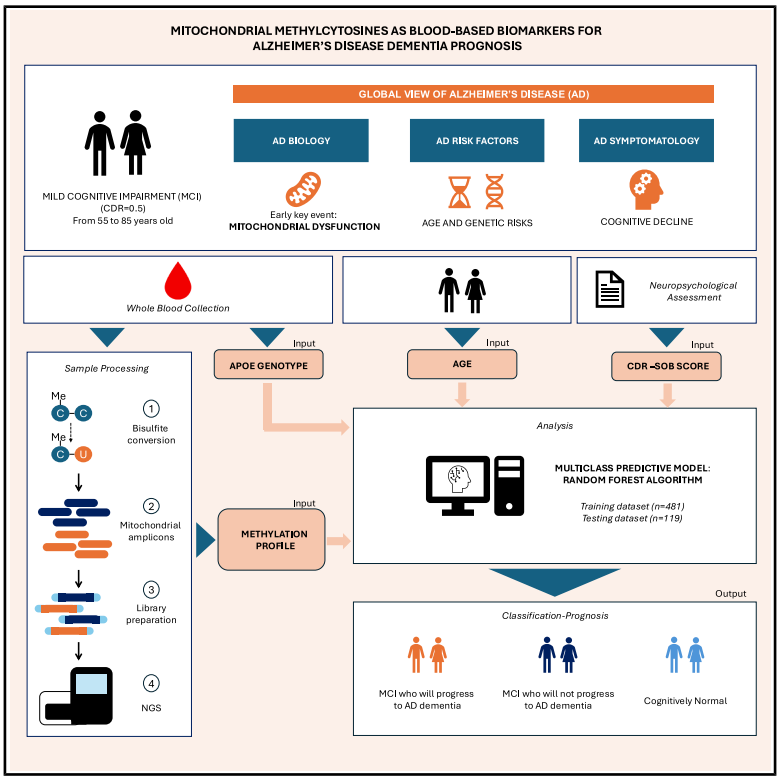
iScience

Article

Mitochondrial methylcytosines as blood-based biomarkers for Alzheimer's disease dementia prognosis

Graphical abstract



Highlights

- Mitochondrial DNA methylation is a blood biomarker for Alzheimer's dementia
- Changes in D-loop and *ND1* methylation are detected in the absence of β-amyloid positivity
- A dementia prognostic AI-algorithm with a holistic view of Alzheimer's dementia
- Our model predicts cognitive decline due to AD in patients with mild cognitive impairment

Authors

Jordi Gascón-Bayarri, Jose Luis Mosquera, Marta Blanch, ..., Russell H. Swerdlow, Ramón Reñé-Ramírez, Marta Barrachina

Correspondence

jlmosquera@admit-therapeutics.com (J.L.M.), mbarrachina@admit-therapeutics.com (M.B.)

In brief

Medicine; Neurology; Neuroscience; Molecular neuroscience

Gascón-Bayarri et al., 2025, iScience *28*, 113418
September 19, 2025
https://doi.org/10.1016/j.isci.2025.113418

CellPress

iScience

CellPress
OPEN ACCESS

Article

Mitochondrial methylcytosines as blood-based biomarkers for Alzheimer's disease dementia prognosis

Jordi Gascón-Bayarri,[1,13] Jose Luis Mosquera,[2,13,*] Marta Blanch,[2,13] Pau Martí,[2] Beatriz Fontal,[2] Carla Trapero,[2] Nuria Rojo,[1] Inma Rico,[1] Jaume Campdelacreu,[1] Cristopher Fowler,[3] Simon M. Laws,[4,5] Adrià Tort-Merino,[6,7] Raquel Sanchez-Valle,[6,7] Joan Bello,[8] Juan Fortea,[7,9] Alberto Lleó,[7,9] Courosh Mehanian,[10,11] Russell H. Swerdlow,[12] Ramón Reñé-Ramírez,[1] and Marta Barrachina[2,14,*]

[1]Functional Unit of Dementia, Service of Neurology, Bellvitge University Hospital, Bellvitge Biomedical Research Institute, IDIBELL, 08907 Barcelona, Spain
[2]ADmit Therapeutics SL, 08980 Barcelona, Spain
[3]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Melbourne, VIC 3010, Australia
[4]Collaborative Genomics and Translation Group, Centre for Precision Health, School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA 6027, Australia
[5]Collaborative Genomics and Translation Group, School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA 6027, Australia
[6]Alzheimer's Disease and Other Cognitive Disorders Unit, Neurology Service, Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, 08036 Barcelona, Spain
[7]Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), 28029 Madrid, Spain
[8]Neurology Service, Complex Hospitalari Moisès Broggi, 08906 L'Hospitalet/Sant Joan Despí, Spain
[9]Memory Unit, Department of Neurology, Institut de Recerca Sant Pau - Hospital de Sant Pau, Universitat Autònoma de Barcelona, 08025 Barcelona, Spain
[10]Global Health Labs, Bellevue, WA 98007, USA
[11]University of Oregon, Eugene, OR 97403, USA
[12]The University of Kansas Alzheimer's Disease Research Center, Fairway, KS 66205, USA
[13]These authors contributed equally
[14]Lead contact
*Correspondence: jlmosquera@admit-therapeutics.com (J.L.M.), mbarrachina@admit-therapeutics.com (M.B.)
https://doi.org/10.1016/j.isci.2025.113418

SUMMARY

Alzheimer's Disease Dementia (ADD) prognosis is an unmet medical need. Mitochondrial dysfunction is an early AD etiopathogenic factor. The present study analyzed mitochondrial DNA (mtDNA) methylation patterns in blood samples from patients with mild cognitive impairment (MCI) who progressed to ADD (P), MCI remained stable (NP), and Cognitively Normal (CN) individuals. Differentially methylated sites were identified in the D-loop region in both CN vs. NP and NP vs. P comparisons, even before β-amyloid positivity. A Random Forest model was developed using mtDNA methylation data combined with cognitive and risk factor features. Model's performance was assessed by cross-validation and tested on an independent set, achieving 84.4% accuracy in training and 83.2% (95% CI: 75.2%–89.4%) in testing. For identifying P patients, sensitivity and specificity were 95.1% and 70.7%, respectively. The AUC-ROC was 90.3%. The developed model demonstrates predictive capacity in distinguishing cognitive decline and stability in MCI individuals, independently of their β-amyloid status.

INTRODUCTION

Early and accurate diagnosis of or prognostication in Alzheimer's Disease (AD) is an unmet medical need. AD is a human neurodegenerative disease that progresses over time, manifesting three clinical stages: preclinical, prodromal, and dementia. The preclinical period can last nearly twenty years. Mild Cognitive Impairment (MCI) is considered the prodromal stage and is followed by dementia.[1] A diagnosis of AD is definitively confirmed when β-amyloid (senile and neuritic plaques) and tau (neurofibrillary tangles) are found in the postmortem brain[2] or *in vivo* by biomarkers.[3]

Patients with MCI are identified as having AD in the presence of a positive β-amyloid PET Scan or reduced β-amyloid levels in cerebrospinal fluid (CSF) and/or increased p-Tau and total tau in CSF. These patients are of interest to clinical trials as they are in the early AD stages.[4] However, the invasiveness and cost of these techniques or their insufficient accessibility limit their use.[5] Great efforts are being made to develop blood-based

biomarkers for AD. Blood β-amyloid levels can, to some extent, help predict the presence of cerebral β-amyloid.[6] Interestingly, the measurement of blood p-Tau217 also informs the status of brain amyloidosis.[7] However, there is no blood-based biomarker that predicts the progression to AD dementia (ADD) in patients with MCI in early stages, i.e., patients with MCI who are β--amyloid negative initially but who progress to ADD after a clinical follow-up and who subsequently convert to β-amyloid positivity.

Several studies support the "Mitochondrial Cascade Theory" which claims that mitochondrial dysfunction is one of the first events that precedes the appearance of β-amyloid and tau aggregates.[8–10] Mitochondria are cellular organelles that control the cell's metabolic energetics, and they have their own genome, the mitochondrial DNA (mtDNA). The regulation of this circular genome is an emerging field, and several reports show it is epigenetically regulated.[11] Our group described the existence of different mtDNA methylation profiles across different stages of the disease by analyzing human postmortem brains with AD-related pathology.[12]

In the present study, we show the presence of a different mtDNA methylation profile in blood samples from patients with MCI who progress to ADD (P) compared to patients with MCI who do not progress to ADD (NP) or Cognitively Normal (CN) subjects, and we show this altered epigenetic pattern even in patients with MCI who were β-amyloid negative before progressing to ADD with β-amyloid positivity. These mitochondrial methylation levels were used to derive principal components, which, together with non-invasive clinical features, were integrated into a Random Forest-based classification model. The resulting multiclass model classifies patients into three groups: CN, MCI who will not progress to ADD, and MCI who will progress to ADD, representing an innovative tool for improving patient recruitment for clinical trials or for the prognosis of ADD progression in Memory Units with a blood test.

RESULTS

Blood samples from 600 patients with MCI and CN were collected at the baseline visit to determine the levels of mtDNA methylation in two loci, the D-loop and the *ND1* (see Experimental model and study participant details section). The patients with MCI were labeled as progressed to ADD (P) or as MCI non-progressed to ADD (NP) based on their diagnosis at the conclusion of their follow-up period (Figure 1 and see *Clinical criteria of progression to dementia* in Experimental model and study participant details section). CN subjects were also included in the study with a minimum of five years of clinical follow-up. The clinical data at the baseline visit, and the cohorts recruited are shown in Tables 1 and S1, respectively. Among the P group (n = 307) there were 62 patients with a negative or missing PET and/or CSF datapoint. However, 30 of these 62 are associated with a positive evaluation at some point during the clinical follow-up (Table S2).

Exploration data analysis of methylation patterns

The initial exploration data analysis of methylation patterns across the D-loop and *ND1* loci showed distinct variations in methylation levels and variability when comparing the CN, NP, and P groups (Table S3). For the D-loop amplicon, notable changes in methylation patterns were observed across the CpG, CHG, and CHH contexts. In the CN group, methylation levels were moderate, with an average of 0.42 ± 0.12 across sites. The coefficient of variation (CV) of 0.29 indicates moderate relative variability in methylation, suggesting some heterogeneity among cytosine sites. The NP group showed a reduction in methylation levels to 0.35 ± 0.17, which is indicative of potential hypomethylation events associated with early stages of cognitive impairment. The increased CV of 0.49 reflects higher variability, suggesting more diverse methylation patterns in the NP group. The P group exhibited an increase in methylation levels to 0.44 ± 0.21, surpassing both NP and CN groups, while the CV remained high at 0.48, indicating continued variability in methylation patterns. That is, the methylation levels in the P stage not only recover from the decline seen in the NP group, but also exceed those in the CN group. Conversely, the ND1 amplicon showed consistently high methylation levels across all groups and contexts, suggesting a potentially different functional role or inherent epigenetic stability, regardless of group.

Boxplots were generated to visualize the distribution of methylation levels across cytosine sites (Figures 2, S2, and S3). Figure 2 reveals distinct differences in the methylation patterns across groups, supporting the findings of the descriptive statistics. Notably, the patients with NP show a lower global mean methylation level compared to the CN and P groups, consistent with the hypomethylation observed also in the descriptive analysis. The distribution of methylation levels in patients with NP also shows a more concentrated interquartile range, indicating reduced variability and a tendency toward lower methylation values across sites.

Apolipoprotein E and mitochondrial DNA methylation levels

To gain a better understanding of the distribution of methylation levels, the relationship between methylation and *APOE* genotypes was investigated. To this end, boxplots were generated to examine the distribution of methylation levels across cytosine sites, stratified by *APOE* genotype and Group (Figures 3, S4, and S5). Genotypes E2.E3, E3.E3, and E3.E4 show relatively symmetric distributions with low variability for the CN group, whose methylation levels are generally centered around the global mean. In contrast, the NP group displays a substantial decrease in methylation levels compared to the global mean, accompanied by wider variability, as evidenced by larger interquartile ranges and whiskers. The P group shows a recovery in methylation levels, with means slightly above those of the CN group and significantly increased variability. This group-dependent pattern is particularly evident in E2.E3, where the NP group displays a sharper fall in methylation levels and a greater increase in methylation levels of the P when compared to E3.E3 and E3.E4, suggesting distinct epigenetic regulatory dynamics for this genotype. The limited number of patients with E2.E2, E2.E4, and E4.E4 genotypes prevents descriptive conclusions from being derived. However, the available data suggest subtle trends, such as a slight overall decrease in methylation levels across groups in E2.E4 and comparable methylation distributions for the NP and P groups in E4.E4.

