental State Examination (MMSE): a renowned tool for brief cognitive assessment, the MMSE covers various cognitive disorders. Its origins trace back to the work by McKhann et al. [27] and it functions as an aid for clinicians to diagnose conditions like dementia and delirium. With 11 domains of evaluation, scores can range from 0 to 30, with a higher score implying better cognitive performance [29]. In this study, various statistical data points concerning the MMSE score were mentioned, including scores for the normal group, the cognitive decline group, and the overall average, alongside the average age of all patients and age differences across various groups;
- Montreal Cognitive Assessment (MoCA): this test is a standard tool to explore diverse facets of cognitive function, especially in older adults. The MoCA assessment, which lasts about 10 minutes, offers scores ranging between 0 and 30, with higher scores denoting better cognitive faculties [12].

Performance in daily living activities: two scales were employed to evaluate the participants' abilities in their daily routines:

- Barthel Index (BI): this index assesses elderly individuals' independence when executing daily chores, encompassing ten distinct variables. The scoring metric is straightforward: 0 indicates inability, 1 signifies needing assistance, and 2 denotes independence. Hence, the total score can vary from 0 to 20, where a higher score equates to greater independence [30];
- Lawton Instrumental Activities of Daily Living (IADL) Scale: specifically tailored for Iranians, this validated test assesses competencies in more intricate daily tasks. Spanning eight questions, the IADL scale delves into abilities ranging from telecommunication to financial management. Its score ranges from 0 (low function) to 8 (high function). The IADL evaluation offers insights into cognitive or physical function, although a combination of both is generally required [31].

The participant selection process for our study is summarized in the flowchart presented in Fig. 1. This figure delineates the stages of participant recruitment, the criteria for inclusion and exclusion, and the final distribution of the participants whose diagnoses were confirmed. It provides a visual breakdown of the cohort, highlighting the strict methodology employed to refine the dataset and the resulting composition of the study population.

### Data set description

This study's data facilitates a holistic understanding of factors influencing dementia and its subtypes, providing a multifaceted approach to discern the interplay of genetics, lifestyle, comorbidities, and cognitive performance in the progression of neurodegenerative disorders. The s depth and breadth of this dataset, encompassing both quantitative and categorical variables, make it suitable for rigorous machine-learning or statistical analyses aiming to identify predictors and patterns associated with dementia-related outcomes. In all the dataset comprises 41 features. It

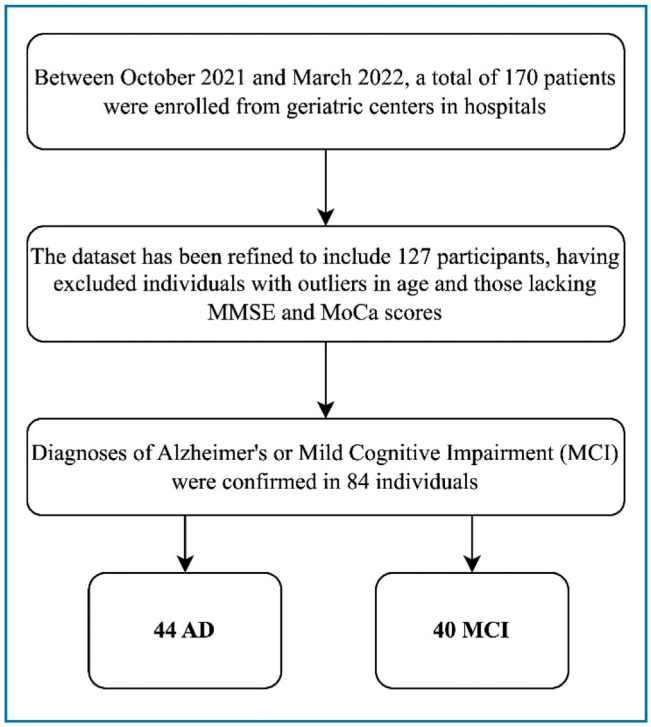


**Figure 1.** Participant selection flowchart. The flowchart illustrates the process of enrolling patients from geriatric centres in hospitals, refining the dataset to exclude outliers and incomplete records, and confirming diagnoses of Alzheimer's or MCI.

comprises comprehensive demographic, clinical, and cognitive assessments of participants potentially at risk of or diagnosed with dementia-related disorders. Key variables include unique identifiers (before and after data cleaning), demographic attributes such as age, gender, marital status, and educational level, along with medical histories including familial predispositions to AD. Additionally, lifestyle factors like smoking, alcohol consumption, and sleep patterns are documented. Crucially, the dataset evaluates participants' cognitive function through tools like the FAST, MoCA, MMSE, GDS, and measures of daily living activities, both basic (ADL) and instrumental (IADL). The dataset culminates in a diagnostic variable, categorizing participants as Normal, AD, MCI, or requiring expert consultation.

Table 1 provides a detailed analysis between AD and MCI groups, incorporating demographic variables (age, gender, age at onset), educational background, and cognitive assessments (MoCA score, MMSE score, ADL score, IADL score). The inclusion of gender distribution and age at onset offers a nuanced understanding of the participant demographics, revealing a higher prevalence of females in the MCI group and a later average age at onset for AD, suggesting potential gender-related differences and timing differences in disease manifestation. The significant differences in MoCA and IADL scores between the groups underscore cognitive and functional disparities, with MCI participants demonstrating better performances, indicative of their milder impairment. Although age, education, MMSE, and ADL scores do not show significant statistical differences, they provide an essential context to the overall analysis. This comprehensive comparison facilitates a deeper understanding of the characteristics distinguishing AD from MCI, contributing to more precise diagnostic criteria and potentially pointing the way to targeted intervention strategies.

**Table 1** Comparative analysis of clinical and demographic characteristics between AD and MCI groups, highlighting key differences in cognitive functioning and daily living activities.

| Variable | AD Mean | MCI Mean | *P*-value |
|---|---|---|---|
| Age | 73.39 | 69.98 | 0.0062 |
| Education | 6.15 | 7.92 | 0.0021 |
| MoCA Score | 11.35 | 19.06 | < 0.0001 |
| MMSE Score | 15.39 | 18.50 | 0.0011 |
| ADL Score | 4.64 | 5.55 | 0.0088 |
| IADL Score | 4.71 | 6.81 | 0.0012 |
| GDS Score | 3.45 | 2.86 | 0.069 |
| Gender - Female (%) | 56.82 | 70.00 | NA |
| Gender - Male (%) | 43.18 | 30.00 | NA |
| Age at Onset | 71.05 | 66.86 | 0.0038 |

## Data pre-processing steps

### Handling missing data

Handling missing data was paramount to ensuring the integrity and reliability of our dataset. Given the multifaceted nature of the data, we adopted a two-pronged approach for imputation:

- demographic information: for missing demographic attributes like age and numbers of children, we utilized both mean and mode imputation techniques. Mean imputation replaces missing values with the mean value of that particular column, while mode imputation replaces missing values with the most frequently occurring value. This dual strategy ensured that the demographic attributes retained their fundamental structure and significance [32];
- scores: our dataset exhibited missing values in pivotal cognitive and functional scores, including the FAST, MoCA, IADL, GDS, and MMSE metrics. To handle these missing scores, we used regression imputation. This technique involves training regression models on the available data and then using these models to predict and fill in the missing scores. This method ensures that the imputed values are consistent with the relationships observed in the rest of the data [33].

### Feature selection

Feature selection is the process of selecting the most pertinent features that contribute to the predictive power of the model [18]. Our strategy hinged on the use of Recursive Feature Elimination (RFE) with a Random Forest Classifier. Through this method, we identified critical features, such as the FAST score, MoCA score, IADL score, and GDS score, among others. These features exhibited varying degrees of importance, with the FAST score emerging as the most pivotal. By concentrating on these vital attributes, we were able to optimize the efficiency and accuracy of our machine-learning models.

### Data scaling

Given the diverse range of scores and attributes in our dataset, scaling was essential to ensure that no single feature disproportionately influenced the model. We employed the Min—Max scaling method, which transforms features by scaling them to a given range, typically between 0 and 1. This normalization ensures that all features contribute equally to the model's performance, preventing any single attribute from overshadowing the rest [34].

### Data encoding

To handle categorical variables in our dataset, we adopted a data encoding approach. Encoding was used to convert nominal categorical variables into a format that could be fed to machine-learning algorithms for more effective processing. By converting these categorical variables into a series of binary columns, we ensured that our model could interpret and manage this data efficiently.

The data pre-processing phase was instrumental in refining our dataset for optimal model training. By meticulous handling of missing data, feature selection, scaling, and encoding, we laid the way for successful Alzheimer's disease diagnosis and MCI identification using machine-learning algorithms.

## Feature engineering

Feature engineering plays a vital role in enhancing the predictive performance of machine-learning models [35,36]. By refining and transforming features, it is possible to provide more informative inputs for our models, enabling them to make more accurate predictions. In the context of our research on AD and MCI diagnosis, the significance of the features plays a pivotal role in the successful differentiation between the two conditions explored.

### Data augmentation with SMOTE

Recognizing the challenges with class imbalances and limited diversity in our dataset, the Synthetic Minority Oversampling Technique (SMOTE) was employed. This technique generates synthetic data, ensuring the model has a richer understanding of the data distribution [37]. By augmenting the data in this manner, the diversity of the dataset was improved, leading to enhanced model generalization and predictive accuracy.

### Analysis of the importance of different features

For feature selection, it was crucial to quantify the significance of each selected feature. Using the Random Forest Classifier, we gauged the importance of each feature. The outcomes were as follows:

- FAST Score (0.276): as the most crucial feature, the FAST score provided insights into the level of functional impairment, offering a direct window on disease progression;
- MoCA Score (0.116): assessing cognitive performance, the MoCA score appeared as the second most important feature, emphasizing the significance of cognitive functioning in the classification process;
- IADL Score (0.072): representing the ability to perform essential daily tasks, IADL scores pinpointed the functional aspects that correlate with disease progression;
- GDS Score (0.069) and MMSE Score (0.069): the two scores, measuring depression levels and cognitive impairment respectively, shared equal importance, underscoring their pivotal roles in the diagnosis process.

The analysis the importance of different features provided a clear hierarchy of features, guiding the modelling process to prioritize them accordingly. By including these key features and utilizing advanced engineering techniques, the foundation was laid for building a robust and accurate classification model.

Fig. 2 displays a series of histograms overlaid with kernel density estimates, representing the distribution of various clinical and demographic variables in this study. The variables include age, age at onset, level of education, FAST score, MoCA score, MMSE score, GDS score, ADL score, and IADL score. Fig. 3 shows a heatmap of correlation coefficients, which is used in statistical analysis to show the strength and direction of relationships between variables. This heatmap represents the correlation across various factors such as age, age at onset, education, and different cognitive and functional scores (FAST, MoCA, MMSE, GDS, ADL, IADL) with each other and with the diagnosis of Alzheimer's disease. The colour gradient from dark purple to dark red indicates the range of correlations from −1 (perfect negative correlation) to +1 (perfect positive correlation), with darker shades indicating stronger relationships and white indicating no correlation.

Fig. 4 shows matrix plots showing the relationship between various variables. The diagonal plots, which are kernel density estimations, show the distribution of each variable, such as age, age at onset, education, and several cognitive test scores like FAST, MoCA, MMSE, GDS, ADL, and IADL. Red dots represent patients with AD and blue dots represent patients with MCI. The off-diagonal plots are scatter plots showing the correlation between the pairs of these variables, separated by diagnosis. As shown in Fig. 4, we can see that for some variables, the distributions between AD and MCI overlap significantly (e.g. education), while others show more distinction (e.g. MoCA score, MMSE score). The scatter plots reveal certain trends and correlations, such as a possible negative correlation between cognitive test scores (MoCA, MMSE) and the severity of impairment (AD is more severe than MCI). Higher FAST scores, which imply greater functional decline, seem to be associated with AD rather than MCI. GDS scores, which measure depression, do not show a clear differentiation between AD and MCI, suggesting that depression levels may not be particularly correlated with the type of cognitive impairment. ADL and IADL scores, indicative of daily living abilities, show a clear trend where AD patients typically have lower scores, denoting more significant daily living challenges. These patterns suggest that while some variables might be better at distinguishing between MCI and AD, others may not, indicating the complexity of diagnosing and differentiating these conditions.

Fig. 5 displays a horizontal bar chart which represents the relative importance of features in the predictive model.

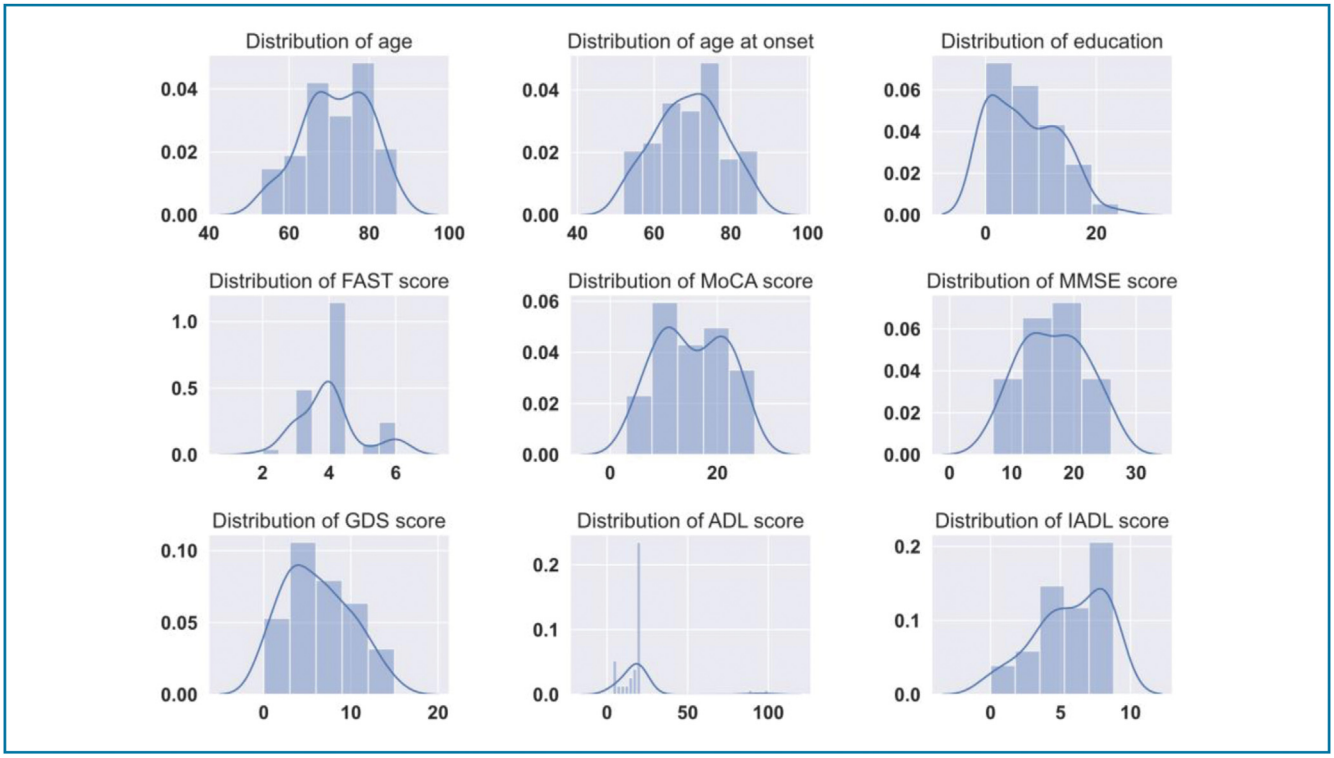


**Figure 2.** Skewness in the distribution of various clinical assessment scores, including age, age at onset, level of education, FAST score, MoCA score, MMSE score, GDS score, ADL score, and IADL score.

The features listed on the y-axis, from top to bottom, include 'Education', 'ADL score', 'Age at onset', 'Age', 'MMSE score', 'GDS score', 'IADL score', 'MoCA score', and 'FAST score'. The x-axis represents the importance of these features in the model, ranging from 0.00 to greater than 0.25. The 'FAST score' appears to be the most important feature, showing the highest value on the importance scale, followed by the 'MoCA score'. The least important feature appears to be 'Education', with a noticeably smaller bar.

Fig. 6 is a boxplot representation of various features. These features are variables or scores used in predictive modelling. The boxplot for each feature displays the median (the line in the middle of the box), the interquartile range (the box itself), and potential outliers (the individual dots beyond the 'whiskers'). The whiskers extend to the highest and lowest values within 1.5 times the interquartile range from the box. Outliers are considered as individual data dots lying beyond the ends of the whiskers.

The boxplots suggest varying degrees of dispersion among the features, with 'age' having the broadest spread and 'MMSE score' and 'MoCA score' appearing more tightly grouped. The presence of outliers, particularly in 'age' and 'IADL score', could indicate atypical cases or errors in data collection that could need to be addressed. The median lines within the boxplots for 'MMSE score' and 'MoCA score' are lower, which is consistent with the use of these tests in detecting cognitive impairments, where lower scores suggest greater impairment, potentially associated with Alzheimer's or MCI. The distribution and central tendency of these features are critical for the predictive models mentioned in the paper title, as they influence the accuracy and reliability of such models. Understanding the distribution of these features can help in refining the models and selecting appropriate machine-learning algorithms for prediction.

## Methodology

In this study, we employed an observational and cross-sectional design to investigate the predictive capacity of various cognitive scores and demographic characteristics for the diagnosis of AD and MCI. This approach allowed us to capture and analyse data from participants at a single time point, providing a snapshot of the relationship between clinical variables and diagnostic outcomes without inferring causality. Such a design is instrumental in identifying potential markers and trends associated with AD and MCI, contributing to the broader understanding of these conditions.