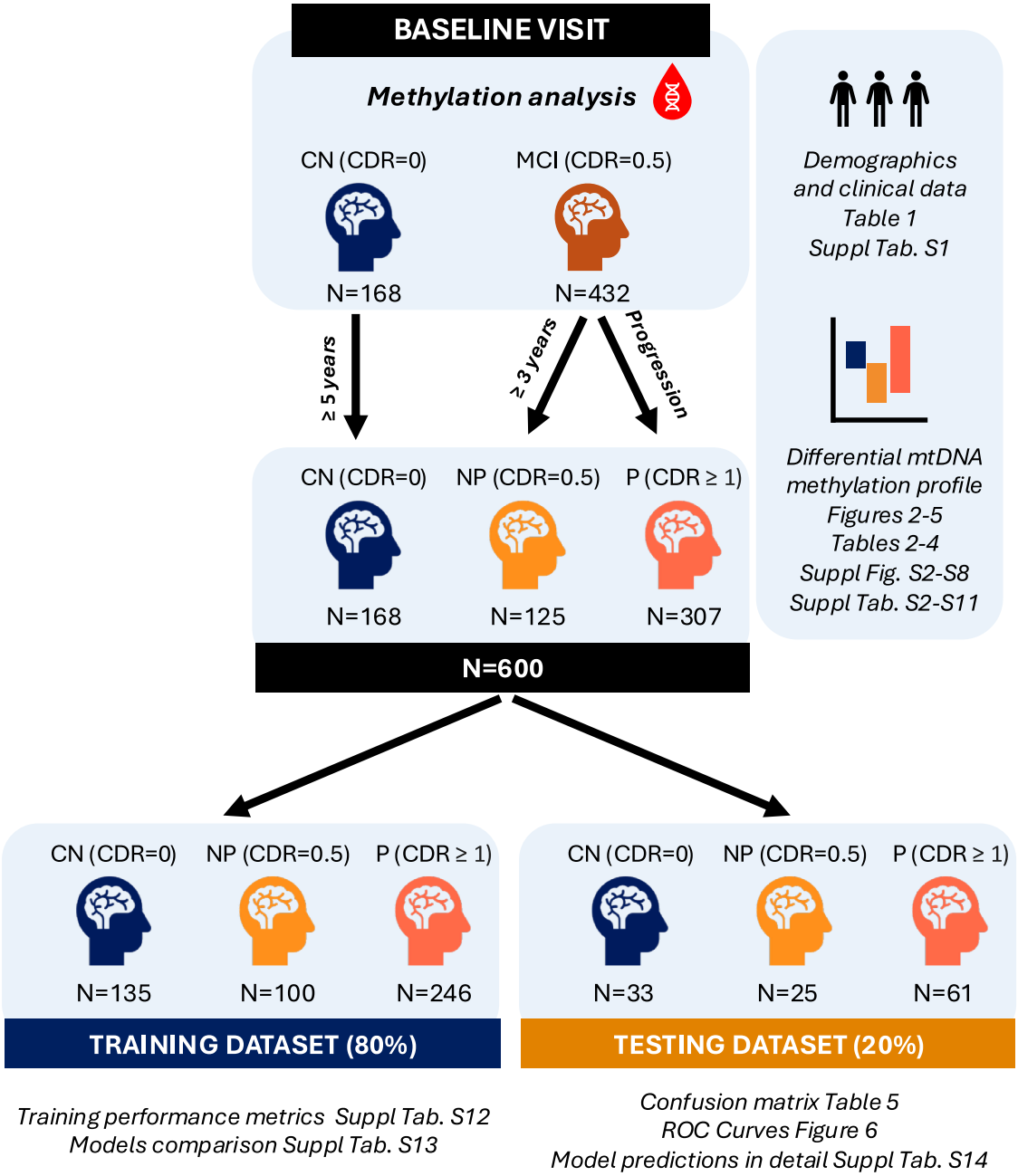


Figure 1. Overview of the clinical study approach and the model development

Mitochondrial DNA methylation levels analysis in negative Aβ-PET Scan subjects

Figures 4A and S6 illustrate the boxplots of the methylation frequencies across cytosine sites in CpG, CHG and CHH contexts of D-loop for the three groups of subjects (CN, NP, and P) with a negative Aβ-PET Scan at the baseline visit. Descriptively, the behavior of these distributions is similar to that described in Figure 2. The CN group has comparatively stable methylation levels across cytosine sites. The NP group exhibits a substantial decrease in methylation levels compared to CN, with its global mean being lower than that of the CN group. The variability in NP is greater than in CN, as evidenced by wider interquartile ranges and longer whiskers, coupled with a tendency to skew to the right (i.e., to the top). In contrast, the P group demonstrates a recovery in methylation levels with respect to the NP group, with its global mean closely aligned to that of the CN group. The P group also exhibits the greatest variability among the three groups, as evidenced by wider interquartile ranges. These findings are also consistent with the trends observed in the *APOE* E4 non-carriers in the APOE-stratified analysis.

Table 1. Clinical data at the baseline visit

Groups	CN	NP[a] *(MCI non progressed)*	P *(MCI progressed to ADD)*
CDR	0	0.5	0.5
Sample size (*N*)	168	125	307
Variables, mean ± SD (% missing)			
Clinical follow up (years)	5–15	3–16	Progression to ADD in 1–14
Education (years)	12.64 ± 2.79 (44%)	10.38 ± 3.88	9.99 ± 4.24 (22%)
Age (years)	66 ± 6	71 ± 7	74 ± 6
MMSE	28.97 ± 1.20	27.18 ± 1.97	25.92 ± 2.78
SOB	0	1.18 ± 0.97	2.04 ± 1.23
Sex, *N* (%)			
Female	89 (53%)	52 (42%)	150 (49%)
Male	79 (47%)	73 (58%)	157 (51%)
E4 carrier, *N* (%)			
Non-carrier	130 (77%)	88 (70%)	131 (43%)
Carrier	38 (23%)	37 (30%)	176 (57%)
PET-CSF, *N* (%)			
Negative PET and CSF	14 (8%)	18 (14%)	4 (1%), 0 out of 4[b]
Positive PET and/or CSF	0 (0%)	31 (25%)	245 (80%)
Negative PET and Missing CSF	64 (38%)	38 (30%)	18 (6%), 2 out of 18[b]
Negative CSF and Missing PET	69 (41%)	19 (15%)	5 (2%), 3 out of 5[b]
Missing PET and CSF	21 (12%)	19 (15%)	35 (11%), 25 out of 35[b]

ADD, Alzheimer's disease dementia; CDR, clinical dementia rating; CN, cognitively normal; CSF, cerebrospinal fluid; MCI, mild cognitive impairment; MMSE, mini-mental state examination; NP, MCI non-progressed; P, MCI progressed; PET, positron emission tomography; SD, standard deviation; SOB, CDR-sum of boxes.

[a]It was mandatory that the NP group showed a clinical follow-up equal or longer than 36 months without showing symptoms of progression.

[b]Number of patients who have shown positivity by Aβ-PET Scan or CSF (βA42, p-Tau181, and/or t-Tau) at a later clinical visit.

Figures 4B and S7 illustrate the boxplots of the methylation frequencies across cytosine sites in CpG, CHG, and CHH contexts of D-loop for the three groups of subjects (CN, NP, and P) who were both negative Aβ-PET Scan and negative CSF (Aβ42, p-Tau181, and t-Tau) at the baseline visit. Compared to the negative Aβ-PET Scan described in Figure 4A, these distributions show a more pronounced pattern. In the CN group, the frequency methylation levels remain stable across cytosine sites, consistent with previous observations. In contrast, the NP group shows a substantial global reduction in methylation levels, with a markedly lower mean compared to CN. The interquartile range is considerably larger than in CN, and a distinct asymmetry is observed in each boxplot. The median is closer to the first quartile, suggesting a right skew. The lower whisker is relatively short, whereas the upper whisker extends well beyond the maximum values observed in CN. In group P, in comparison to both CN and NP groups, the methylation frequency levels display a notable increase, with a global mean that exceeds that of CN. The interquartile range is also wider than in CN, although it is not as extreme as in NP. The distribution is inversely asymmetric, with the median near the third quartile, and a left skew. Furthermore, the lower whisker reaches the minimum observed in CN, while the upper whisker remains short, highlighting a restricted upper tail. Nevertheless, it is important to note that the P group consists of only four patients. This pattern is consistently observed across the three methylation contexts.

Interestingly, methylation frequencies across cytosine sites within the three different contexts of the D-loop were analyzed for the three groups of subjects: (1) subjects who had a negative Aβ-PET Scan at the baseline visit and remained negative throughout their clinical follow-up, and (2) patients with a negative Aβ-PET Scan at their baseline visit who developed a positive scan during a follow-up assessment (Figures 5 and S8). More specifically, negativity was checked by PET and/or CSF, as defined by a binary classification (positive or negative) that combines the results of four assessments: Aβ-PET Scan, Aβ42 in CSF, p-Tau181, and t-Tau in CSF, each of which is also binary (positive or negative). Thus, an individual was classified as PET-CSF positive if at least one of the four assays demonstrated positivity at any time point during the clinical follow-up. In contrast, an individual was classified as negative if all four tests were consistently negative during the clinical follow-up.

Figure 5B shows a profile comparable to the one shown in Figure 4A. The NP group shows decreased methylation levels in comparison to the CN group. However, it has left-skewed interquartile ranges, characterized by a long whisker extending to the right, but its median values nearly align with those of the P group. The P group exhibits greater variability and marginally elevated mean methylation levels relative to NP. The CN group demonstrates substantially reduced variability and a higher global mean relative to the other two groups. Table S2 reports the number of subjects with a negative or no Aβ-PET Scan at

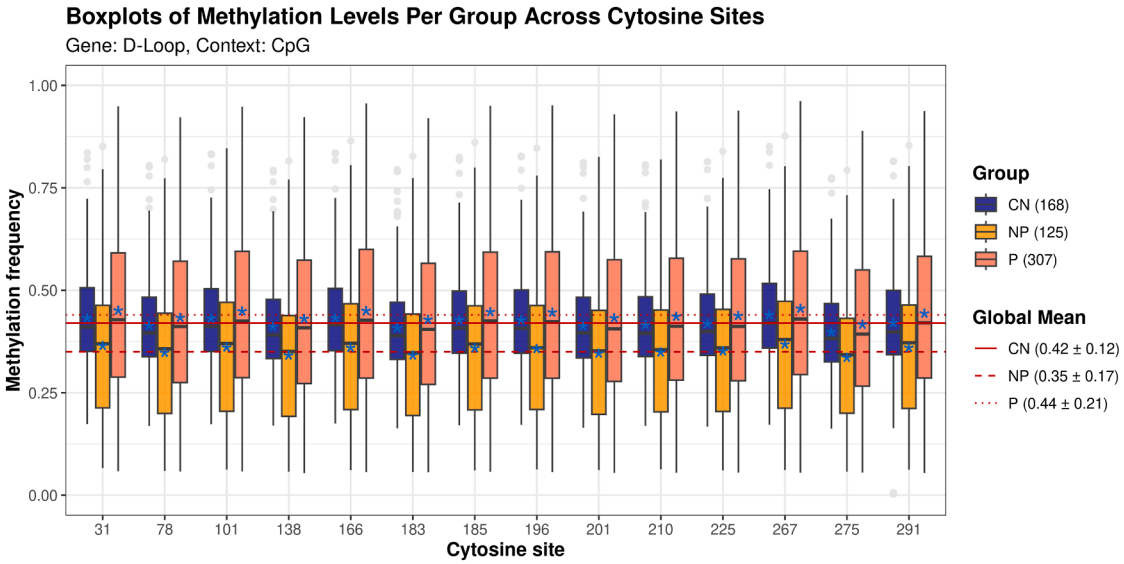


Figure 2. Methylation levels per group across cytosine sites within the CpG context in amplicon D-loop

The *x*-axis represents base positions relative to the start of amplicon D-loop, with each position corresponding to a cytosine site analyzed within the CpG context. The *y* axis indicates methylation frequency (%). Boxplots are color-coded to differentiate between the three groups: blue for cognitively normal (CN), light orange for MCI non-progressed to ADD (NP), and dark orange for MCI progressed to ADD (P) patients. The box spans from Q1 to Q3 (the interquartile range, IQR), with a line indicating the median. The whiskers extend to the most extreme data points within 1.5 times the IQR from the quartiles. Any values beyond that range are shown as individual outliers. The solid red line (CN) dashed red line (NP), and dotted red line (P), indicate the group-specific global means. Blue stars indicate the mean methylation level for each group. On the right side, the labels display the group-specific global mean ± standard deviation and the number of patients in each group. CHG and CHH contexts are shown in Figure S2. See Figure S3 for CpG, CHG and CHH contexts in amplicon ND1.

the baseline visit, and a positive Aβ-PET Scan and or positive CSF during the follow-up period.

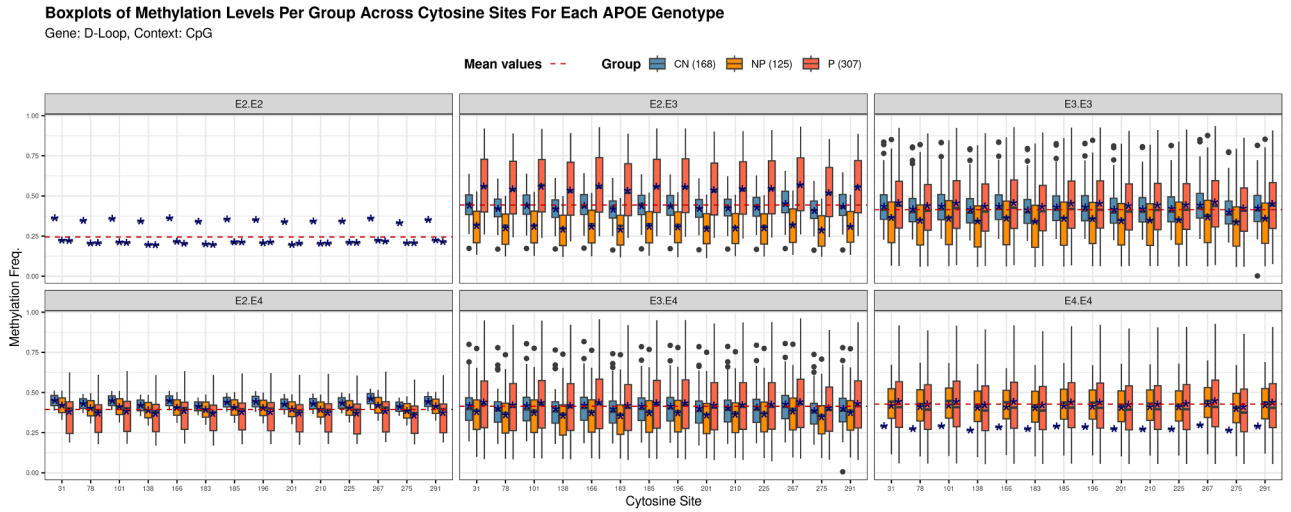
Differential methylation analysis

Differential methylation analysis (DMA) was conducted for each methylation context (CpG, CHG, and CHH) across D-loop and ND1 amplicons, comparing the three groups: CN, NP, and P (Tables 2, 3, 4, S4, and S5) in blood samples from the baseline visit. The analysis revealed differentially methylated sites (DMS) in D-loop CpGs for two of the three comparisons (CN vs. NP, and NP vs. P) using the Benjamini-Yekutieli (BY) adjusted *p*-value threshold of <0.05 (Tables 2 and 3). For instance, in the CpG context, for the CN vs. NP comparison, all cytosine loci were detected as DMS, with log fold changes (logFC) ranging from 0.325 to 0.377, consistently indicating hypomethylation in the patients with NP with respect to CN subjects. Interestingly, in the comparison between patients with NP and P, all cytosine loci were identified as DMS, with logFC values ranging from −0.477 to −0.450, consistently highlighting hypomethylation in the NP with respect to the patients with P. In contrast, no significant DMS were observed for the CN vs. P comparison, as all BY were equal to 1, suggesting no evidence of significant methylation differences between these groups (Table 4). The results for CHG and CHH contexts are provided in Table S4.