In the development of our predictive models, the selection of variables was guided by a comprehensive literature review, expert clinical input, and preliminary statistical analyses. Variables such as age, gender, education level, cognitive scores (e.g., MoCA, MMSE), and functional assessments (e.g. ADL, IADL scores) were chosen for their established correlation with neurocognitive disorders in existing research. Additionally, we performed an exploratory data analysis to identify variables with significant differences between diagnostic groups, ensuring that our models identified factors with strong discriminative power. This multifaceted approach enabled us to construct models that are both theoretically grounded and empirically validated, enhancing their predictive accuracy and clinical relevance.

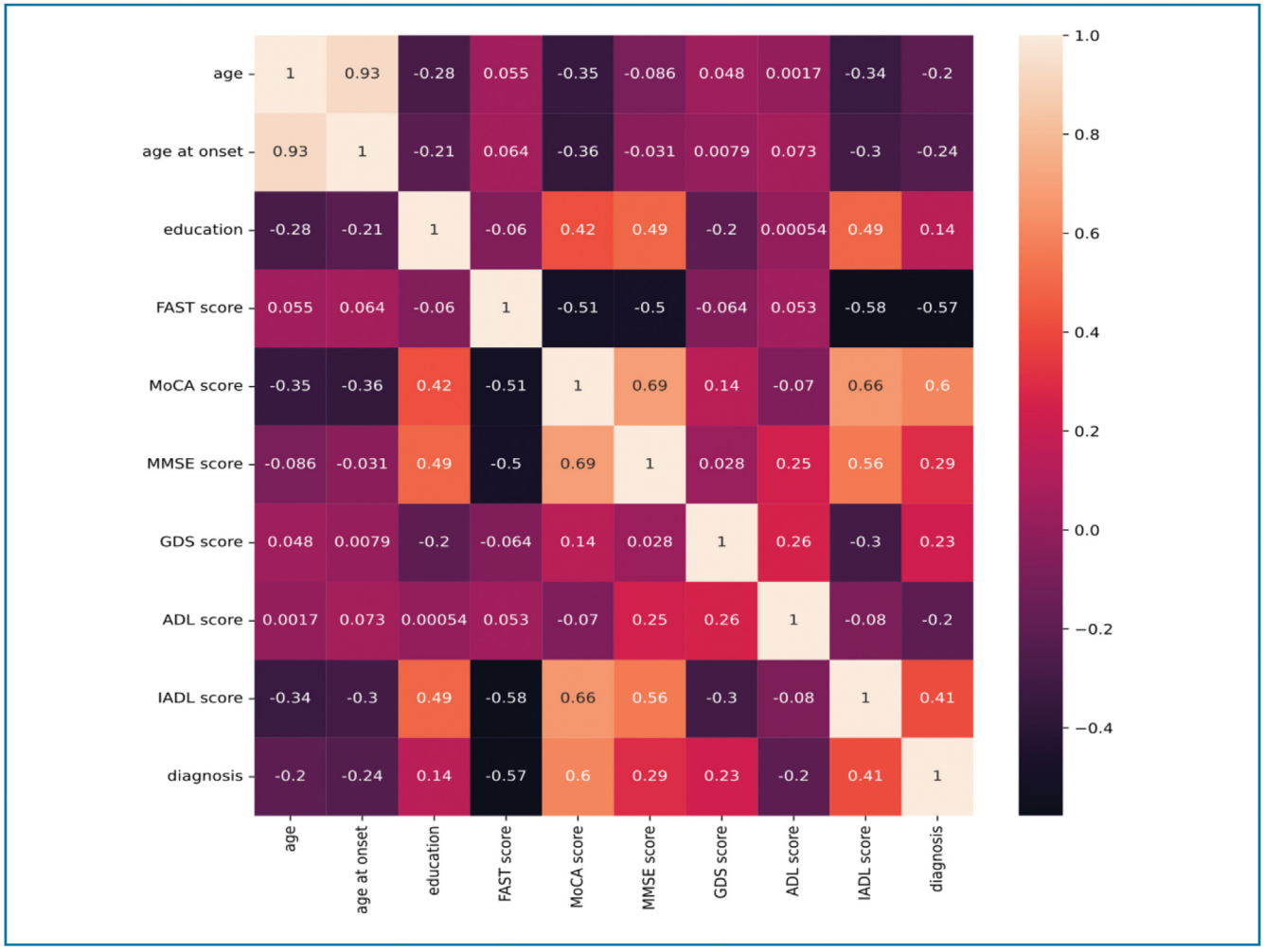


**Figure 3.** Heatmap showing the statistical relationships between different variables.

In this section, we delineate the specific algorithms and techniques utilized in our study. Our methodology encompasses a range of machine-learning models to ensure comprehensive analysis. The chosen algorithms vary in complexity and underlying principles, enabling robust comparisons.

## Machine-learning algorithms

Machine-learning algorithms are the cornerstone of predictive modelling, identifying data patterns to make informed predictions or decisions without explicit programming. The selection of the appropriate algorithm is crucial, as it can significantly impact a model's performance. In this section, we delve into the specific algorithms employed in our study.

### Decision tree

Decision trees split the data into subsets on the basis of the feature values. This process is repeated recursively, resulting in a tree-like model of decisions [38]. Decision trees are interpretable and can capture non-linear relationships in the data, which can be pivotal in detecting Alzheimer's disease and MCI [39].

### Random forest

Random forest is an ensemble learning method that operates by constructing multiple decision trees during the training phase and outputting the class that is the mode of the classes produced by individual trees for classification tasks [40]. In the context of Alzheimer's disease and MCI identification, the Random forest algorithm assesses the importance of various factors and makes predictions based on the collective decision of its trees. Given the potential complexity and non-linearity associated with disease predictors, Random forest provides a robust approach that can capture intricate patterns in the data, often outperforming simpler models like logistic regression [41].

### AdaBoost

AdaBoost, or Adaptive boosting, is an iterative ensemble method that adjusts the weights of misclassified elements, ensuring that subsequent classifiers focus on them. AdaBoost can increase the performance of the base classifiers, making it suitable for our task where small improvements in diagnostic accuracy can be critical [42].

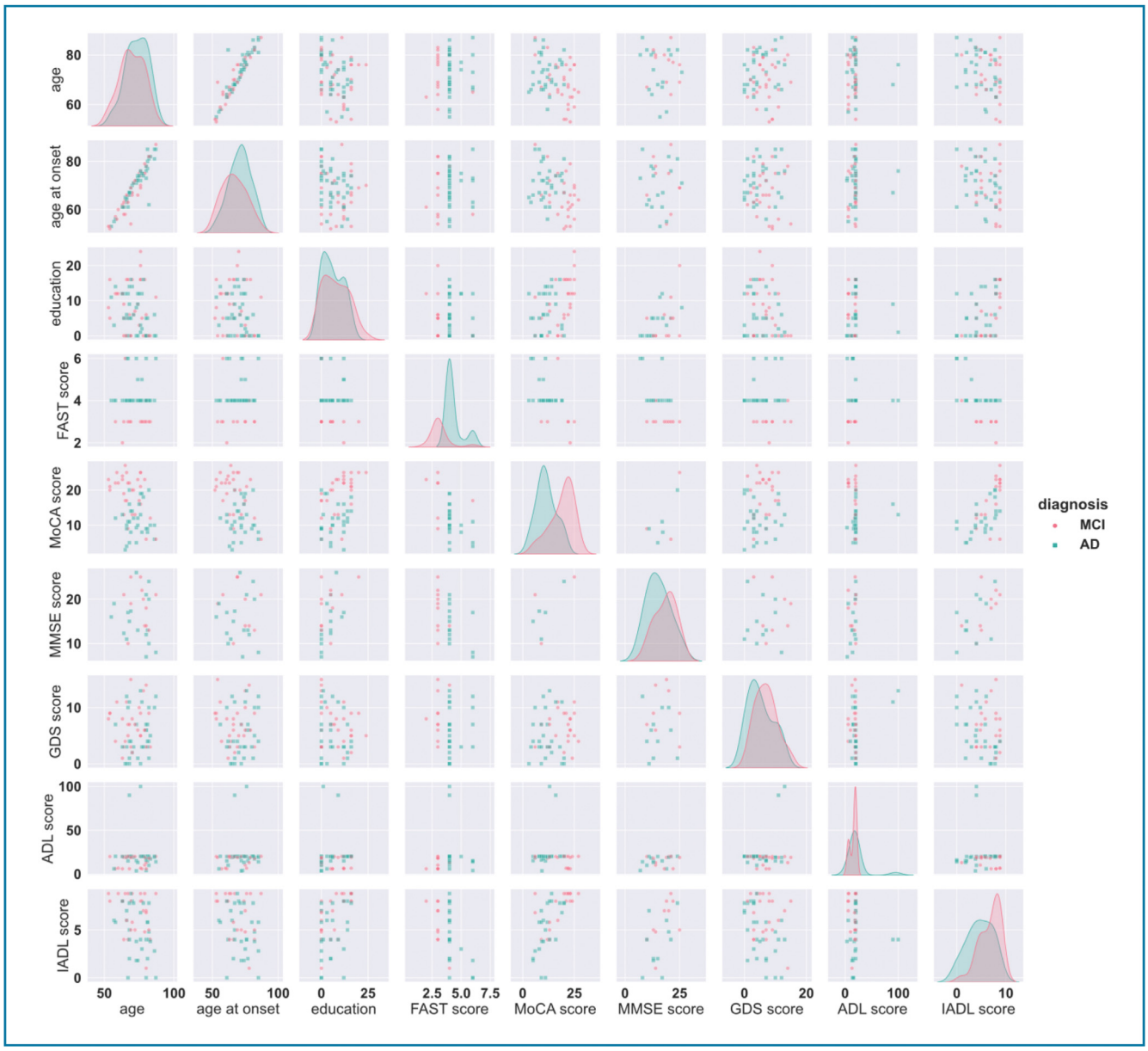


**Figure 4.** Pairplot of features with strong correlations.

### Extra trees

The Extra trees algorithm builds multiple decision trees and merges their outputs for a final result. Unlike random forests, it uses the whole dataset when building each tree and chooses split points entirely at random [43]. Extra Trees can capture more intricate patterns in data, potentially identifying subtle signals related to Alzheimer's disease and MCI [44].

### Bagging

Bagging, or Bootstrap aggregating, involves creating multiple subsets of the original dataset (with replacement) and training a model on each. The final prediction is an ensemble of the predictions from each model. As mentioned in the Cross-Validation section, Bagging with Decision trees was particularly successful in our research. It reduces variance and prevents overfitting, making it a strong choice for our task [45].

### SGD (Stochastic Gradient Descent)

SGD is an iterative optimization algorithm used to find the values of parameters that minimize a loss function. In the context of this research, it is used for large-scale learning. Given the potentially vast amount of data involved in this study, SGD can be an efficient way to train a model, especially if faced with computational constraints [46].

## Model training

Model training involves feeding the pre-processed and augmented data into each of the algorithms mentioned above, tuning their parameters for optimal performance.

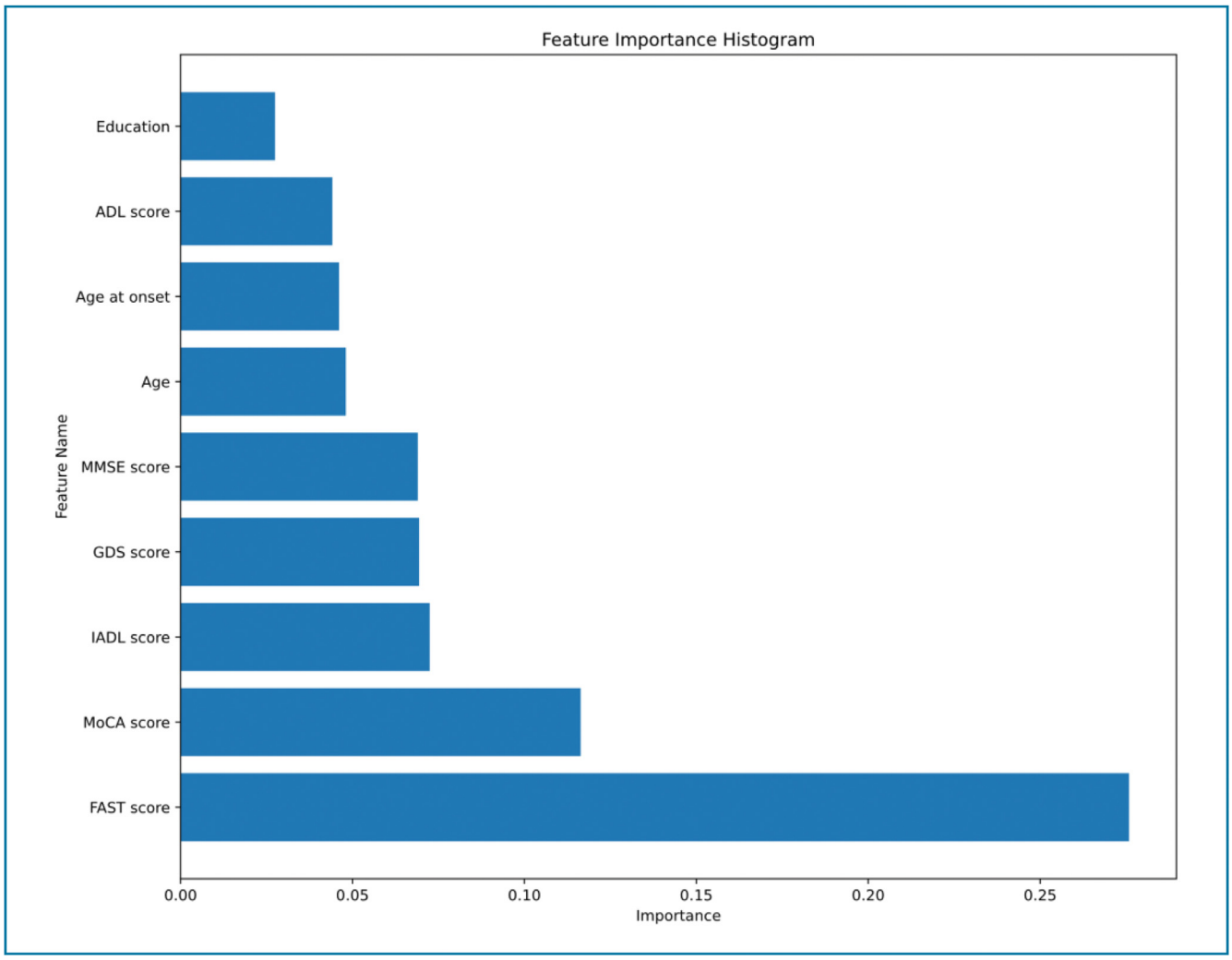


**Figure 5.** Histogram for feature importance.

Procedure:

- data partitioning: before training, the data is split into training and validation sets. This ensures we have a separate set of data to evaluate the model's performance;
- hyperparameter tuning: for each algorithm, a grid search or random search is employed to identify the best hyperparameters that result in the highest validation performance;
- model fit: using the training set and the optimal hyperparameters identified, each model is trained. During this phase, the model learns the underlying patterns and relationships in the data;
- early stopping: to prevent overfitting, especially in algorithms like SGD, an early stopping criterion can be used. This stops the training process if the model's performance on the validation set starts deteriorating;
- ensemble methods: for ensemble techniques like AdaBoost, Bagging, and Extra trees, multiple base models are trained. Their predictions are then combined in a specified manner (e.g. weighted average, majority vote) to produce the final output;
- regularization: techniques such as L1 or L2 regularization can be applied to prevent overfitting, especially in models like SGD.

By the end of the training process, each algorithm will have generated a model that can predict whether a given set of cognitive scores indicates Alzheimer's Disease, MCI, or a healthy state. The performance of these models is then evaluated using the validation set and metrics mentioned in the original research.

## Model evaluation

### Performance metrics

To ensure our models were performing optimally, we implemented a range of performance metrics to evaluate their efficacy. These metrics served as a yardstick to determine how well our models classified AD and MCI based on the given feature set [47]:

- accuracy: this metric gave us an overall understanding of how often the model made correct predictions across both classes;
- precision: precision allowed us to measure how many of the predicted positive cases (either AD or MCI, depending on the classification task) were actual positives;