Notably, DMA for the ND1 amplicon revealed limited significant differences across all comparisons and methylation contexts, suggesting that overall methylation patterns in this region remain stable across groups (Table S5). In the CpG context, although no significant DMS were detected with a threshold of BY < 0.05, some cytosine sites showed moderate evidence of differential methylation in specific comparisons when alternative criteria [e.g., Benjamini-Hochberg (BH) < 0.1] were applied. For instance, in the CN vs. NP comparison, sites 119, 123, and 179 displayed logFC ranging from −0.316 to −0.270 (Table S5). Overall, our findings demonstrate the existence of a differential mitochondrial epigenetic pattern among blood samples from patients with MCI who progress to patients with ADD and MCI who remain as MCI.

We performed a DMA in the subset of patients with MCI who showed a negative Aβ-PET Scan (Tables S6, S7, S8, S9, S10, and S11). In the CpG context (Table S6), significant DMS were identified when comparing CN to NP, and NP to P using the BY < 0.05. For the CN vs. NP comparison, all cytosines were identified as significant DMS, with logFC ranging from 0.707 to 0.761, consistently indicating NP hypomethylation relative to CN. Similarly, for the NP vs. P comparison, all cytosine sites were identified as DMS, with logFC values ranging from −0.715 to −0.692, consistently showing hypomethylation in the NP group with respect to the P group. In contrast, no significant DMS were observed for the CN vs. P comparison, as all BY adjusted *p*-values were equal to 1. Similar behavior of significant differences was observed in the CHG and CHH contexts (Tables S7 and S8).

The consistency of significant DMS across several *p*-value correction approaches was assessed to further evaluate the robustness of the findings. Both BH and BY adjustments identified the same significant DMS for the CN vs. NP, and NP vs. P comparisons, highlighting the robustness of the results despite



Number of Subjects

Group	E2.E2	E2.E3	E3.E3	E2.E4	E3.E4	E4.E4	Total
CN	1	23	106	2	35	1	168
NP	2	9	77	2	31	4	125
P	1	15	115	5	131	40	307
Total	4	47	298	9	197	45	600

Figure 3. Methylation levels per group across cytosine sites within the CpG context in amplicon D-loop, stratified by *APOE* genotype

The panels represent the six *APOE* genotypes: E2.E2, E2.E3, E3.E3 (top row), and E2.E4, E3.E4, E4.E4 (bottom row). For each cytosine site, boxplots show methylation levels for CN (blue), NP (light orange), and P (dark orange) groups. The box spans from Q1 to Q3 (IQR), with a line indicating the median. The whiskers extend to the most extreme data points within 1.5 times the IQR from the quartiles. Any values beyond that range are shown as individual outliers. Blue stars indicate the mean methylation level for each group, while the red dashed line represents the global mean methylation frequency within each *APOE* genotype. The table summarizes the sample sizes for each *APOE* genotype and group. CHG and CHH contexts are shown in Figures S4 and S5, respectively.

the application of less rigorous (BH) or more conservative (Bonferroni) criteria. Bonferroni notably supported the significance of these DMS, particularly in the NP vs. P comparison, confirming the consistency of the observed differences.

DMA for the ND1 amplicon showed no significant DMS across the CpG context in any comparison with a BY < 0.05 (Table S9). This outcome was also observed in the CHG and CHH contexts (Tables S10 and S11), which further supports the absence of significant methylation differences in this region across the three groups.

To refine the analysis, subjects who entered the study with a negative Aβ-PET Scan were stratified according to their subsequent PET-CSF status (remained negative or converted to positive). DMA was performed separately for each PET-CSF subgroup across the D-loop and ND1 amplicons, analyzing the three methylation contexts and the three group comparisons (CN vs. NP, CN vs. P, and NP vs. P). Subjects in the persistently negative PET-CSF showed clear and different patterns of differential methylation. Significant DMS were identified for both the CN vs. NP, and NP vs. P comparisons, with all cytosine sites passing the BY < 0.05 threshold (Table S6). For the CN vs. NP comparison, all cytosines were significant, with logFC values ranging from 0.804 to 0.847, indicating consistent hypomethylation in the NP group relative to the CN group. Similarly, for the NP vs. P comparison, all cytosines were also significant, with logFC values ranging from −0.979 to −0.957, demonstrating hypomethylation in the NP group relative to the P group. No significant DMS were detected for the CN vs. P comparison using BY < 0.05. Conversely, for subjects classified as PET-CSF positive after a clinical follow-up, no significant DMS were identified across any of the three comparisons among groups, using the BY < 0.05 threshold (Table S6). Relaxing the cutoff to BY < 0.1 or using alternative adjustment methods, such as BH, revealed several loci with moderate evidence of differential methylation in the CN vs. NP comparison, characterized by logFC values indicative of hypomethylation in the NP group with respect to the CN group. However, these findings should be interpreted cautiously due to the more permissive statistical criteria. Results for other methylation contexts (CHG and CHH) displayed similar trends and are provided in Tables S7 and S8.

Unlike the findings for the D-loop amplicon, the ND1 amplicon had no significant DMS in all group comparisons, regardless of PET-CSF status (positive or negative) and across all methylation contexts (Tables S9, S10, and S11). Although some loci showed modest evidence of differential methylation when less severe thresholds were applied (e.g., BY < 0.1 or BH < 0.05), the results were inconsistent and not robust.

Multiclass classification model

After confirming the existence of differential mitochondrial epigenetic patterns among blood samples from patients with P and NP, we analyzed the pre-processing data. After creating dummy variables, no zero and near zero variance variables were detected. Data splitting resulted in two distinct subsets:

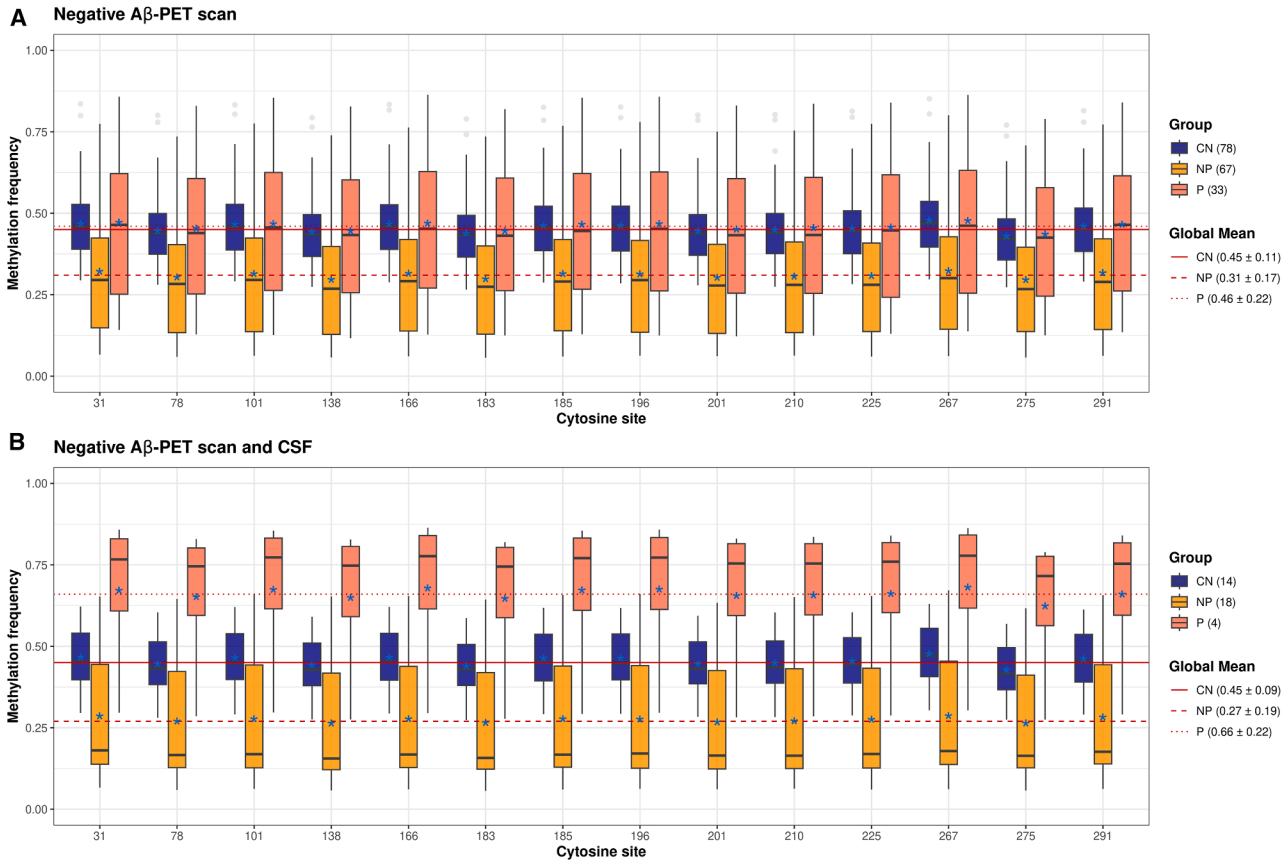


Figure 4. Methylation levels per Group across cytosine sites within the CpG context in amplicon D-loop at the baseline visit

The x axis represents base positions relative to the start of the D-loop amplicon, with each position corresponding to a specific cytosine site within the CpG context. The y axis indicates methylation frequency (%). Boxplots are color-coded to differentiate between the three groups: blue for cognitively normal (CN), light orange for MCI non-progressed to ADD (NP), and dark orange for MCI progressed to ADD (P) patients. The box spans from Q1 to Q3 (IQR), with a line indicating the median. The whiskers extend to the most extreme data points within 1.5 times the IQR from the quartiles. Any values beyond that range are shown as individual outliers. The solid red line (CN) dashed red line (NP), and dotted red line (P), indicate the group-specific global means. Blue stars indicate the mean methylation level for each group. On the right side, the labels display the group-specific global mean ± standard deviation and the number of patients in each group.

(A) Methylation frequency across cytosine sites within the CpG context in D-loop amplicon is shown in subjects who exhibited a negative Aβ-PET Scan at the baseline visit. CHG and CHH contexts are shown in Figure S6. See also Tables S6, S7, S8, S9, S10, and S11.

(B) Methylation frequency across cytosine sites within the CpG context in D-loop amplicon is shown in subjects who exhibited a negative Aβ-PET Scan and a CSF at the baseline visit. CHG and CHH contexts are shown in Figure S7. See also Tables S6, S7, S8, S9, S10, and S11.

the training dataset, consisting of 481 subjects, and the independent test dataset, consisting of 119 subjects. The independent test dataset was completely hidden during model development and was only used to assess the final model's general performance. The feature selection process, aimed at identifying the optimal number of features, was conducted on the training dataset using a tri-repeated 10-fold cross-validation. Throughout this process, numerous models were evaluated. Notably, the models demonstrating the best performance often incorporated dimensionality reduction techniques, specifically combining genetic algorithms (GA) with principal component (PC) analysis to enhance effectiveness.

The training results indicated that the model, which incorporated the first five principal components (PC1-PC5) along with the variables Age, *APOE* (represented using the one-hot encoding), and SOB, achieved an average accuracy of 84.4% across all folds and repetitions during the CV, with a corresponding Kappa statistic of 73.9%. That is, the final model was:

$$Group \sim PC1 + PC2 + PC3 + PC4 + PC5 + Age + E2.E3 + E3.E3 + E3.E4 + E4.E4 + SOB \quad \text{(Model 1)}$$

where E2.E3, E3.E3, E3.E4, E4.E4 are the dummy features derived from the expansion of the *APOE* variable through the one-hot encoding.

This average accuracy and the kappa statistic metrics demonstrate the model's robust predictive ability and substantial agreement beyond chance. The mean sensitivity, averaged across all groups, was 79.3%, and the mean specificity was 90.6%. Table S12 provides a summary of the descriptive statistics for various performance metrics of the Random Forest (RF)

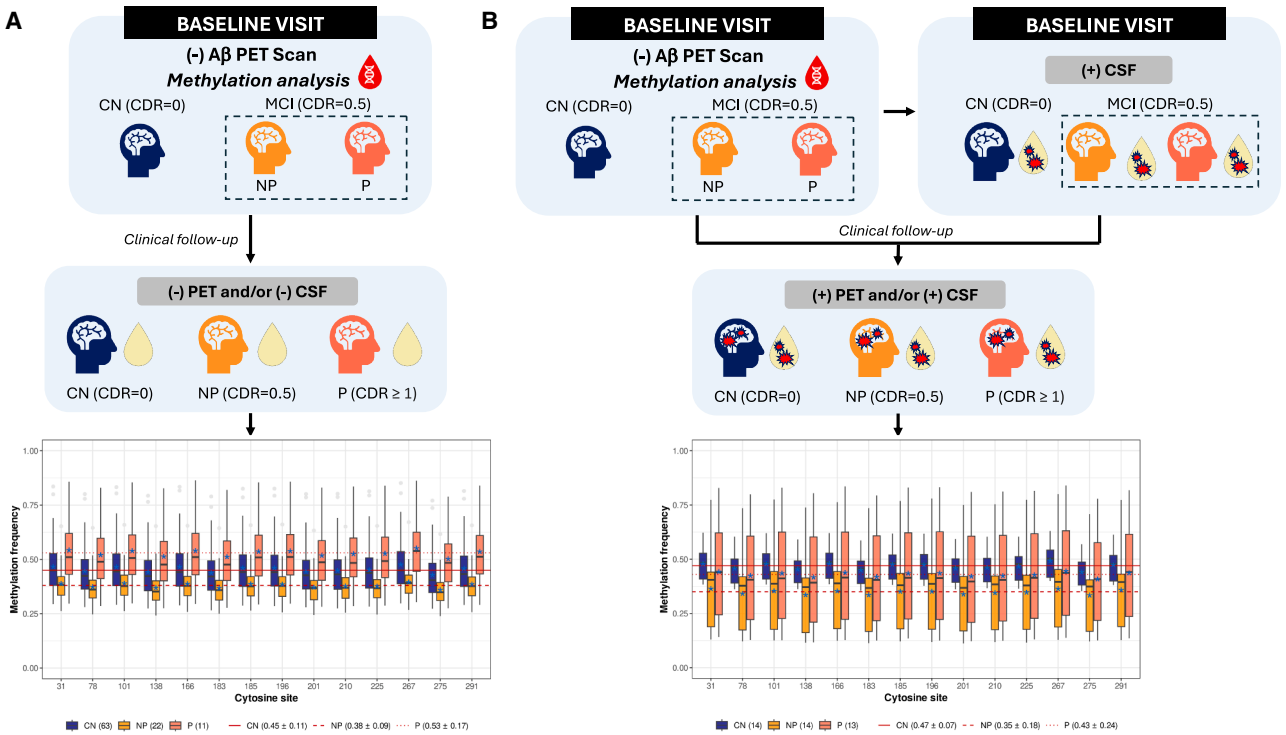


Figure 5. Methylation levels per Group across cytosine sites within the CpG context in amplicon D-loop, stratified by PET-CSF classification

(A) Boxplots of methylation levels on subjects classified as negative PET-CSF. The box spans from Q1 to Q3 (IQR), with a line indicating the median. The whiskers extend to the most extreme data points within 1.5 times the IQR from the quartiles. Any values beyond that range are shown as individual outliers. All groups were negative by Aβ PET Scan at the baseline visit when the methylation analysis in the subjects' blood samples was performed. All the subjects were negative by Aβ PET Scan and/or CSF (Aβ42, p-Tau181, and t-Tau) during the clinical follow-up. See also Tables S6, S7, S8, S9, S10, and S11.

(B) Boxplots of methylation levels in subjects classified as positive PET-CSF. All groups were negative by Aβ PET Scan at the baseline visit when the methylation analysis on the subjects' blood samples was performed. During the clinical follow-up, all the subjects presented PET-CSF positivity. Subjects were designated as positive PET-CSF if at least one of the four assessments (Aβ-PET Scan, Aβ42 in CSF, p-Tau181 and t-Tau in CSF) revealed positivity at any point during the clinical follow-up. Some of these patients were already CSF-positive at the baseline visit. The use of color-coding in boxplots enables the differentiation of the three groups: blue for cognitive normal (CN), light orange for MCI non-progressed to ADD (NP), and dark orange for MCI progressed to ADD (P) patients. The global mean is represented by a solid red line for CN, a dashed red line for NP, and a dotted red line for P. Blue stars indicate the mean methylation level for each group. The legend accompanying the boxplots provides the group-specific global mean ± standard deviation and the number of patients in each group. See also Tables S6, S7, S8, S9, S10, and S11. CHG and CHH contexts are shown in Figure S8.

model training process. The confusion matrix generated by the classification that the model performed on the test dataset is shown in Table 5, and the associated performance metrics are summarized in Table S13 (column Model 1). The model yielded a test accuracy of 83.2% (95% CI: 75.2%–89.4%) and a Kappa statistic of 71.1%, demonstrating strong generalizability beyond the training data. In classifying patients with MCI that progressed to ADD, the model achieved a sensitivity of 95.1%, specificity of 70.7%, PPV of 77.3%, and NPV of 93.2%. With an F1-score of 85.3%, the model exhibits a solid balance between accurately identifying positive cases and its overall prediction accuracy. Similarly, the metrics for CN subjects and patients with NP were also calculated to ensure a comprehensive understanding of the model's capabilities across different categories. Table S14 details predictions from the testing data.

Post hoc analysis using Youden's index identified optimal probability thresholds for group P, group NP, and group CN of 56.3% (95% CI: 41.0%–64.9%), 23.7% (95% CI: 10.5%–43.7%), and 1.1% (95% CI: 0.9%–4.8%), respectively. These probability thresholds derive from the test dataset and maximize the balance between sensitivity and specificity for each class.

The model demonstrates a strong balance between training and test metrics, suggesting that it neither overfits nor underfits the data. The training accuracy of 84.4% and test accuracy of 83.2% are closely aligned, indicating that the model generalizes well to unseen data. Similarly, the Kappa statistic is consistent across datasets (73.9% in training and 71.1% in testing). The sensitivity of 95.1% and specificity of 70.7% for group P, along with the AUC of 90.3%, further support the model's ability to capture meaningful patterns while maintaining generalizability. These findings highlight the model's robustness, suggesting that its complexity is appropriate for the data and does not compromise generalizability.

Figure 6 displays ROC curves and associated sensitivity and specificity analyses, with the optimal probability cutoff points determined using the Youden's Index. The top left plot shows the ROC curves for P vs. Other, NP vs. Other, and CN vs. Other

Table 2. Differential methylation analysis results for CpG sites within the D-loop amplicon in the comparison CN vs. NP

Amplicon	Context	Site	logFC	AveMeth	T	P.Value	BH	BY	Bonferroni	B
D-loop	CpG	31	0.3559903	−3.229E-17	3.0586647	0.0022303	0.0025336	0.0082382	0.0312241	−1.6674557
D-loop	CpG	78	0.362679	1.335E-16	3.1161333	0.0018386	0.0025336	0.0082382	0.0257405	−1.4983251
D-loop	CpG	101	0.3712463	−4.119E-17	3.1897442	0.0014292	0.0025336	0.0082382	0.0200095	−1.2770888
D-loop	CpG	138	0.3720153	8.025E-17	3.1963509	0.001397	0.0025336	0.0082382	0.0195574	−1.2569797
D-loop	CpG	166	0.3765606	−1.448E-17	3.235404	0.0012194	0.0025336	0.0082382	0.0170721	−1.1372632
D-loop	CpG	183	0.3627133	2.034E-17	3.1164282	0.0018368	0.0025336	0.0082382	0.0257147	−1.497449
D-loop	CpG	185	0.3611516	6.8E-17	3.1030102	0.001922	0.0025336	0.0082382	0.0269083	−1.537224
D-loop	CpG	196	0.3671031	−2.596E-17	3.1541456	0.0016154	0.0025336	0.0082382	0.0226155	−1.3847245
D-loop	CpG	201	0.3612038	−2.171E-17	3.1034585	0.0019191	0.0025336	0.0082382	0.0268676	−1.5358979
D-loop	CpG	210	0.3575354	−7.637E-17	3.07194	0.0021336	0.0025336	0.0082382	0.0298699	−1.6286659
D-loop	CpG	225	0.3623463	1.449E-16	3.1132752	0.0018565	0.0025336	0.0082382	0.0259908	−1.5068109
D-loop	CpG	267	0.3732578	3.033E-17	3.2070267	0.0013462	0.0025336	0.0082382	0.0188468	−1.2243975
D-loop	CpG	275	0.3541216	4.976E-17	3.0426084	0.0023527	0.0025336	0.0082382	0.0329372	−1.7141469
D-loop	CpG	291	0.3251866	−1.385E-16	2.7939998	0.0052179	0.0052179	0.0169663	0.0730507	−2.4057322

This table presents the results of the differential methylation analysis performed using the limma package in R. Each row corresponds to a specific cytosine site within the analyzed amplicon. The columns report the amplicon name (Amplicon), the cytosine methylation context (Context), and the genomic position of the site (Site). Differential methylation results are provided as the log2 fold-change (logFC) in methylation between conditions CN and NP, the average methylation level across all samples (AveMeth), and the moderated t-statistic from the limma analysis (t). The table also includes the raw *p*-value (P.Value) and multiple testing corrections: Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) adjustments for false discovery rate (FDR) control, as well as the Bonferroni correction for family-wise error rate (FWER) control. The last column (B) represents the log-odds of differential methylation as estimated by limma's empirical Bayes method. See also Tables S4 and S5.

comparative scenarios, with AUC values of 90.3% (95% CI: 84.7%–95.9%), 84.9% (95% CI: 77.3%–92.6%), and 99.8% (95% CI: 99.4%–100.0%), respectively. The remaining plots detail the sensitivity and specificity for each comparison, with optimal probability cutoff points indicated on each sensitivity curve by a dotted gray line. The sensitivity and specificity values shown in these plots (along with their 95% confidence intervals) correspond to the optimal cutoff points determined by Youden's index.

Model comparison

A comparative analysis of 14 models, each representing different combinations of predictors derived from the selected base

Table 3. Differential methylation analysis results for CpG sites within the D-loop amplicon in the comparison NP vs. P

Amplicon	Context	Site	logFC	AveMeth	T	P.Value	BH	BY	Bonferroni	B
D-loop	CpG	31	−0.4571306	−3.229E-17	−4.3726068	1.243E-05	1.45E-05	4.714E-05	0.000174	2.8437781
D-loop	CpG	78	−0.4683209	1.335E-16	−4.4796459	7.576E-06	1.227E-05	3.989E-05	0.0001061	3.3072051
D-loop	CpG	101	−0.4668511	−4.119E-17	−4.465587	8.09E-06	1.227E-05	3.989E-05	0.0001133	3.2456973
D-loop	CpG	138	−0.4773634	8.025E-17	−4.5661402	5.039E-06	1.227E-05	3.989E-05	7.055E-05	3.6898702
D-loop	CpG	166	−0.471961	−1.448E-17	−4.5144646	6.435E-06	1.227E-05	3.989E-05	9.008E-05	3.4603686
D-loop	CpG	183	−0.4743395	2.034E-17	−4.5372163	5.78E-06	1.227E-05	3.989E-05	8.092E-05	3.5610915
D-loop	CpG	185	−0.4654893	6.8E-17	−4.4525606	8.595E-06	1.227E-05	3.989E-05	0.0001203	3.1888799
D-loop	CpG	196	−0.469119	−2.596E-17	−4.4872802	7.31E-06	1.227E-05	3.989E-05	0.0001023	3.3406859
D-loop	CpG	201	−0.4682133	−2.171E-17	−4.4786166	7.613E-06	1.227E-05	3.989E-05	0.0001066	3.3026954
D-loop	CpG	210	−0.4684606	−7.637E-17	−4.4809822	7.529E-06	1.227E-05	3.989E-05	0.0001054	3.3130613
D-loop	CpG	225	−0.4650566	1.449E-16	−4.4484216	8.762E-06	1.227E-05	3.989E-05	0.0001227	3.1708612
D-loop	CpG	267	−0.4552313	3.033E-17	−4.3544399	1.35E-05	1.454E-05	4.727E-05	0.000189	2.7662367
D-loop	CpG	275	−0.4609427	4.976E-17	−4.4090711	1.051E-05	1.338E-05	4.35E-05	0.0001472	3.0003921
D-loop	CpG	291	−0.4497109	−1.385E-16	−4.3016351	1.715E-05	1.715E-05	5.575E-05	0.0002401	2.5426844

This table presents the results of the differential methylation analysis performed using the limma package in R. Each row corresponds to a specific cytosine site within the analyzed amplicon. The columns report the amplicon name (Amplicon), the cytosine methylation context (Context), and the genomic position of the site (Site). Differential methylation results are provided as the log2 fold-change (logFC) in methylation between conditions NP and P, the average methylation level across all samples (AveMeth), and the moderated t-statistic from the limma analysis (t). The table also includes the raw *p*-value (P.Value) and multiple testing corrections: Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) adjustments for false discovery rate (FDR) control, as well as the Bonferroni correction for family-wise error rate (FWER) control. The last column (B) represents the log-odds of differential methylation as estimated by limma's empirical Bayes method. See also Tables S4 and S5.

Table 4. Differential methylation analysis results for CpG sites within the D-loop Amplicon in the comparison CN vs. P

Amplicon	Context	Site	logFC	AveMeth	t	P.Value	BH	BY	Bonferroni	B
D-loop	CpG	31	−0.1011402	−3.229E-17	−1.0695938	0.2848331	0.3371363	1	1	−4.6684553
D-loop	CpG	78	−0.1056419	1.335E-16	−1.1172008	0.2639406	0.3371363	1	1	−4.6409741
D-loop	CpG	101	−0.0956048	−4.119E-17	−1.0110542	0.3120198	0.3371363	1	1	−4.700607
D-loop	CpG	138	−0.1053481	8.025E-17	−1.1140932	0.2652712	0.3371363	1	1	−4.6428045
D-loop	CpG	166	−0.0954004	−1.448E-17	−1.008893	0.3130552	0.3371363	1	1	−4.7017594
D-loop	CpG	183	−0.1116263	2.034E-17	−1.1804871	0.2378401	0.3371363	1	1	−4.6025896
D-loop	CpG	185	−0.1043377	6.8E-17	−1.103408	0.2698818	0.3371363	1	1	−4.6490592
D-loop	CpG	196	−0.1020159	−2.596E-17	−1.0788541	0.2806839	0.3371363	1	1	−4.6632035
D-loop	CpG	201	−0.1070095	−2.171E-17	−1.1316635	0.2578084	0.3371363	1	1	−4.6323886
D-loop	CpG	210	−0.1109252	−7.637E-17	−1.1730729	0.2408	0.3371363	1	1	−4.6071958
D-loop	CpG	225	−0.1027103	1.449E-16	−1.0861972	0.2774231	0.3371363	1	1	−4.6590068
D-loop	CpG	267	−0.0819735	3.033E-17	−0.8668988	0.3860223	0.3860223	1	1	−4.7720683
D-loop	CpG	275	−0.1068211	4.976E-17	−1.1296711	0.2586473	0.3371363	1	1	−4.6335779
D-loop	CpG	291	−0.1245243	−1.385E-16	−1.3168879	0.1879123	0.3371363	1	1	−4.5126696

This table presents the results of the differential methylation analysis performed using the limma package in R. Each row corresponds to a specific cytosine site within the analyzed amplicon. The columns report the amplicon name (Amplicon), the cytosine methylation context (Context), and the genomic position of the site (Site). Differential methylation results are provided as the log2 fold-change in methylation between conditions CN and NP (logFC), the average methylation level across all samples (AveMeth), and the moderated t-statistic from the limma analysis (t). The table also includes the raw *p*-value (P.Value) and multiple testing corrections: Benjamini-Hochberg (BH) and Benjamini-Yekutieli (BY) adjustments for false discovery rate (FDR) control, as well as the Bonferroni correction for family-wise error rate (FWER) control. The last column (B) represents the log-odds of differential methylation as estimated by limma's empirical Bayes method. See also Tables S4 and S5.

model (1), revealed notable differences in performance metrics across training and test datasets. Table S13 provides a comprehensive summary of all metrics across the models and facilitates a direct comparison of their performance.

The second-best model,

$$Group \sim PC1 + PC2 + PC3 + PC4 + PC5 + Age + SOB \quad \text{(Model 2)}$$

achieved a test accuracy of 82.4% (95% CI: 74.3%–88.7%) and a Kappa statistic of 70.8%, closely mirroring the performance of the base model (1). However, it exhibited slightly lower sensitivity for group P (86.9%) compared to the base model (95.1%), suggesting a marginal loss in identifying positive cases for this group. This reduction in sensitivity was accompanied by a higher specificity (77.6%) relative to the base model (70.7%), reflecting a trade-off between false positive and false negative rates. The AUC values for group P (89.4%) and group NP (84.3%) further confirm the robustness of this model, although they remain slightly lower than those of the base model (1). These results suggest that while the second-best model provides strong overall performance, the inclusion of *APOE* genotypes in the base model contributes to improved sensitivity for group P.