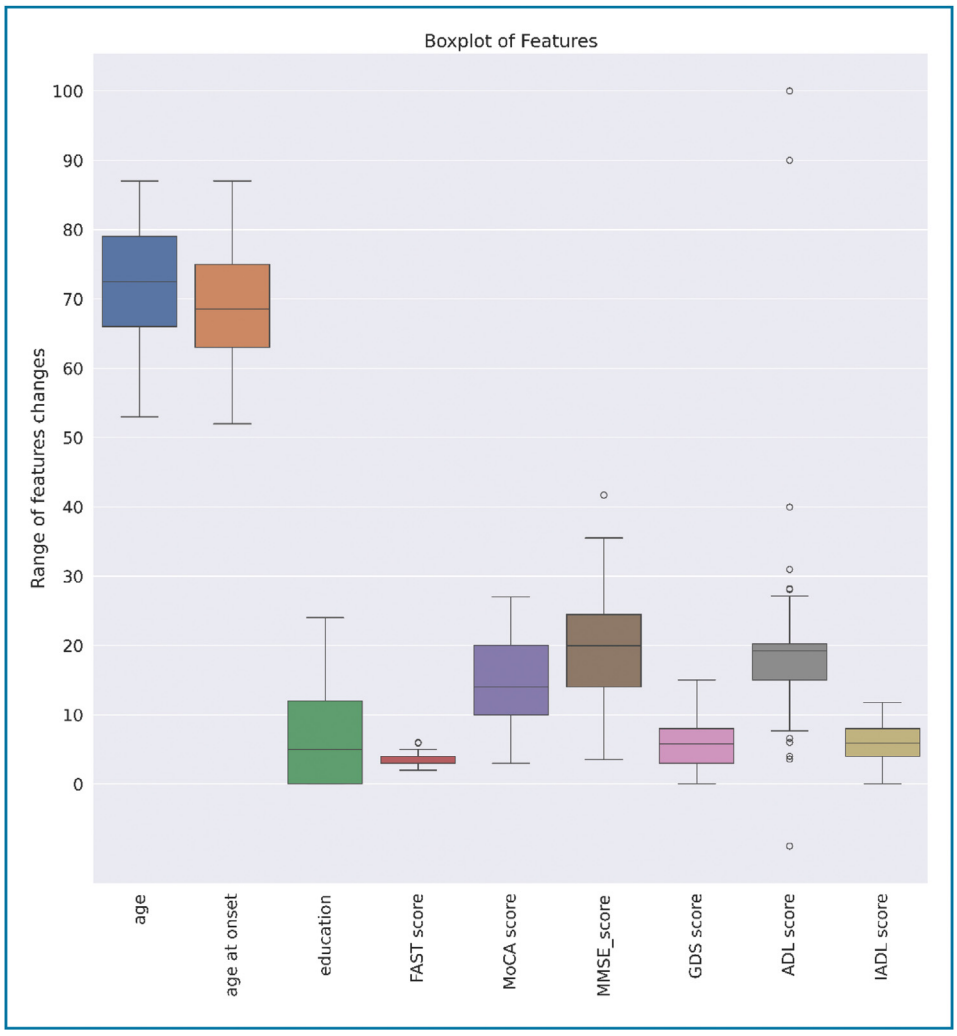


**Figure 6.** Boxplot for feature importance.

- recall: recall, or sensitivity, quantified how many of the actual positive cases were correctly predicted by our models;
- F1 score: as a weighted average of precision and recall, the F1 score provided a metric that balanced the trade-offs between false positives and false negatives;
- area under curve (AUC): this metric, derived from the Receiver Operating Characteristic (ROC) curve, gave us an indication of the model's ability to distinguish between the positive and negative classes, with values closer to 1 indicating superior performance.

## Cross-validation

Recognizing the risk of overfitting and the need for a robust assessment of our model's generalization abilities, we utilized a 5-subset cross-validation strategy:

- the entire dataset was partitioned into five equally-sized subsets;
- for each iteration of the validation, four of these subsets served as the training set while the remaining one was used for testing;
- by cycling through all possible combinations of training and testing subsets, we ensured that every subset was used both for training and validation. This approach reduced bias and variance, leading to more reliable performance metrics.

The metrics obtained from each stage were then averaged to produce a comprehensive evaluation of the model's performance.

Fig. 7 displays five confusion matrices corresponding to five different stages of the cross-validation process used in evaluating the performance of predictive models for Alzheimer's disease diagnosis and MCI identification. Each matrix is a 2 $\times$ 2 grid representing the counts of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions made by the model. Class '0' represents the negative class (no disease/MCI), and class '1' the positive class (presence of disease/MCI). The level 0, for instance, shows a perfect classification with no misclassifications. 1 to 4 displays various degrees of misclassifications, indicating some variability in the model's performance across different subsets of the data.

The confusion matrices suggest that the model generally performs well, with a large number of correct predictions (TP and TN) across all levels. Level 0 shows 100% accuracy,

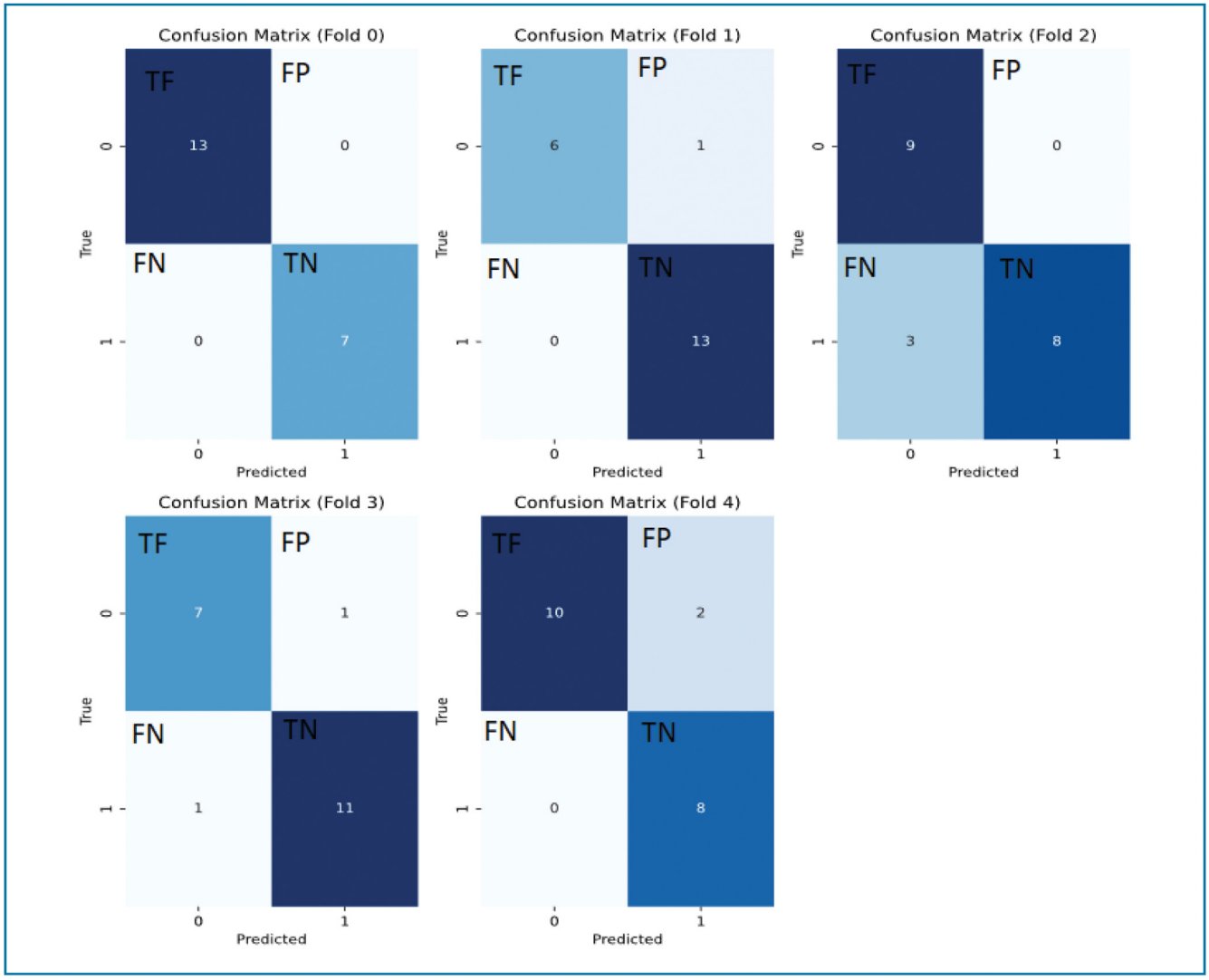


**Figure 7.** Series of five confusion matrices representing the results of a binary classifier tested on five different subsets of data.

which may indicate overfitting or a particularly favourable split of the data. However, other levels exhibit some errors, predominantly in the form of false negatives (e.g. 2 with 3 FNs) rather than false positives. The variability from one subset to another highlights the importance of cross-validation in assessing model robustness since it reveals that the model's performance can fluctuate depending on the data subset it is tested on. The models seem to have a stronger predictive capability for the positive class (class '1') as evidenced by lower false positives compared to false negatives, but there is a concern about consistency, as some subsets like 3 show a more balanced misclassification.

## ROC curves

The Receiver Operating Characteristic (ROC) curve is a graphical representation that plots the true positive rate (recall) against the false positive rate for various threshold settings of a classification model [16]. It provides a comprehensive view of the model's performance across a range of decision thresholds.

- by assessing the area under the ROC curve (AUC), we gained insights into the model's discriminant power. An AUC value of 1 would indicate perfect classification, while 0.5 suggests the model is no better than random guessing;
- through the ROC curves, we also identified optimal threshold values for our classification tasks, which could be critical when balancing the trade-off between sensitivity and specificity, especially in medical applications where false negatives and false positives can have significant consequences.

Our model evaluation methodology was designed to be comprehensive, ensuring the models not only fit the training data but also possess strong generalization scope when faced with new data. The combination of performance metrics, cross-validation, and ROC analysis laid a solid foundation for accurate and reliable AD and MCI classification.

Fig. 8 shows a ROC curve, which is commonly used to evaluate the performance of a binary classifier system. The graph plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. There are five curves representing five different data subsets in a

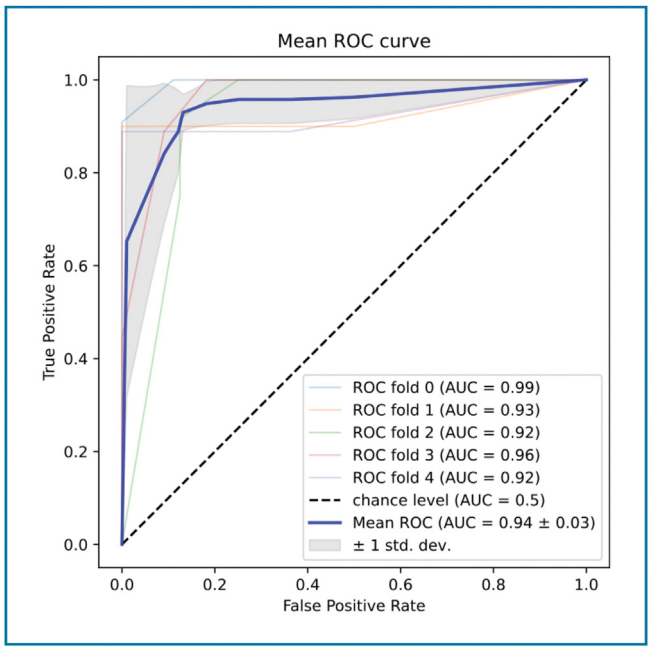


**Figure 8.** ROC for the best model.

cross-validation process, each with its AUC value ranging from 0.92 to 0.99, which indicates high performance. The dashed line represents the chance level (AUC = 0.5), which is the expected result of a random classifier. The blue line represents the mean ROC curve across all datasets with an AUC of $0.94 \pm 0.03$, suggesting a strong and stable model across different data partitions. The shaded area around the mean ROC curve indicates the standard deviation across datasets.