In contrast, models without principal components, for example,

$$Group \sim Age + SOB \quad \text{(Model 8)}$$

$$Group \sim Age + E2.E3 + E3.E3 + E3.E4 + E4.E4 \quad \text{(Model 12)}$$

consistently demonstrated signs of underfitting, as evidenced by their uniformly lower metrics across training and test datasets. For instance, model (8) achieved a training accuracy of 76.7% and a test accuracy of 73.9% (95% CI: 65.1%–81.6%). For Group NP, the sensitivity remained unchanged at 32.0%, while the specificity dropped to 85.1% compared to 96.8% in the base model (1). This reduction in specificity for the less prevalent group, coupled with a lower Kappa statistic of 57.2%, indicates that the exclusion of PCs compromises the model's ability to effectively generalize across all groups. Similarly, model (12) exhibited the poorest overall performance, with test accuracy dropping to 58.8% (95% CI: 49.4%–67.8%) and a Kappa statistic of 30.0%. For Group P, sensitivity decreased to 72.1%, compared to 95.1% in the base model (1), and specificity dropped further to 58.6%, compared to 70.7%. The AUC for Group P also declined to 72.1% (95% CI: 62.9%–81.3%), reflecting reduced discriminatory power. For Group NP, sensitivity dropped to 0%, indicating a complete inability to correctly classify true positive cases, while specificity was perfect at 100.0%, with no false positives. The AUC for NP was 54.4% (95% CI: 42.0%–66.8%), demonstrating the model's inability to adequately differentiate this group. Furthermore, for the CN group sensitivity was 78.8%, while specificity fell to 70.9%, with an AUC of 62.5% (95% CI: 52.0%–73.0%). These results highlight that models excluding PCs fail to achieve robust sensitivity and specificity, leading to significant underfitting and poor discrimination, particularly for less prevalent groups such as the NP group.

Overall, the comparative analysis confirms that neither the base model (1) nor the second-best model (2) shows substantial signs of overfitting, as their training and test metrics remain closely aligned. In contrast, models excluding PCs consistently underfit the data, failing to capture key variability between groups and subtle patterns within the data that are necessary for accurate group discrimination.

Table 5. Confusion matrix of the model

	Reference		
Prediction	CN	NP	P
CN	33	0	0
NP	0	8	3
P	0	17	58

See also Table S14, which shows the model predictions in the test dataset.
CN, cognitively normal; NP, MCI non-progressed; P, MCI progressed.

DISCUSSION

Cerebral mitochondrial dysfunction has been implicated in the pathophysiology of AD and this has been demonstrated in an *in vivo* study.[13] Moreover, mitochondrial impairment has also been detected at systemic level.[14,15] In fact, a study revealed high similarity between the transcriptome of blood and brain from patients with AD.[16] Interest in mtDNA methylation has developed over the past decade, not only in the brain[17] but also in blood from patients with AD.

The D-loop region was found to be hypomethylated in peripheral blood DNA and in plasma cell-free DNA from patients with late-onset AD relative to CN subjects,[18,19] and hypermethylation was described in a small study of patients with MCI ($n = 14$) relative to CN using a methylation sensitive-high resolution melting analysis.[20] These studies selected well-defined clinical phenotypes, differentiating among patients with MCI and AD, while our study is focused on 432 patients with MCI (CDR = 0.5) designated as P or NP after a clinical follow-up period. Then, based on the loci we previously identified as differentially methylated in human postmortem brains with AD-related pathology,[12] we studied mtDNA from blood samples at a baseline visit from patients with CN subjects and MCI.

Our results revealed that although the *ND1* locus has a higher percentage of methylation than the D-loop, only several cytosines differ significantly between the groups. By contrast, we showed the D-loop locus has similar methylation profiles between CN subjects and the P group, while the NP group shows hypomethylated levels. The D loop region contains the Conserved Sequence Blocks I-III (CSB) that form part of the Origin of replication of the H strand (O_H). This region, which resides on the L strand of mtDNA, contains the site of RNA to DNA transition, and generates a 7S RNA that acts as an mtDNA replication primer. The CSB II region was found to be enriched in methylcytosines in endothelial cells and was accompanied by increased 7S DNA,[21] speculating a relationship between DNA methylation and 7S abundance, which means that hypomethylation in the CSB II region should boost mtDNA replication. Our D loop amplicon is adjacent to CSB I-II of the L strand. It has been suggested that the methylation of the L strand, which displays a complementary sequence to the RNA primer that leads to the RNA to DNA transition, may play a role in the generation or stabilization of the 7S DNA fragments, and could regulate the replication and the transcription of mtDNA.[22] Therefore, we propose that hypomethylation in the D-loop region in the NP group may be a compensatory response that preserves mitochondrial function, avoiding the progression to ADD. A recent systematic review showed that increased mtDNA copy number avoids mitochondrial dysfunction.[23] Additional studies are required to unravel the molecular mechanisms involved in D-loop methylation. Interestingly, a recent study shows the presence of a differential mitochondrial transcriptomic profile in blood samples from P and NP patients.[24]

Sex differences in both the incidence of AD and in DNA methylation patterns, including those measured in blood samples, have been well documented.[25–27] Although our study did not report sex-stratified analyses, additional analyses performed for internal validation did not indicate a significant contribution of Sex to the classification performance, nor did Sex demonstrate associations with site-specific methylation patterns in our dataset. Nevertheless, it is important to acknowledge that Sex remains a relevant factor in the context of epigenetic studies of AD using blood-based biomarkers. Future studies with larger or independent cohorts may help to clarify whether Sex-specific methylation signatures could further enhance the prognostic potential of methylation-based classifiers in this context.

The impact of *APOE* genotype on mtDNA methylation levels was also studied due to the link of APOE with AD pathogenesis.[28] *APOE* is the major genetic risk factor for AD, and there are three alleles: E2 (lowers AD risk), E3 (neutral AD risk), and E4 (increases AD risk). The protein is a cholesterol carrier in the brain but also affects systemic metabolism. Interestingly, reduced cytochrome oxidase activity was found in platelets of *APOE* E4 carrier patients with AD.[29] The mechanism through which *APOE* affects AD risk is unknown. A new theory that links mitochondrial dysfunction with AD pathogenesis, the ''MAM Hypothesis,'' argues that AD is a lipid disorder.[30] It is based on the characterization of lipid rafts in the endoplasmic reticulum (ER) known as mitochondria-associated ER membranes (MAM) that communicate physically and biochemically with mitochondria. Interestingly, an *in vitro* study showed that the ApoE4 protein increases MAM activity.[31] Our study shows the hypomethylation of the D-loop region in the NP group compared to the P group, with this profile being more pronounced in *APOE* E4 non-carriers. Furthermore, the presence of an E2 allele increases this difference, whereas the presence of an E4 allele decreases intergroup differences. Further experiments should be performed to determine whether MAM activity alters mtDNA methylation.

We analyzed D-loop methylation profiles in blood samples from β-amyloid negative individuals. This analysis also showed NP group hypomethylation. It is important to note we defined ADD progression based on clinical criteria (shift of CDR, SOB or FAQ score), with CDR = 0.5 representing the sole inclusion criterion and without considering β-amyloid or tau biomarker status. Therefore, these observations can be questioned by arguing that P group individuals with a negative Aβ-PET Scan progressed to a non-AD dementia.[3] Accordingly, we performed the same analysis while stratifying the three groups (CN, NP, and P) to consider subjects who were biomarker positive by PET or CSF after a clinical follow-up and subjects who remained negative throughout. The analysis again showed hypomethylation in members of the NP group whose AD biomarkers remained negative. This pattern also applied to individuals whose AD biomarkers transitioned to positive during the follow-up period. These results need to be validated in a larger cohort and are

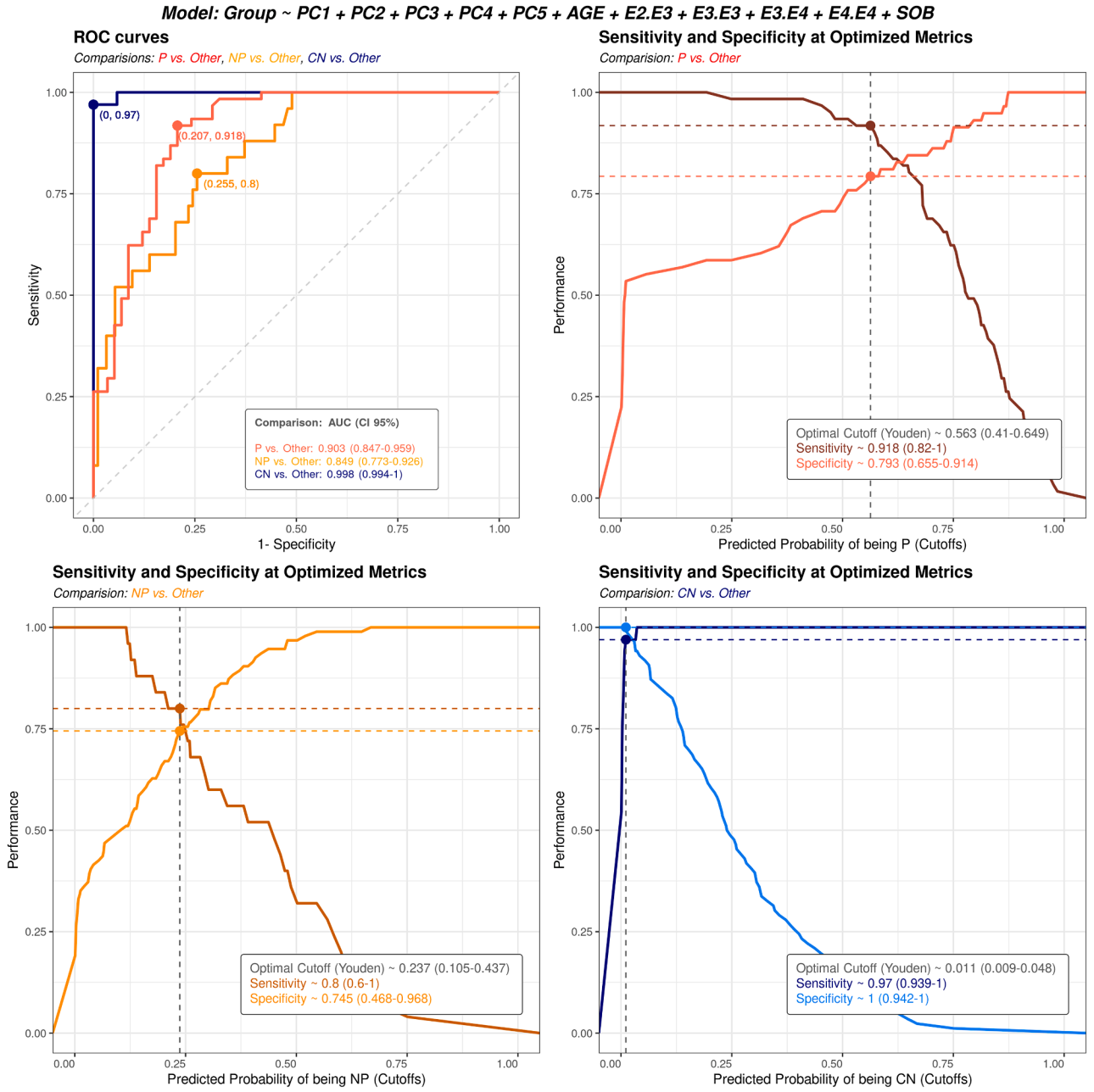


Figure 6. Receiver operating characteristic (ROC) curves and sensitivity-specificity analyses for the multiclass classification model

The top left panel shows One-vs-Other ROC curves for P vs. Other (dark orange), NP vs. Other (light orange), and CN vs. Other (blue), with corresponding AUC values (see also Table S13). The remaining panels display sensitivity and specificity across probability thresholds, with optimal cutoff points determined using Youden's Index, indicated by dotted gray lines.

potentially important as they suggest changes in a non-β-amyloid biomarker may precede changes in β-amyloid biomarkers.

Following the differential methylation analysis, a multiclass classification model was developed to distinguish between CN, NP and P. The best-performing model integrated the methylcytosine data from both the ND1 and D-loop amplicons, as well as the age, SOB, and *APOE* genotype information. This model demonstrated robust generalization capacity and achieved an accuracy of 84.4% (kappa = 73.9%) in the training and 83.2% (kappa = 71.1%) in the independent test datasets. The minimal decrease in accuracy and kappa between training and testing indicates the model effectively captured relevant patterns without overfitting. Typically, the discrepancy between training and test performance would be significantly greater when overfitting occurs, while underfitting would lead to a general decrease in accuracy and kappa values. Therefore, the consistency observed

here shows the selected feature set is reliable and able to differentiate among the three groups.

However, the classification of the NP group is particularly challenging due to the definitions of the groups NP and P. Although both groups initially meet criteria for the MCI syndrome, our DMA demonstrates differences in their methylation patterns. Nevertheless, the underlying methylation profiles may overlap, resulting in less evident separation. This overlap, combined with the need to differentiate NP from P individuals, may contribute to the observed reduction in sensitivity (32%). An important factor to consider is the dynamic nature of disease progression. At the time of model development, patient classification into groups CN, NP, or P was determined by expert clinical assessment. These classifications were used as labels for training and evaluating the model. However, while the data used both molecular and clinical remained fixed, patient follow-up after developing the model revealed that some patients with MCI initially classified as NP were later reclassified as P. This suggests that some of the observed misclassifications may not be true errors but rather a consequence of disease progression over time. In other words, the model may have captured an intrinsic biological distinction that was not yet reflected in the initial clinical labels. If these updated classifications were incorporated into model training and evaluation, its predictions might align more closely with redefined clinical labels.

Our results highlight the relevance of using probabilistic outputs rather than discrete classifications in clinical decision-making. Our data additionally demonstrates the potential value of complementing clinical assessments with patient stratification.

MCI patient heterogeneity can potentially confound the outcomes of clinical trials that focus on that group.[32] Over 65% of current AD/MCI clinical trials do not target β-amyloid or tau, but rather neurotransmitter receptors, inflammation, metabolism, epigenetic regulators, neurogenesis, and synaptic plasticity.[33] Moreover, we propose the use of this prognostic classifier model as a triaging tool that can integrate into clinical practice. Patients with MCI without β-amyloid pathology, but who are predicted to decline by our classifier model, would particularly warrant clinical follow-up.

In conclusion, this study establishes mtDNA methylcytosine analysis as a blood-based biomarker that predicts cognitive decline versus stability in persons meeting MCI syndrome criteria. Its predictive capacity appears to apply to MCI individuals with positive or negative Aβ or Aβ/tau biomarkers.

Limitations of the study

Despite promising findings, some limitations of our study should be addressed or clarified.

Current diagnostic approaches for AD rely on identifying the presence of neuropathologic changes associated with the appearance of amyloid plaques or AD tauopathy in the brain, recently through blood tests.[3] In the present study, 10.42% of P patients had no record of positivity for Aβ and/or tau at the baseline visit and at follow-up (see Table 1). In most cases, there is missing data on the presence of these proteins, and therefore, their positivity cannot be confirmed. However, while the detection of Aβ and tau proteinopathy serves as a core biomarker of the disease, the present study was addressed to avoid their significant limitation, lying in the lack of direct correlation between Aβ and/or tau positivity with cognitive decline. For this reason, this study presents a disruptive solution, independent from Aβ, that addresses the disease from a holistic perspective using artificial intelligence, including the AD biology (mitochondrial dysfunction), but also the symptomatology (neuropsychological assessment) and the major risk factors (age and genetic risk). All the MCI cases included followed the MCI criteria defined by Albert et al.[34]

The time factor represents a significant limitation of the study's approach. Our research diverges from most AD studies by not comparing well-defined phenotypes, i.e., CN vs. MCI vs. ADD. Instead, the present findings are derived from a complex study design in which MCI patients were analyzed at their baseline visit but then were labeled according to their subsequent progression or not to ADD. To classify patients as stable MCI (i.e., NP), a minimum clinical follow-up of 36 months was required. Blood samples from these patients are complex to obtain due to their limited availability and the restricted access to the biorepositories for non-amyloid/tau research approaches. On the other hand, according to the continuum of AD, there is no clearly defined standard time frame for progression from MCI to ADD. This poses a key limitation for evaluating the model's performance, particularly regarding the specificity of the test. Some patients labeled as NP based on the available follow-up period may progress to ADD in the future. In such cases, even if the model correctly predicts that these patients will progress (i.e., classifies them as P), they would still be considered false positives due to the current follow-up status. In other words, the model may appear to make incorrect predictions not because it is inaccurate, but because the follow-up period is not long enough to confirm the true progression status of some patients.

The use of machine learning results in a new set of capabilities in healthcare. However, the use of predictive models based on machine learning presents an important limitation: the communication of the uncertainty or confidence level associated with their predictions in clinical practice.[35,36] In real-world clinical practice, there are probability scores associated with the prediction's ambiguity (probabilities extremely low or extremely high indicate high confidence in the prediction, while probabilities near 0.5 indicate a lack of confidence[36]). Ideally, our predictive model should be able to effectively indicate "Indeterminate or unconclusive result" or "I am highly certain about this prediction." Although this is a technical limitation of the present ADD prognosis model, our team will make efforts to determine a cut-off to ensure high confidence in its predictions, to abstain from providing erroneous prognosis, and to communicate the indeterminate results in the clinical practice.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Marta Barrachina (mbarrachina@admit-therapeutics.com).

Materials availability

This study did not generate new materials.

Data and code availability

The data supporting the findings of this study were obtained from

(1) the Australian Imaging Biomarkers and Lifestyle Study of Aging (AIBL), which is available from the AIBL database (https://aibl.csiro.au/) upon registration and compliance with the data use agreement.
(2) the Knight Alzheimer Disease Research Center—Washington University School of Medicine, which is available from the Knight ADRC (https://knightadrc.wustl.edu/) upon request and compliance with the data use agreement.
(3) the Fundación CITA-Alzhéimer Fundazioa, provided by the Basque Biobank/Biodonostia Node (CITA), which is available at www.biobancovasco.org, upon request and compliance with the data use agreement.

This article does not report the original code, as it is part of a proprietary algorithm currently protected under patent application number PCT/EP2023/056250.

Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all the participants who took part in this study and the clinicians who referred participants. We thank the samples from the AIBL study (aibl.org.au), that is a collaboration between CSIRO, Edith Cowan University (ECU), The Florey Institute of Neuroscience and Mental Health (FINMH), National Aging Research Institute (NARI) and Austin Health. It also involves support from CogState Ltd., Hollywood Private Hospital, and Sir Charles Gairdner Hospital. The AIBL study has received partial financial support from the Alzheimer's Association (US), the Alzheimer's Drug Discovery Foundation, an Anonymous Foundation, the Science and Industry Endowment Fund, the Dementia Collaborative Research Centres, the Victorian Government's Operational Infrastructure Support program, the Australian Alzheimer's Research Foundation (now Alzheimer's Research Australia), the National Health and Medical Research Council (NHMRC), and the Yulgilbar Foundation. Numerous commercial interactions have supported AIBL data collection and analyses. In-kind support has also been provided by Sir Charles Gairdner Hospital, Cogstate Ltd, Hollywood Private Hospital, The University of Melbourne, and St Vincent's Hospital. Finally, we thank all AIBL investigators, not listed as authors, who contributed to the design and implementation of the resource but did not actively contribute data to, or participate in the development, analysis, interpretation or writing of this current study. A complete listing of AIBL investigators can be found at https://aibl.org.au/about/our-researchers/. We also thank the Washington University, the ADRC, and Program Project grants (P50AG05681, P01AG03991, and P01AG026276) that made support the Knight ADRC cohorts. R.H.S. is supported by P30AG072093. This project received funding from European Union's Horizon 2020 research and innovation program under grant agreement No. 960327; Ayuda Torres Quevedo 2018 (PTQ2018-009873), Ayuda Neotec 2019 (EXP-00123430/SNEO-20191199), and funding provided for by the Alzheimer's Drug Discovery Foundation Diagnostics Accelerator. The Diagnostics Accelerator, created in July 2018, is a partnership of funders with funding commitments totaling nearly $100 million over 3 years from partners including ADDF Co-Founder Leonard A. Lauder, Bill Gates, Jeff Bezos, MacKenzie Scott, the Dolby family, the Charles and Helen Schwab Foundation, The Association for Frontotemporal Degeneration, among others, to develop novel biomarkers for the early detection of Alzheimer's disease and related dementias. This research initiative is dedicated to accelerating the development of affordable and accessible biomarkers to diagnose Alzheimer's disease and related dementias and advance the clinical development of more targeted treatments. Through translational research awards and access to consulting support from industry experts, this program will challenge, assist, and fund the research community in both academia and industry to develop novel peripheral and digital biomarkers. To learn more about the initiative, visit the website at www.alzdiscovery.org/accelerator.

AUTHOR CONTRIBUTIONS

Based on CrediT Taxonomy, J.G.-B.: investigation, methodology, project administration, resources, and review and editing. J.L.M.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, writing original draft, and review and editing. M.Bl.: conceptualization, data curation, investigation, methodology, project administration, supervision, validation, and review and editing. P.M.: data curation, formal analysis, investigation, and review. B.F.: investigation and review. C.T.: data curation, project administration, supervision, visualization, and review and editing. N.R.: resources and review. I.R.: resources and review. J.C.: investigation, resources, and review. C.F.: resources and review and editing. S.M.L.: resources and review and editing. A.T.-M.: resources and review. R.S.-V.: resources and review. J.B.: resources and review. J.F.: resources and review. A.L.: resources and review and editing. C.M.: validation, supervision, and review. R.H.S.: supervision and review and editing. R.R.-R.: investigation, methodology, project administration, resources, and review and editing. M. Ba.: conceptualization, funding acquisition, methodology, project administration, supervision, validation, writing original draft, and review and editing.

DECLARATION OF INTERESTS

J.G.-B., R.R.-R., and M.Ba. are co-founders of ADmit Therapeutics and hold shares of the company. J.L.M., M.Bl., P.M., B.F., C.T., and M.Ba. are employees of ADmit Therapeutics. C.M. and R.H.S. are members of the ADmit's Scientific Advisory Board. The data presented form part of a patent, with the application number PCT/EP2023/056250. A.L. has served as a consultant or on advisory boards for Almirall, Fujirebio-Europe, Roche, Biogen, Grifols, Novartis, Eisai, Lilly, and Nutricia, outside the submitted work.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Inclusion and exclusion criteria for participants
  - Clinical criteria for dementia progression
- METHOD DETAILS
  - DNA extraction
  - Bisulfite treatment
  - Primer design
  - PCR for D-loop and *ND1*
  - Amplicon library preparation
  - High-throughput sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Methylation data processing and analysis
  - Bioinformatics workflow for extracting methylation data
  - Exploratory data analysis of methylation patterns
  - Differential methylation analysis
  - Development of the classification model
  - Exploratory data analysis
  - Data preprocessing
  - Model training
  - Model performance evaluation

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.113418.

Received: April 9, 2025
Revised: April 16, 2025
Accepted: August 18, 2025
Published: August 21, 2025

REFERENCES

1. Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Kaye, J., Montine, T.J., et al. (2011). 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. Alzheimer's Dement. *7*, 280–292. https://doi.org/10.1016/j.jalz.2011.03.003.
2. Braak, H., and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. *82*, 239–259. https://doi.org/10.1007/BF00308809.
3. Jack, C.R., Andrews, J.S., Beach, T.G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., et al. (2024). Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimer's Dement. *20*, 5143–5169. https://doi.org/10.1002/alz.13859.
4. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., et al. (2010). Alzheimer's disease neuroimaging initiative (ADNI). Neurology *74*, 201–209. https://doi.org/10.1212/WNL.0b013e3181cb3e25.
5. Hanson, M., Liu, Y., and Mattke, S. (2025). Geographic access to amyloid PET scans in the US: Implications for clinical trials and real-world treatment. J. Prev. Alzheimers Dis. *12*, 154.
6. Meyer, M.R., Kirmess, K.M., Eastwood, S., Wente-Roth, T.L., Irvin, F., Holubasch, M.S., Venkatesh, V., Fogelman, I., Monane, M., Hanna, L., et al. (2024). Clinical validation of the PrecivityAD2 blood test: a mass spectrometry-based test with algorithm combining %p-tau217 and Aβ42/40 ratio to identify presence of brain amyloid. Alzheimer's Dement. *20*, 3179–3192. https://doi.org/10.1002/alz.13764.
7. Ashton, N.J., Brum, W.S., Di Molfetta, G., Benedet, A.L., Arslan, B., Jonaitis, E., Langhough, R.E., Cody, K., Wilson, R., Carlsson, C.M., et al. (2024). Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for alzheimer disease pathology. JAMA Neurol. *81*, 255–263. https://doi.org/10.1001/jamaneurol.2023.5319.
8. Swerdlow, R.H., and Khan, S.M. (2004). A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. Med. Hypotheses *63*, 8–20. https://doi.org/10.1016/j.mehy.2003.12.045.
9. Swerdlow, R.H., Burns, J.M., and Khan, S.M. (2014). The Alzheimer's disease mitochondrial cascade hypothesis: progress and perspectives. Biochim. Biophys. Acta *1842*, 1219–1231. https://doi.org/10.1016/j.bbadis.2013.09.010.
10. Swerdlow, R.H. (2023). The Alzheimer's disease mitochondrial cascade hypothesis: a current overview. J. Alzheimers Dis. *92*, 751–768. https://doi.org/10.3233/JAD-221286.
11. Donato, L., Mordà, D., Scimone, C., Alibrandi, S., D'Angelo, R., and Sidoti, A. (2024). From powerhouse to regulator: The role of mitoepigenetics in mitochondrion-related cellular functions and human diseases. Free Radic. Biol. Med. *218*, 105–119. https://doi.org/10.1016/j.freeradbiomed.2024.03.025.
12. Blanch, M., Mosquera, J.L., Ansoleaga, B., Ferrer, I., and Barrachina, M. (2016). Altered mitochondrial DNA methylation pattern in Alzheimer disease-related pathology and in Parkinson disease. Am. J. Pathol. *186*, 385–397. https://doi.org/10.1016/j.ajpath.2015.10.004.
13. Venkataraman, A.V., Mansur, A., Rizzo, G., Bishop, C., Lewis, Y., Kocagoncu, E., Lingford-Hughes, A., Huiban, M., Passchier, J., Rowe, J.B., et al. (2022). Widespread cell stress and mitochondrial dysfunction occur in patients with early Alzheimer's disease. Sci. Transl. Med. *14*, eabk1051. https://doi.org/10.1126/scitranslmed.abk1051.
14. Delbarba, A., Abate, G., Prandelli, C., Marziano, M., Buizza, L., Arce Varas, N., Novelli, A., Cuetos, F., Martinez, C., Lanni, C., et al. (2016). Mitochondrial alterations in peripheral mononuclear blood cells from Alzheimer's disease and mild cognitive impairment patients. Oxid. Med. Cell. Longev. *2016*, 5923938. https://doi.org/10.1155/2016/5923938.
15. Lunnon, K., Keohane, A., Pidsley, R., Newhouse, S., Riddoch-Contreras, J., Thubron, E.B., Devall, M., Soininen, H., Kłoszewska, I., Mecocci, P., et al. (2017). Mitochondrial genes are altered in blood early in Alzheimer's disease. Neurobiol. Aging *53*, 36–47. https://doi.org/10.1016/j.neurobiolaging.2016.12.029.
16. Iturria-Medina, Y., Khan, A.F., Adewale, Q., and Shirazi, A.H.; Alzheimer's Disease Neuroimaging Initiative (2020). Blood and brain gene expression trajectories mirror neuropathology and clinical deterioration in neurodegeneration. Brain *143*, 661–673. https://doi.org/10.1093/brain/awz400.
17. Devall, M., Mill, J., and Lunnon, K. (2014). The mitochondrial epigenome: a role in Alzheimer's disease? Epigenomics *6*, 665–675. https://doi.org/10.2217/epi.14.50.
18. Stoccoro, A., Siciliano, G., Migliore, L., and Coppedè, F. (2017). Decreased methylation of the mitochondrial D-loop region in late-onset Alzheimer's disease. J. Alzheimers Dis. *59*, 559–564. https://doi.org/10.3233/JAD-170139.
19. Ding, B., Zhang, X., Wan, Z., Tian, F., Ling, J., Tan, J., and Peng, X. (2023). Characterization of mitochondrial DNA methylation of Alzheimer's disease in plasma cell-free DNA. Diagnostics *13*, 2351. https://doi.org/10.3390/diagnostics13142351.
20. Stoccoro, A., Baldacci, F., Ceravolo, R., Giampietri, L., Tognoni, G., Siciliano, G., Migliore, L., and Coppedè, F. (2022). Increase in mitochondrial D-loop region methylation levels in mild cognitive impairment individuals. Int. J. Mol. Sci. *23*, 5393. https://doi.org/10.3390/ijms23105393.
21. Bianchessi, V., Vinci, M.C., Nigro, P., Rizzi, V., Farina, F., Capogrossi, M.C., Pompilio, G., Gualdi, V., and Lauri, A. (2016). Methylation profiling by bisulfite sequencing analysis of the mtDNA non-coding region in replicative and senescent endothelial cells. Mitochondrion *27*, 40–47. https://doi.org/10.1016/j.mito.2016.02.004.
22. Bellizzi, D., D'Aquila, P., Scafone, T., Giordano, M., Riso, V., Riccio, A., and Passarino, G. (2013). The control region of mitochondrial DNA shows an unusual CpG and non-CpG methylation pattern. DNA Res. *20*, 537–547. https://doi.org/10.1093/dnares/dst029.
23. Harerimana, N.V., Paliwali, D., Romero-Molina, C., Bennett, D.A., Pa, J., Goate, A., Swerdlow, R.H., and Andrews, S.J. (2023). The role of mitochondrial genome abundance in Alzheimer's disease. Alzheimer's Dement. *19*, 2069–2083. https://doi.org/10.1002/alz.12812.
24. Huang, Y.-L., Tsai, T.-H., Shen, Z.-Q., Chan, Y.-H., Tu, C.-W., Tung, C.-Y., Wang, P.-N., and Tsai, T.-F. (2025). Transcriptomic predictors of rapid progression from mild cognitive impairment to Alzheimer's disease. Alzheimers Res. Ther. *17*, 3. https://doi.org/10.1186/s13195-024-01651-0.
25. Silva, C., Zhang, W., Young, J.I., Gomez, L., Schmidt, M.A., Varma, A., Chen, X.S., Martin, E.R., and Wang, L. (2022). Distinct sex-specific DNA methylation differences in Alzheimer's disease. Alzheimers Res. Ther. *14*, 1–21. https://doi.org/10.1186/s13195-022-01070-z.
26. Zhang, L., Young, J.I., Gomez, L., Silva, T.C., Schmidt, M.A., Cai, J., Chen, X., Martin, E.R., and Wang, L. (2021). Sex-specific DNA methylation differences in Alzheimer's disease pathology. Acta Neuropathol. Commun. *9*, 77–79. https://doi.org/10.1186/s40478-021-01177-8.
27. Govender, P., Ghai, M., and Okpeku, M. (2022). Sex-specific DNA methylation: impact on human health and development. Mol. Genet. Genomics. *297*, 1451–1466. https://doi.org/10.1007/s00438-022-01935-w.
28. Mahley, R.W., and Huang, Y. (2012). Apolipoprotein E sets the stage: response to injury triggers neuropathology, including Alzheimer's disease. Neuron *76*, 871–885. https://doi.org/10.1016/j.neuron.2012.11.020.Apolipoprotein.
29. Wilkins, H.M., Wang, X., Menta, B.W., Koppel, S.J., Bothwell, R., Becker, A.M., Anderson, H., Schwartz, E., Pei, D., Yellapu, N.K., et al. (2021). Bioenergetic and inflammatory systemic phenotypes in Alzheimer's disease APOE ε4-carriers. Aging Cell *20*, e13356. https://doi.org/10.1111/acel.13356.