### Enhanced description of the ultimate selected model: the Bagging classifier

In our study, the Bagging classifier emerged as the most effective model for diagnosing Alzheimer's disease and identifying mild cognitive impairment. The Bagging (Bootstrap aggregating) algorithm enhances model accuracy and prevents overfitting by combining the predictions of multiple base decision tree models trained on different subsets of the original dataset. These subsets are created through random sampling with replacement.

Operational details:

- algorithm type: ensemble learning method, specifically Bootstrap aggregating (Bagging);
- base model: decision trees;
- training process: multiple decision tree models are independently trained on randomly sampled subsets of the training data. Each subset is created with replacement, ensuring diverse training scenarios for the base models;
- decision making: the Bagging classifier aggregates the predictions from all individual decision trees to make a final decision. For classification tasks, this usually involves a simple majority vote mechanism;
- hyperparameters and settings: while specific hyperparameters such as the number of trees in the ensemble, the depth of individual trees, and the criteria for splitting nodes are crucial, they were optimized for our dataset to achieve the highest predictive performance.

This methodology ensures robustness against overfitting, making the Bagging classifier particularly suited for our predictive modelling task. It demonstrated superior performance across various evaluation metrics, including accuracy, precision, recall, F1 score, and the area under the ROC curve, outperforming other tested algorithms such as Random forest, AdaBoost, Extra trees, Decision tree, and SGD classifiers.

Our application of the Bagging classifier led to a customised configuration of hyperparameters optimized through cross-validation to maximize the model's performance on our dataset. This involved fine-tuning the number of decision trees and their maximum allowed depth, among other parameters, to ensure an optimal balance between bias and variance, thereby enhancing the model's generalizability to new data.

## Experimental results

In this section, we present the outcomes of our experiments conducted to assess the efficacy of our proposed models. We detail the process of data partitioning, performance on diagnosing Alzheimer's disease and on identifying MCI, and the comparison with other algorithms. We also explore a thorough interpretation of the results obtained.

### Exploratory data analysis and data splitting

To ensure the validity and generalizability of our models, we split the dataset into training and testing subsets. The splitting was performed in a manner that ensured both subsets contained a representative distribution of AD and MCI instances. Using a stratified approach, we allocated 80% of the data for training and the remaining 20% for testing. The stratified approach was crucial to maintain the proportionality of AD and MCI instances, ensuring that both the training and testing datasets were representative of the overall data.

### Model performance on MCI identification

For the classification of MCI, Random forest, AdaBoost, and Extra trees classifiers were top performers, with Extra trees slightly edging out the others. Decision tree and SGD provided a slightly lower performance in terms of MCI identification. The Bagging classifier, which uses a combination of models to make decisions, presented itself as a particularly promising tool, achieving a good balance between accuracy, precision, recall, and AUC.

### Comparison of algorithms

Upon close examination of the performance metrics, we found that the Bagging classifier achieved the highest mean accuracy and F1 score, making it the most robust model in our experimentation. Random forest, AdaBoost, and Extra trees were close contenders, offering competitive results. Decision tree and SGD trailed behind but still presented respectable performances. The disparity in results underscores

**Table 2** Comparison of performance metrics across various machine-learning classifiers.

| Classifier | Mean accuracy | Mean F1 Score | Mean precision | Mean AUC |
|---|---|---|---|---|
| Random forest | 0.88 | 0.87 | 0.90 | 0.88 |
| AdaBoost | 0.88 | 0.88 | 0.87 | 0.88 |
| Extra trees | 0.89 | 0.89 | 0.90 | 0.89 |
| Bagging | 0.9 | 0.90 | 0.90 | 0.90 |
| Decision tree | 0.84 | 0.84 | 0.86 | 0.84 |
| SGD | 0.85 | 0.85 | 0.84 | 0.84 |

the importance of choosing the right classifier for specific tasks, especially in critical domains. Table 2 compares the performance of various machine-learning classifiers. The classifiers listed include Random forest, AdaBoost, Extra trees, Bagging, Decision tree, and SGD (Stochastic Gradient Descent). For each classifier, four metrics are reported: Mean accuracy, Mean F1 score, Mean precision, and Mean AUC (Area under the ROC curve). These metrics are commonly used to evaluate the performance of predictive models, with higher values indicating better performance. The Bagging classifier shows the highest Mean accuracy and Mean F1 score, suggesting it may be the most effective model among those listed for the prediction.

Table 2 indicates that ensemble methods like Random forest, Extra trees, and Bagging are performing well on the task, with Bagging slightly outperforming the others in terms of accuracy and F1 score, both of which are critical for evaluating the performance of classification models in medical diagnostics. A high F1 score suggests a balance between precision and recall, which is important in the medical field where both false positives and false negatives can have serious consequences. The AUC values are also high for these models, indicating good discrimination between positive and negative classes. The Decision tree and SGD have lower performance metrics across the board, which could suggest that the dataset is complex and benefits from the more nuanced decision boundaries that those ensemble methods and more sophisticated algorithms can provide. It is important to note, however, that the difference in performance metrics is relatively small, suggesting that all models have merit and that further validation is necessary to determine the most practical model for deployment in a clinical setting.

Fig. 9 presents the performance metrics for different classifiers. The performance metrics are shown on the vertical axis, ranging from 0.0 to above 0.8, indicating the scores of the classifiers. There are four different coloured bars for each classifier, representing Accuracy, Mean F1 Score, Mean Precision, and Mean AUC, respectively. All classifiers appear to perform similarly on these metrics, with each of the scores generally above 0.8, suggesting that the models are fairly good at predicting outcomes.

### Interpretation of results

The results highlight the potential of machine-learning algorithms in assisting with the diagnosis of Alzheimer's disease and MCI. The accuracy, precision, recall, and AUC values indicate a promising direction for the adoption of these models in clinical settings. Particularly, the Bagging classifier, with its ensemble approach, emerged as a frontrunner, indicating the power of combining multiple models for improved performance. Furthermore, the critical features identified, such as FAST Score and MoCA Score, emphasize the importance of cognitive assessments in the early detection of AD and MCI. The models' performances also underscore the significance of a comprehensive approach to data handling, pre-processing, and augmentation. While all models presented satisfactory results, the variations in their performances suggest that a combination of models, perhaps in an ensemble manner, might be the optimal approach for future research and applications in the field of Alzheimer's disease diagnosis.

## Discussion

### Interpretation of findings

The findings from the study provide insight into the factors influencing AD and MCI. The data indicates a higher prevalence of AD in females, as shown in the gender distribution across AD and MCI diagnoses. This suggests that gender may be a significant factor in the progression of these conditions. The close ages of onset in the AD group imply that Alzheimer's symptoms develop after their emergence. In contrast, the MCI group, with a younger average onset age and higher education levels, suggests that educational background might play a role in early diagnosis or in delaying the progression of cognitive impairments.

The analysis of various clinical and demographic variables through histograms, heatmaps, and matrix plots reveals intricate relationships between these factors and their impact on AD and MCI. Cognitive scores like MoCA and MMSE are skewed towards the lower end, indicating significant cognitive impairments in the sample. Interestingly, these cognitive assessments correlate negatively with the FAST score, which measures stages of AD, indicating that lower cognitive functioning is associated with advanced disease stages. Additionally, the positive correlation between educational levels and cognitive scores hints at the potential protective role of education against cognitive decline.