30. Area-Gomez, E., and Schon, E.A. (2024). Towards a unitary hypothesis of Alzheimer's disease pathogenesis. J. Alzheimers Dis. *98*, 1243–1275. https://doi.org/10.3233/JAD-231318.

31. Tambini, M.D., Pera, M., Kanter, E., Yang, H., Guardia-Laguarta, C., Holtzman, D., Sulzer, D., Area-Gomez, E., and Schon, E.A. (2016). ApoE4 upregulates the activity of mitochondria-associated ER membranes. EMBO Rep. *17*, 27–36. https://doi.org/10.15252/embr.201540614.

32. Mielke, M.M., Anderson, M., Ashford, J.W., Jeromin, A., Lin, P.-J., Rosen, A., Tyrone, J., Vandevrede, L., Willis, D.R., Hansson, O., et al. (2024). Recommendations for clinical implementation of blood-based biomarkers for Alzheimer's disease. Alzheimer's Dement. *20*, 8216–8224. https://doi.org/10.1002/alz.14184.

33. Cummings, J., Zhou, Y., Lee, G., Zhong, K., Fonseca, J., and Cheng, F. (2024). Alzheimer's disease drug development pipeline:2024. Alzheimer's Dement. *10*, e12465. https://doi.org/10.1002/trc2.12465.

34. Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement. *7*, 270–279. https://doi.org/10.1016/j.jalz.2011.03.008.

35. Chua, M., Kim, D., Choi, J., Lee, N.G., Deshpande, V., Schwab, J., Lev, M. H., Gonzalez, R.G., Gee, M.S., and Do, S. (2023). Tackling prediction uncertainty in machine learning for healthcare. Nat. Biomed. Eng. *7*, 711–718. https://doi.org/10.1038/s41551-022-00988-x.

36. Kompa, B., Snoek, J., and Beam, A.L. (2021). Second opinion needed: communicating uncertainty in medical machine learning. npj Digit. Med. *4*, 4. https://doi.org/10.1038/s41746-020-00367-3.

37. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

38. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

39. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.

40. R Core Team (2024). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). https://www.R-project.org/.

41. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686. https://doi.org/10.21105/joss.01686.

42. Albing, C., Vossen, J., and Newham, C. (2007). In Bash Cookbook: Solutions and Examples for bash Users, M. Loukides, ed. (O'Reilly Media, Inc).

43. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

44. Fowler, C., Rainey-Smith, S.R., Bird, S., Bomke, J., Bourgeat, P., Brown, B.M., Burnham, S.C., Bush, A.I., Chadunow, C., Collins, S., et al. (2021). Fifteen years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer's disease. J. Alzheimers Dis. Rep. *5*, 443–468. https://doi.org/10.3233/ADR-210005.

45. Alcolea, D., Clarimón, J., Carmona-Iragui, M., Illán-Gala, I., Morenas-Rodríguez, E., Barroeta, I., Ribosa-Nogué, R., Sala, I., Sánchez-Saudinós, M.B., Videla, L., et al. (2019). The Sant Pau Initiative on Neurodegeneration (SPIN) cohort: a data set for biomarker discovery and validation in neurodegenerative disorders. Alzheimer's Dement. *5*, 597–609. https://doi.org/10.1016/j.trci.2019.09.005.

46. Morris, J.C. (1993). The clinical dementia rating (CDR): current version and scoring rules. Neurology *43*, 2412–2414. https://doi.org/10.1212/wnl.43.11.2412-a.

47. McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement. *7*, 263–269. https://doi.org/10.1016/j.jalz.2011.03.005.

48. Lansdall, C.J., McDougall, F., Butler, L.M., Delmar, P., Pross, N., Qin, S., McLeod, L., Zhou, X., Kerchner, G.A., and Doody, R.S. (2023). Establishing clinically meaningful change on outcome assessments frequently used in trials of mild cognitive impairment due to Alzheimer's disease. J. Prev. Alzheimers Dis. *10*, 9–18. https://doi.org/10.14283/jpad.2022.102.

49. Pfeffer, R.I., Kurosaki, T.T., Harrah, C.H., Chance, J.M., and Filos, S. (1982). Measurement of functional activities in older adults in the community. J. Gerontol. *37*, 323–329. https://doi.org/10.1093/geronj/37.3.323.

50. Benjamini, Y., and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. J. Educ. Behav. Stat. *25*, 60–83. https://doi.org/10.3102/10769986025001060.

51. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Statist. *29*, 1165–1188. https://doi.org/10.1214/aos/1013699998.

52. Dunn, O.J. (1961). Multiple comparisons among means. J. Am. Stat. Assoc. *56*, 52–64. https://doi.org/10.1080/01621459.1961.10482090.

53. Youden, W.J. (1950). Index for rating diagnostic tests. Cancer *3*, 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Whole blood samples from cognitively normal (CN) subjects and patients with mild cognitive impairment (MCI)	Prospective recruitment for the present study. Hospital Universitari de Bellvitge, Hospital Clínic de Barcelona, and Hospital General de L'Hospitalet.	MAP-AD Study
Whole blood aliquots from CN subjects	Fundación CITA-Alzhéimer Fundazioa, Basque Biobank/Biodonostia Node	http://www.biobancovasco.org
Total DNA from MCI patients	Sant Pau Initiative on Neugeneration (SPIN) cohort—Hospital de la Santa Creu i Sant Pau (Alcolea et al., 2019)	https://santpaumemoryunit.com/our-research/spin-cohort/
Total DNA from CN and MCI subjects	Australian Imaging, Biomarker and Lifestyle Study biobank	http://aibl.org.au
Total DNA from MCI subjects	Knight Alzheimer Disease Research Center—Washington University School of Medicine biorepository	https://knightadrc.wustl.edu/
Total DNA from MCI patients	Hospital Clínic de Barcelona biorepository	https://www.clinicbarcelona.org/idibaps/areas-y-programas/neurociencias-clinicas-y-experimentales/enfermedad-de-alzheimer-y-otros-trastornos-cognitivos
Critical commercial assays		
EZ DNA Methylation™ Kit	Zymo Research	Cat#D5001
IDT® for Illumina® DNA/RNA UD Indexes Set A, B, C and D	Illumina	Cat#20091654; Cat#20091656; Cat#20091658; Cat#20091660
Illumina® PhiX Control v3 Library	Illumina	Cat#FC-110-3001
MiSeq® Reagent Kit v3 600 cycles	Illumina	Cat#MS-102-3003
Oligonucleotides		
D-loop-Forw: 5'TCGTCGGCAGCGTCAGATGTGTATA AGAGACAGYAYTTGGGGGTA GYTAAAGTGAAYTG3'	This paper	N/A
D-loop-Rev: 5'GTCTCGTGGGCTCGGAGATGTGTATAA GAGACAGTCCTACAARCATTAATTAATTAACACAC 3'	This paper	N/A
ND1-Forw: 5'TCGTCGGCAGCGTCAGATGTGTATAAG AGACAGATAAAAYT TAAAAYTTTAYAGTYAGAG 3'	This paper	N/A
ND1-Rev: 5'GTCTCGTGGGCTCGGAGATGTGTATAA GAGACAGTTRARTTTRATRCTCACCCTRATCA 3'	This paper	N/A
Software and algorithms		
FastQC (v0.11.9)	Andrews, 2010[37]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Trimmomatic (v0.39)	Bolger et al., 2014[38]	http://www.usadellab.org/cms/?page=trimmomatic
Bismark	Krueger and Andrews, 2011[39]	https://www.bioinformatics.babraham.ac.uk/projects/bismark/
R stats package (v4.0.0)	R Core Team, 2024[40]	https://www.r-project.org/
Tidyverse (v2.0.0) collection packages	Wickham et al. 2019[41]	https://www.tidyverse.org/
Bash scripting	Albing et al. 2007[42]	N/A
Bioconductor package (v3.60.2)	Ritchie et al., 2015[43]	https://www.bioconductor.org/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Training and testing datasets include data from 600 subjects aged from 55 to 85, recruited from seven international cohorts. Sixty four out of 600 individuals were collected from the Knight Alzheimer Disease Research Center—Washington University School of Medicine (Knight, United States), 248 from the Australian Imaging, Biomarker and Lifestyle (AIBL, Australia; see aibl.org.au for further details) cohort,[44] 125 from the Hospital Universitari de Bellvitge (HUB, Spain), 79 from the Hospital Clínic de Barcelona (HCB, Spain), 75 from the Fundación CITA-Alzhéimer Fundazioa, provided by the Basque Biobank/Biodonostia Node (CITA, Spain; see www.biobancovasco.org for further details), 5 from the Hospital de la Santa Creu i Sant Pau (HSCSP, Spain),[45] and 4 from the Hospital General de L'Hospitalet (HGH, Spain). All volunteers gave written informed consent before participating. The operating protocols and the informed consent were approved by the Research Ethics Committees from the HUB [CEIm PR107/21(CSI 21/25)], Ethics Committee for Research with Medicines of Euskadi (CEIm-E PI2021087), and Ethics Committee for Research with Medicines of the HSCSP [CEIm 22/267 (OBS)]. The informed consent for the Knight cohort samples was approved by the Washington University School of Medicine Institutional Review Board and Ethics Committee (IRB 201104178) and the institutional ethics committees of Austin Health, St Vincent's Health, Hollywood Private Hospital and Edith Cowan University approved the AIBL study.

Blood and DNA samples were obtained from two groups, CN subjects [Clinical Dementia Rating (CDR) score equal to 0] and MCI (CDR=0.5), across the different cohorts. Specifically, the CN group includes samples from consecutively recruited patients stored in private registered repositories from the AIBL and CITA biobanks. The MCI group includes samples longitudinally recruited (prospective samples) in HUB, HCB, HGH, and samples from the HCB, HSCSP, AIBL, and Knight repositories. The sex, age, years of schooling, and clinical data of each group are reported in Table 1. Table S1 shows the number and percentages of subjects recruited from each cohort and sample collection type.

Inclusion and exclusion criteria for participants

The inclusion criteria for MCI patients were: (1) Written informed consent, (2) A global CDR of 0.5,[46] amnestic and not qualifying for a syndromic assignment of mild AD dementia, (3) Meeting the National Institute on Aging- Alzheimer's Disease Association criteria for MCI,[34] based on the consensus opinion of the evaluating team members, and (4) between 55 and 85 years of age. The inclusion criteria for CN subjects were: (1) Subjects who have previously signed informed consent for participation in future research studies at the time of sample collection, (2) CDR=0 and not qualifying for a syndromic diagnosis of MCI.

Patients were excluded from participation in this study if at least one of the following criteria were met: (1) Patients who couldn't adequately perform the neuropsychological tests or participate in the clinical assessment, and (2) Patients with any of the following conditions: Familial AD with a monogenic autosomal dominant pattern, history of severe traumatic brain injury, a history of major intracranial haemorrhage, major depression or another severe psychiatric disease, history of active cancer (excluding cutaneous squamous cell carcinoma), normal pressure hydrocephalus, clinically significant vascular brain disease, known HIV infection, and systemic or metabolic diseases that in the investigator's opinion could cause cognitive impairment and/or interfere with the conduct of the study.