The predictive models evaluated through cross-validation and various performance metrics demonstrate a robust ability to distinguish between AD and MCI. The ensemble methods like Random forest and Bagging show particularly strong performances, suggesting that these approaches are effective for handling the complexities of AD and MCI prediction. The ROC curves further corroborate the models' efficacy, with high AUC values indicating their strong discriminant ability. However, the room for improvement, as

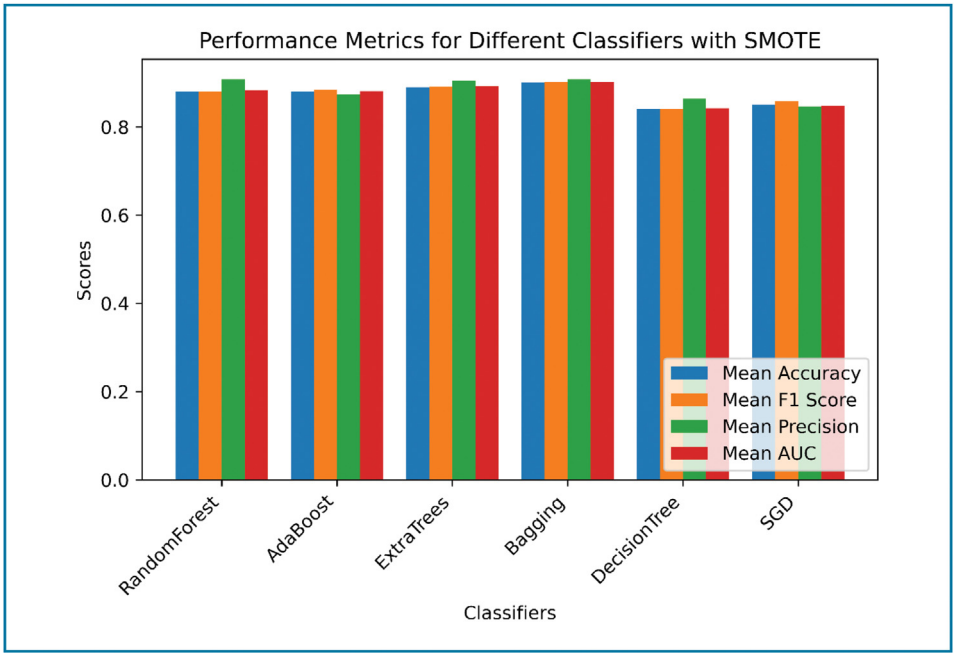


**Figure 9.** Model accuracy scores using barplots.

indicated by the AUC values not being perfect, points towards the need for ongoing refinement of these models, possibly by integrating more diverse data or optimizing algorithms.

## Implications for Alzheimer's disease diagnosis

The research presented in the paper offers significant implications for the diagnosis of AD and MCI. First and foremost, the use of various predictive models, as indicated by the diverse set of machine-learning classifiers evaluated, shows promise in enhancing the accuracy and timeliness of AD and MCI diagnosis. The inclusion of demographic factors like age and gender, alongside clinical assessments such as cognitive scores (FAST, MoCA, MMSE), implies a more comprehensive approach to diagnosis. The gender differences observed, particularly the higher prevalence of AD among females and the balanced distribution of MCI across genders, suggest that gender-specific factors might be crucial in understanding and diagnosing these conditions. Furthermore, the relationship between educational level and cognitive scores could indicate the potential role of cognitive resources in delaying the onset or progression of symptoms, a valuable observation for early intervention strategies.

The statistical analyses and visualizations such as correlation heatmaps and boxplots, underscore the complexity of diagnosing AD and MCI. The correlation between various cognitive and functional scores reveals the intricate interplay of factors contributing to these conditions. The diversity in cognitive scores and the prominence of functional decline, as seen in the significance of the FAST score in predictive modelling, highlight the need for multifaceted diagnostic tools that can capture the nuanced progression of AD and MCI. Moreover, the consistent performance of ensemble methods in machine-learning classifiers suggests that a combination of algorithms could be more effective in handling the complexity of data associated with AD and MCI, thereby improving diagnostic accuracy.

The performance metrics of the predictive models, especially the high AUC values observed in the ROC curves, demonstrate the potential of artificial intelligence algorithms in enhancing diagnostic processes. These models not only provide a framework for early and accurate diagnosis but also enable possibilities for personalized medicine approaches in treating AD and MCI. The ability to distinguish between AD and MCI with considerable accuracy could lead to more targeted and effective treatment plans, tailored to the individual's specific condition and stage of progression. However, it is important to note that while the models show promise, the presence of outliers and variability in performance across different subsets indicate the need for continuous refinement and validation of these models in diverse and larger populations to ensure their reliability and applicability in clinical settings.

## Implications for MCI identification

The data presented in this study has significant implications for the identification of MCI. The balanced gender distribution for MCI, as shown in Fig. 1, suggests that the progression from MCI to AD might be influenced by factors other than gender, such as genetic, environmental, or lifestyle factors. This finding is crucial for developing more targeted predictive models that consider these factors. Additionally, the demographic insights from Table 1, indicating a younger average age and higher educational level in the MCI group, suggest that earlier diagnosis and intervention could be possible and beneficial. The implication here is that individuals with higher education levels might be more aware of cognitive changes and seek medical advice sooner, thus getting diagnosed with MCI earlier. This highlights the importance of raising awareness and education about cognitive health

across all demographic groups to facilitate early detection and management of MCI.

The analysis of various clinical and demographic variables in Figs. 2 and 4 further reinforces the complexity of MCI identification. The distributions of cognitive scores, functional status, and educational levels indicate that a multitude of factors play a role in the onset and progression of MCI. This complexity is underscored by the varied performance of machine-learning classifiers in Table 2 and Fig. 8, which suggests that no single model or variable can conclusively predict MCI. The high relevance of the FAST and MoCA scores in the predictive model, as indicated in Fig. 5, points to the significance of functional decline and cognitive performance in MCI detection. However, the minimal influence of education level on the model's predictions implies that clinical assessments are more crucial than demographic factors in diagnosing MCI. The consistent performance across different cross-validation subsets in Fig. 7 indicates that the models have good generalizability, making them reliable tools for MCI identification. This underscores the potential of artificial intelligence algorithms in enhancing predictive accuracy for MCI, thereby facilitating early intervention and possibly delaying or preventing the progression to AD.

## Comparison with existing models in literature

In this study, the Bagging classifier demonstrated superior performance in diagnosing Alzheimer's disease and identifying mild cognitive impairment, with notable accuracy, sensitivity, specificity, and AUC values. To contextualize our model's performance, we compared these results with similar studies in the field. For instance, Sarica et al. [41] explored the Random forest algorithm for classifying neuroimaging data in Alzheimer's disease, achieving an accuracy of 88% and an AUC of 0.87, which is comparable to our findings with the Bagging classifier (accuracy of 90%, AUC of 0.90). However, our model shows a slight improvement, possibly due to the diverse dataset and advanced feature selection process employed.

Furthermore, the study by Grassi [48] using an ensemble-based machine-learning algorithm reported an accuracy of 85%, highlighting the potential of ensemble methods in enhancing predictive performance. Our research supports this notion, with the Bagging classifier outperforming individual models like Decision trees and SGD, which aligns with the findings by Özçift [40] in cardiac arrhythmia diagnosis using ensemble classifiers.

By comparing these metrics, we highlight the robustness of our model within the landscape of Alzheimer's disease diagnostics. This comparison not only validates the effectiveness of our approach but also contributes to the ongoing dialogue on the most promising machine-learning techniques for early and accurate AD/MCI detection.

## Limitations of the study

The study, while offering valuable insights into the predictive modelling of AD and MCI, presents several limitations. Firstly, the gender distribution in the dataset leans more towards females, potentially limiting the generalizability of the findings across genders. Secondly, the skewness towards higher education in the participant pool could have influenced the predictive model's performance, as this might not accurately reflect the broader population's education levels. Additionally, the presence of outliers on certain variables like age and IADL score could impact the robustness of the predictive models. The variability in model performance across different cross-validation subsets, particularly the occurrence of false negatives, points to potential issues in consistency and reliability. Furthermore, while the models show high predictive performance, they are not perfect, suggesting a need for further refinement, possibly through the inclusion of more diverse demographic and clinical variables, or by optimizing the algorithms used. The study's reliance on specific clinical scores, though informative, may not capture the entire spectrum of factors influencing AD and MCI, highlighting the need for a more holistic approach in future research.

While our study has provided valuable insights into the predictive accuracy of various machine-learning models for AD and MCI diagnosis, we acknowledge a limitation in the scope of our validation process. The current study did not include external validation on a completely independent sample. This step is crucial for assessing the generalizability of our findings across diverse populations and settings. Future research efforts will aim to incorporate this critical phase of model validation to ensure our algorithms maintain their predictive accuracy and reliability when applied to broader, novel datasets.

## Future research directions

The study's findings create opportunities for future research in the diagnosis and progression modelling of AD and MCI. A critical area to explore is the role of gender and other demographic factors in the progression from MCI to AD, as the current data suggests gender-based differences in prevalence and disease characteristics. Further investigation into the protective effects of education and cognitive resources on disease onset and progression is warranted, given the observed correlations. Additionally, expanding the dataset to include more diverse populations and variables, such as genetic markers or lifestyle factors, could enhance the predictive models' accuracy and generalizability. The application of advanced machine-learning techniques, like deep learning, could also be explored to capture more complex patterns in the data. Longitudinal studies would provide deeper insights into the progression dynamics of AD and MCI, potentially leading to earlier and more precise diagnoses and tailored intervention strategies.

We plan to extend our research to include external datasets that encompass a broader spectrum of dementia conditions, such as vascular dementia, Frontotemporal dementia (FTD), and Lewy body disease. This future work aims to enhance the model's discriminative power, enabling more precise differentiation between AD, MCI, and other dementias. This clarification and our commitment to ongoing research address the current limitations and outline our approach to enhancing the model's applicability to a wider range of neurodegenerative diseases.

## Conclusion

This study has provided significant insights into AD and MCI through the application of machine-learning models. The analysis highlighted the distribution of AD and MCI across genders, with a higher incidence of AD among females. Key demographic factors, such as age, education, and onset age, along with cognitive scores, were identified as influential in predicting the progression from MCI to AD. The ensemble methods, particularly Bagging, demonstrated superior performance in predicting AD and MCI, emphasizing the complexity and nuanced nature of these conditions. The study's integration of cognitive scores and artificial intelligence algorithms for AD and MCI identification is a significant contribution to the field. By employing various machine-learning classifiers and evaluating their performance through confusion matrices, ROC curves, and importance plots, the research provides a comprehensive view of how these models can enhance our understanding and prediction of cognitive impairments. The findings have practical applications in the medical field, particularly in early diagnosis and intervention strategies for AD and MCI. The predictive models can assist clinicians in identifying individuals at higher risk, potentially leading to earlier intervention and better management of these conditions. Additionally, the insights into demographic and cognitive factors can inform public health policies and educational campaigns targeting at-risk groups. In conclusion, this study underscores the potential of combining cognitive assessments with advanced AI techniques to improve the diagnosis and understanding of Alzheimer's disease and mild cognitive impairment. While the findings are promising, they also highlight the need for further research and refinement of these models to ensure their efficacy and reliability in diverse clinical settings.

## Disclosure of interest

The authors declare that they have no competing interest.

## Références

[1] Sadegh-Zadeh S-A. Computational methods toward early detection of neuronal deterioration. Thesis in Philosophy in Computer Science, University of Hull, January 2019. file:///C:/Users/566693/Downloads/content-hull_17355a.pdf.
[2] Sadegh-Zadeh S-A, Kambhampati C. Computational investigation of amyloid peptide channels in Alzheimer's disease. J Basel 2018;2(1):1—14.
[3] Morabito FC, Ieracitano C, Mammone N, et al. An explainable artificial intelligence approach to study MCI to AD conversion via HD-EEG processing. Clin EEG Neurosci 2023;54(1):51—60.
[4] Culberson JW, Kopel J, Sehar U, et al. Urgent needs of caregiving in ageing populations with Alzheimer's disease and other chronic conditions: support our loved ones. Ageing Res Rev 2023;90:102001.
[5] Elzayat EM, Shahien SA, El-Sherif AA, et al. miRNAs and stem cells as promising diagnostic and therapeutic targets for Alzheimer's disease. J Alzheimers Dis 2023;94(s1):S203—25.
[6] Khosroazad S, Abedi A, Hayes MJ, et al. Sleep signal analysis for early detection of Alzheimer's disease and related dementia (ADRD). IEEE J Biomed Health Inform 2023;27(5):2264—75.
[7] Deng Y, Wang H, Gu K, et al. Alzheimer's disease with frailty: prevalence, screening, assessment, intervention strategies and challenges. Biosci Trends 2023;17(4):283—92.
[8] Roy RG, Mandal PK, Maroon JC, et al. Oxidative stress occurs prior to amyloid Aβ plaque formation and tau phosphorylation in Alzheimer's disease: role of glutathione and metal ions. ACS Chem Neurosci 2023;14(17):2944—54.
[9] Sadegh-Zadeh S-A, Kambhampati C, Davis DN. Ionic imbalances and coupling in synchronization of responses in neurons. J Basel 2019;2(1):17—40.
[10] Pang Y, Kukull W, Sano M, et al. Predicting progression from normal to MCI and from MCI to AD using clinical variables in the national Alzheimer's coordinating center uniform data set version 3: Application of machine learning models and a probability calculator. J Prev Alzheimers Dis 2023;10(2):301—13.
[11] Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. Int Psychogeriatr 1992;4(3):55—69.
[12] Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695—9.
[13] McDowell I, Kristjansson B, Hill GB. Community screening for dementia: the mini mental state exam (MMSE) and modified mini-mental state exam (3MS) compared. J Clin Epidemiol 1997;50(4):377—83.
[14] Millán-Calenti JC, Tubío J, Pita-Fernández S, et al. Prevalence of functional disability in activities of daily living (ADL), instrumental activities of daily living (IADL) and associated factors, as predictors of morbidity and mortality. Arch Gerontol Geriatr 2010;50(3):306—10.
[15] Pashmdarfard M, Azad A. Assessment tools to evaluate Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) in older adults: A systematic review. Med J Islam Repub Iran 2020;34:33.
[16] Pocklington C, Gilbody S, Manea L, et al. The diagnostic accuracy of brief versions of the Geriatric Depression Scale: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2016;31(8):837—57.
[17] Sadegh-Zadeh S-A, Qeranqayeh AR, Benkhalifa E, et al. Dental caries risk assessment in children 5 years old and under via machine learning. Dent J Basel 2022;10(9):164.
[18] Sadegh-Zadeh SA, Fakhri E, Bahrami M, et al. An approach toward artificial intelligence Alzheimer's disease diagnosis using brain signals. Diagnostics 2023;13(3):477.
[19] Abed S, Mourad M, Abdelkreem MM, et al. AD Classification and Detection Using Neuroimaging Data. 2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, September 2023, 1—8.
[20] Sheng J, Xin Y, Zhang Q, et al. Predictive classification of Alzheimer's disease using brain imaging and genetic data. Sci Rep 2022;12(1):2405.
[21] Alexander N, Alexander DC, Barkhof F, et al. Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning. BMC Med Inform Decis Mak 2021;21(1):1—13.
[22] Grassi M, Rouleaux N, Caldirola D, et al. A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. Front Neurol 2019;10:463970.
[23] Edmonds EC, Delano-Wood L, Galasko DR, et al. Subtle cognitive decline and biomarker staging in preclinical Alzheimer's disease. J Alzheimers Dis 2015;47(1):231—42.

[24] Porsteinsson AP, Isaacson RS, Knox S, et al. Diagnosis of early Alzheimer's disease: clinical practice in 2021. J Prev Alzheimers Dis 2021;8(3):371—86.
[25] Lee M-T, Jang Y, Chang W-Y. How do impairments in cognitive functions affect activities of daily living functions in older adults? PLoS One 2019;14(6):e0218112.
[26] Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, et al. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's disease neuroimaging initiative. Arch Gen Psychiatry 2011;68(9):961—9.
[27] McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984;34(7):939.
[28] Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7(3):280—92.
[29] Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. J Am Geriatr Soc 1992;40(9):922—35.
[30] Sainsbury A, Seebass G, Bansal A, et al. Reliability of the Barthel Index when used with older people. Age Ageing 2005;34(3):228—32.
[31] Lawton MP. Instrumental activities of daily living (IADL) scale. Psychopharmacol Bull 1988;24:785—7.
[32] Enders CK. Multiple imputation as a flexible tool for missing data handling in clinical research. Behav Res Ther 2017;98:4—18.
[33] Figueroa-García JC, Neruda R, Hernandez—Pérez G. A genetic algorithm for multivariate missing data imputation. Inf Sci (N Y) 2023;619:947—67.
[34] Sadegh-Zadeh S-A, Bahrami M, Najafi A, et al. Evaluation of COVID-19 pandemic on components of social and mental health using machine learning, analysing United States data in 2020. Front Psychiatry 2022;13:933439.
[35] Zheng A, Casari A. Feature engineering for machine learning: principles and techniques for data scientists. Sebastopol, California: O'Reilly Media, Inc.; 2018.
[36] Sadegh-Zadeh SA, Sakha H, Movahedi S, et al. Advancing prognostic precision in pulmonary embolism: A clinical and laboratory-based artificial intelligence approach for enhanced early mortality risk stratification. Comput Biol Med 2023;167:107696.
[37] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321—57.
[38] Saarela M, Jauhiainen S. Comparison of feature importance measures as explanations for classification models. SN Appl Sci 2021;3:1—12.
[39] Kadem M. Interpretable machine learning in Alzheimer's disease dementia. These, 2023. https://macsphere.mcmaster.ca/handle/11375/28503.
[40] Özçift A. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. Comput Biol Med 2011;41(5):265—71.
[41] Sarica A, Cerasa A, Quattrone A, et al. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci 2017;9:329.
[42] Rojas R. AdaBoost and the super bowl of classifiers: a tutorial introduction to adaptive boosting. Freie University, Berlin, Tech. Rep 2009;1(1):1—6 https://www.inf.fu-berlin.de/inst/ag-ki/adaboost4.pdf.
[43] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63:3—42.
[44] Sarica A, Aracri F, Bianco MG, et al. Conversion from mild cognitive impairment to Alzheimer's disease: a comparison of tree-based Machine Learning algorithms for Survival Analysis. In: International Conference on Brain Informatics. New-York: Springer; 2023. p. 179—90.
[45] Breiman L. Bagging predictors. Mach Learn 1996;24:123—40.
[46] Mandt S, Hoffman MD, Blei DM. Stochastic gradient descent as approximate Bayesian inference. J Mach Learn Res 2017;18:1—35.
[47] Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 2015;5(2):1.
[48] Grassi M, Rouleaux N, Caldirola D, et al. A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. Front Neurol 2019;10:463970.