Clinical criteria for dementia progression

An MCI patient was considered to progress to dementia[47] when one of the following criteria was met: CDR change from 0.5 to $\geq$1, or $\geq$2-point deterioration in the CDR-Sum of Boxes (SOB) from baseline,[48] or $\geq$1-point deterioration in at least four instrumental activities of daily living measured by the Functional Activities Questionnaire (FAQ).[49] For the prospective cohort, the clinical follow-up was performed every 12 months or according to the clinical practice of each investigational site. Samples from biobanks showed a clinical follow-up every 18 months.

METHOD DETAILS

DNA extraction

Total DNA was isolated from whole EDTA blood using a Maxwell® RSC Blood DNA Kit (Promega Biotech Ibérica S.L., Madrid, Spain) in Promega's Maxwell® RSC System automated DNA extraction platform according to the manufacturer's instructions. DNA quantity and quality were determined using a NanoDrop™ OneC spectrophotometer (Thermo Fisher Scientific, Madrid, Spain). DNA was stored at -80°C.

Bisulfite treatment

DNA samples (300 ng) were treated with bisulfite using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's protocol (EZ DNA Methylation™ Kit, Instruction Manual Ver.1.2.7). The M-Dilution Buffer step was incubated at 42°C for 30 minutes and the CT conversion reagent incubation was set at 50°C for 15 hours. Finally, the bisulfite converted DNA was resuspended in 30 μL of molecular grade nuclease free water.

Primer design

D-loop and ND1 primers were designed to amplify the following human mitochondrial regions based on the NC_012920.1 GenBank NCBI reference sequence (D-loop: nt. 16,465-230; *MT-ND1,* also known as *ND1*: nt. 3,257-3,682). These primers did not amplify nuclear pseudogenes in a human mtDNA-depleted ρ0 cell line (Figure S1).

To generate the amplicon libraries to be sequenced on the Illumina MiSeq™ System (Illumina, Eindhoven Noord-Brabant, Netherlands), fusion primers were created with an Illumina overhang adapter sequence fused to the 5' of the locus-specific primers according to Illumina's 16S metagenomic library preparation protocol (Illumina, Part # 15044223 Rev. B).

Customized primers were synthesized with desalted purification by Merck Life Science S.L. (Madrid, Spain). The primers, including the Illumina's overhang adapters (underlined part), are shown below:

D-loop-Forw: 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGYAYTTGGGGGTA GYTAAAGTGAAYTG 3'; D-loop-Rev: 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCCTACAARCATTAATTAATTAACACAC 3'; ND1-Forw: 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATAAAAYT TAAAAYTTTAYAGTYAGAG 3'; ND1-Rev: 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTRARTTTRATRCTCACCCTRATCA 3'

The primers contain degenerated bases. The degenerated cytosines in the forward primers are shown as Y (C/T) and the degenerated guanines in the reverse primer are shown as R (G/A).

PCR for D-loop and *ND1*

Five microliters of the bisulfite-treated DNA samples were used for PCR. Amplification of the bisulfite-converted DNA was performed in a reaction volume of 25 μL. Each PCR reaction mix consisted of a final concentration of: 1× FastStart 10× Buffer2, 0.05 U/μL FastStart HiFi Polymerase (Roche Diagnostics S.L., Sant Cugat, Spain), 200 μmol/L of each dNTP (Meridian Bioscience, Cincinnati, USA), and 400 nmol/L of each primer forward and reverse (Merck Life Science S.L). *ND1* amplification required 5% DMSO. Amplifications were performed in a SimpliAmp™ Applied Biosystems™ Thermocycler (Applied Biosystems, Madrid, Spain) using the following conditions:

ND1 amplicon: 94°C for 3 minutes followed by 38 cycles of 94°C for 15 seconds, at 54°C for 45 seconds and 72°C for 1 minute, followed by a final extension step at 72°C for 8 minutes and a final hold at 4°C.

D-loop amplicon: 94°C for 3 minutes followed by 36 cycles at 94°C for 15 seconds, at 58°C for 45 seconds and 72°C for 1 minute, followed by a final extension step at 72°C for 8 minutes and a final hold at 4°C.

Three microliters of each PCR product were checked for amplification using a 1.5% agarose gel stained with Sybr™ Safe DNA Gel Stain (Invitrogen, Madrid, Spain).

Amplicon library preparation

PCR products were purified in a clean-up step using the paramagnetic beads capture reagent Agencourt® AMPure® XP—60 mL (Beckmann Coulter S.L.U, Madrid, Spain) and resuspended in a final volume of 14 μL of Elution buffer (10 mM Tris-Cl, pH 8.5). The amplicons were then dual indexed in a PCR step with a 50 μL final volume. In the reaction, 5 μL of the purified PCR amplicons, 10 μL of unique index identifiers IDT® for Illumina® DNA/RNA UD Indexes Set A, B, C and D (Illumina), 25 μl of high-fidelity polymerase ready mix, KAPA HiFi HotStart Ready Mix (Roche), containing 0.5U KAPA HiFi HotStart DNA Polymerase, 0.3 mM of each dNTP at 1X and MgCl2 (2.5 mM at 1X), and 10 μL of molecular grade nuclease-free water. The index PCR reaction conditions were: 95°C for 3 minutes followed by 8 cycles at 95°C for 30 seconds, at 55°C for 30 seconds and 72°C for 30 seconds, followed by a final extension step at 72°C for 5 minutes and a final hold at 4°C.

After the indexing PCR, a further purification step, the second clean-up, was performed to purify the libraries from excess adapters using the Agencourt® AMPure® XP—60 mL (Beckmann Coulter S.L.U.). The libraries generated were then quantified in a Qubit™ 3 Fluorometer (Invitrogen) using the Qubit™ Double-Stranded DNA Broad Range assay kit (Invitrogen). Libraries were normalized and pooled into a single 10 nM pool. The single pool was denatured and diluted to 7 pM. To introduce variability into the low diversity libraries, the PhiX control v3 (Illumina) was denatured and diluted to 10 pM. Finally, the MiSeq® Reagent Kit v3 600 cycles (Illumina) was loaded with 75 % by volume of the 7 pM library pool and 25 % by volume of the PhiX v3 control at 10 pM for sequencing on an Illumina MiSeq™ System.

High-throughput sequencing

Sequencing was performed on an Illumina MiSeq™ System platform using 300 bp paired-end reads to cover the length of the amplicons. Amplicon libraries from bisulfite-treated DNA resulted in low-diversity libraries. To compensate for this, the Illumina® PhiX Control v3 Library was added to the sequencing run. PhiX provides balanced fluorescent signals that are lacking in low diversity libraries, improving the quality of the run. To ensure accurate quantification of methylation levels in all cytosine contexts, sequencing was designed to achieve in-depth coverage (>5000x per site). The quality of the sequencing run was assessed using Illumina's metrics for the MiSeq® Reagent Kit v3 (2x300bp), which requires a quality score of Q30 or higher for more than 70 % of the bases.

QUANTIFICATION AND STATISTICAL ANALYSIS

Methylation data processing and analysis

Differentially methylated sites in CpG, CHG, and CHH contexts were identified by processing the raw sequencing reads generated from the two mitochondrial amplicons, D-loop and ND1, through an *in-house* bioinformatics workflow organized into three main stages: (1) Extracting methylation information from the sequencing data, (2) performing an exploratory data analysis (EDA) to assess global methylation patterns and potential biases, and (3) conducting a differential methylation data analysis (DMA) to statistically identify significant changes in methylation levels between the groups of patients.

Bioinformatics workflow for extracting methylation data

To guarantee high-quality data for downstream analysis, the first step was to evaluate the raw reads using FastQC (v0.11.9).[37] Trimmomatic (v0.39) was used to filter and trim low-quality bases and adapter sequences.[38] A Phred quality threshold of 28 was applied, and up to 250 bases from the start of each read were retained. Trimmed reads were reassessed using FastQC to verify the enhancement in data quality. Reads were aligned to the human reference genome (GRCh38) with the bisulfite-aware aligner Bismark.[39] The slope parameter was adjusted to -0.6 for the function that determines the minimum alignment score, while all other recommended parameters were adhered to by default. Methylation calling was performed using Bismark to identify methylation states for each site across the three contexts (CpG, CHG, and CHH). Quantification of methylation levels was subsequently carried out using custom *in-house* R (v4.0.0) scripts.[40] This process involved: (1) Extraction of methylation proportions, (2) assessment of the quality of sequence alignment, (3) generation of methylation profiles for each sample, and (4) construction of methylation matrices for each amplicon and context.

Exploratory data analysis of methylation patterns

EDA was performed to describe the methylation measures in detail. For each measure, statistics on central tendency, dispersion, and symmetry were estimated to characterize their distribution. Boxplots were generated for each cytosine site across groups and contexts to visualize distributions, detect potential outliers, and identify trends. Methylation levels were plotted across cytosine sites within the CpG, CHG and CHH contexts, with the x-axis representing base positions relative to the start of each analyzed amplicon rather than absolute mtDNA genomic coordinate, enabling site-specific comparisons across all samples. Overall methylation patterns and potential biases were assessed using R (v4.4.0),[40] along with the tidyverse (v2.0.0) collection packages,[41] and bash scripting.[42]

Differential methylation analysis

To ensure reliable detection of differentially methylated sites (DMS), data standardization was performed prior to the analysis using the R stats package (v4.4.0).[40] This normalization step mitigated systematic variability and potential batch effects, ensuring that methylation levels were comparable across samples and conditions. DMA was then conducted using the limma Bioconductor package (v3.60.2),[43] which applies linear models tailored for high-dimensional data and employs empirical Bayesian methods to stabilize variance estimates and improve the robustness of the results.

To control for multiple testing, three p-value adjustment methods were evaluated: (1) the Benjamini-Hochberg (BH) method,[50] commonly used for high-dimensional datasets under the assumption of independence or positive dependence among tests; (2) the Benjamini-Yekutieli (BY) method,[51] which controls the False Discovery Rate (FDR) in more general settings, offering a more conservative adjustment; and (3) the Bonferroni correction,[52] which controls the Family-Wise Error Rate (FWER) through a stringent threshold. Among these, BY-adjusted p-values were selected as the primary threshold ($BY<0.05$) due to their robustness in controlling false discoveries under potential test dependencies.

Development of the classification model

After confirming the existence of differential mitochondrial epigenetic patterns among blood samples from MCI patients who progressed to ADD (P) and MCI patients who did not progressed to ADD (NP), a workflow involving four main steps was implemented. The process comprised (1) exploratory data analysis, (2) data preprocessing, (3) model training, and (4) model performance evaluation.

Exploratory data analysis

The Exploratory Data Analysis (EDA) was conducted to characterize all study variables. For categorical variables, frequency tables were generated. Measures of central tendency, dispersion, and symmetry were determined to properly analyze continuous variables. Boxplots were produced to enhance the visualization of continuous variable distributions. Boxplots were also built to help depict the distribution of continuous variables.

Data preprocessing

The preprocessing pipeline was designed to prepare the dataset for efficient and accurate model training. One-hot encoding was used to convert categorical variables to numerical ones, ensuring no linear dependencies. Variables with zero or near-zero variance

were removed to avoid instability and computational failure. To mitigate model performance-affecting redundancy and multicollinearity, redundant variables due to linear combinations were also removed.

The dataset was randomly split into two subsets: a training dataset (80% of the data) for model development, and an independent test dataset (20% of the data) for evaluating the model's final performance on unseen data. The split was stratified by group to preserve the overall distribution of groups in both datasets. A subsequent EDA was conducted on the training dataset to confirm the absence of biases. The transformations of centering (mean subtraction) and scaling (division by the standard deviation) were used to normalize the continuous variables in the training dataset. The derived factors were then applied on both the training and testing datasets.

To address the high dimensionality of the dataset, a Genetic Algorithm (GA) was initially employed, followed by a Principal Components Analysis (PCA). During the feature selection process, a supervised binary search was conducted across the predictor space to identify the optimal feature subsets. This optimization was achieved by applying triply repeated 10-fold cross-validation (CV), performed exclusively within the training dataset and assessing model accuracy at each potential cutoff value, thereby ensuring the selection of the most predictive features based on their contribution to model performance.

Model training

Random Forest (RF) was selected as the most suitable algorithm for constructing the classification model after evaluating several classical supervised learning methods. RF was chosen for its superior performance on the dataset, its ability to handle both continuous and categorical variables, its robustness against overfitting, and its capacity to provide insights into variable importance. These attributes made RF particularly well-suited for the structured and diverse nature of the dataset.

The RF model was configured to grow 500 decision trees. Each tree was constructed using a random subset of features, with the number of features sampled at each split (mtry) optimized through grid search. This hyperparameter tuning was performed by evaluating a range of mtry values and selecting the configuration that maximized accuracy during triply repeated 10-fold CV.

During training, group imbalance in the dataset was addressed by applying weights inversely proportional to the group frequencies, ensuring equitable representation of all groups. Weights were normalized to prevent bias during the learning process. The final RF model was then trained on the entire training subset using the optimized hyperparameters.

Model performance evaluation

The performance of the RF model was evaluated using a comprehensive framework encompassing both training and testing processes. During the training process, the model's performance was assessed through the triply repeated 10-fold CV, calculating metrics such as the Accuracy and Cohen's Unweighted Kappa statistic. These metrics were averaged across all resampling iterations to ensure consistent performance and generalizability. Statistical summaries, including mean, median, minimum, maximum and the first and third quartiles, were calculated to assess the consistency of the model's performance across the CV iterations. Additional metrics, such as log-loss, AUC, prAUC, and others, were also calculated.

For the testing dataset, performance metrics including Accuracy, Kappa Statistic, Sensitivity (Recall), Specificity, Positive Predicted Value (PPV, *aka* Precision), Negative Predicted Value (NPV), and F1-score were derived from the confusion matrix. These metrics were used to evaluate the classification performance of the RF model and to compare the proposed model with alternative configurations involving specific clinical variables.

Receiver Operating Characteristic (ROC) curves were generated for each group using a One-vs-Others (OvO) approach. Each group was treated as the positive group, while the remaining groups were considered negative (i.e. PvO, NPvO and CNvO, where PvO corresponds to P vs CN+NP, NPvO to NP vs CN+P, and CNvO to CN vs NP+P). The Area Under the Curve (AUC-ROC) was calculated for each group, providing a measure of the model's ability to distinguish across groups. Optimal cutoff thresholds for each group were determined through post-hoc analysis of the model's performance, which utilized Youden's index method.[53] These thresholds, determined after training using the test dataset, were applied to assess the model's discrimination performance. This post-hoc analysis allowed a deeper investigation of the trade-offs between specificity and sensitivity for each group. Confidence intervals for the AUC values were calculated to assess the reliability of the model's discriminatory capability.