Received: 6 December 2023 | Revised: 3 May 2024 | Accepted: 6 May 2024

DOI: 10.1002/alz.14061

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Blood-based multivariate methylation risk score for cognitive impairment and dementia

Jarno Koetsier[1,2] | Rachel Cavill[3] | Rick Reijnders[1] | Joshua Harvey[4] | Jan Homann[5] | Morteza Kouhsar[4] | Kay Deckers[1] | Sebastian Köhler[1] | Lars M. T. Eijssen[1,6] | Daniel L. A. van den Hove[1,7] | Ilja Demuth[8,9] | Sandra Düzel[10] | for the Alzheimer's Disease Neuroimaging Initiative | Rebecca G. Smith[4] | Adam R. Smith[4] | Joe Burrage[4] | Emma M. Walker[4] | Gemma Shireby[4] | Eilis Hannon[4] | Emma Dempster[4] | Tim Frayling[4] | Jonathan Mill[4] | Valerija Dobricic[11] | Peter Johannsen[12] | Michael Wittig[13] | Andre Franke[13] | Rik Vandenberghe[14] | Jolien Schaeverbeke[14] | Yvonne Freund-Levi[15,16,17] | Lutz Frölich[18] | Philip Scheltens[19] | Charlotte E. Teunissen[20] | Giovanni Frisoni[21] | Olivier Blin[22] | Jill C. Richardson[23] | Régis Bordet[24] | Sebastiaan Engelborghs[25,26] | Ellen de Roeck[25] | Pablo Martinez-Lage[27] | Mikel Tainta[27] | Alberto Lleó[28] | Isabel Sala[28] | Julius Popp[29] | Gwendoline Peyratout[30] | Frans Verhey[1] | Magda Tsolaki[31] | Ulf Andreasson[32] | Kaj Blennow[32,33,34,35] | Henrik Zetterberg[32,36,37,38,39] | Johannes Streffer[40] | Stephanie J. B. Vos[1] | Simon Lovestone[41] | Pieter-Jelle Visser[1,17] | Christina M. Lill[5,42] | Lars Bertram[11] | Katie Lunnon[4] | Ehsan Pishva[1,4]

Funding information: ZonMw Memorabel/Alzheimer Nederland, Grant/Award Number: 733050516; Innovative Medicines Initiative Joint Undertaking, Grant/Award Number: 115372; European Union's Seventh Framework Program, Grant/Award Number: FP7/2007-2013; National Institute for Health and Care Research; Biomedical research Centre; Heisenberg grant of the German Research Foundation, Grant/Award Number: LI 2654/4–1; EU Joint Programme – Neurodegenerative Disease Research 2021″, Grant/Award Numbers: JPND2021, EPIC4ND; Cure Alzheimer's Fund, Grant/Award Number: EPIC4AD; European Union's Horizon Europe research and innovation programme, Grant/Award Number: 101053962; Swedish State Support for Clinical Research, Grant/Award Number: #ALFGBG-71320; Alzheimer Drug Discovery Foundation (ADDF), USA, Grant/Award Number: #201809-2016862; AD Strategic Fund and the Alzheimer's Association, Grant/Award Numbers: #ADSF-21-831376-C, #ADSF-21-831381-C, #ADSF-21-831377-C; Bluefield Project; Olav Thon Foundation; Familjen Erling-Perssons Stiftelse; Stiftelsen för Gamla Tjänarinnor; Marie Skłodowska-Curie, Grant/Award Number: 860197; European Union Joint Programme – Neurodegenerative Disease Research, Grant/Award Number: JPND2021-00694; National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre; UK Dementia Research Institute; Swedish Alzheimer Foundation, Grant/Award Numbers: #AF-930351, #AF-939721, #AF-968270; Hjärnfonden, Sweden, Grant/Award Numbers: #FO2022-0270, #FO2017-0243, #ALZ2022-0006; ALF-agreement, Grant/Award Numbers: #ALFGBG-715986, #ALFGBG-965240; European Union Joint Program for Neurodegenerative Disorders, Grant/Award Number: JPND2019-466-236; Alzheimer's Association 2021 Zenith Award, Grant/Award Numbers: ZEN-21-848495, SG-23-1038904 QC; Alzheimer's Society UK, Grant/Award Number: AS-PG-14-038; Medical Research Council; National Institute of Aging (NIA); National Institutes of Health (NIH), Grant/Award Number: R01AG067015; Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health, Grant/Award Number: U01 AG024904; DOD ADNI (Department of Defense award, Grant/Award Number: W81XWH-12-2-0012; National Institute on Aging; National Institute of Biomedical Imaging and Bioengineering; Swedish Research Council, Grant/Award Numbers: #2022-01018, #2019-02397, #2017-00915, #2022-00732

Correspondence
Ehsan Pishva, Department of Psychiatry and Neuropsychology, Mental Health and Neuroscience Research Institute (MHeNs), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, 6200 MD Maastricht, The Netherlands.
Email: e.pishva@maastrichtuniversity.nl

Abstract

INTRODUCTION: The established link between DNA methylation and pathophysiology of dementia, along with its potential role as a molecular mediator of lifestyle and environmental influences, positions blood-derived DNA methylation as a promising tool for early dementia risk detection.

METHODS: In conjunction with an extensive array of machine learning techniques, we employed whole blood genome-wide DNA methylation data as a surrogate for 14 modifiable and non-modifiable factors in the assessment of dementia risk in independent dementia cohorts.

RESULTS: We established a multivariate methylation risk score (MMRS) for identifying mild cognitive impairment cross-sectionally, independent of age and sex ($P = 2.0 \times 10^{-3}$). This score significantly predicted the prospective development of cognitive impairments in independent studies of Alzheimer's disease (hazard ratio for Rey's Auditory Verbal Learning Test (RAVLT)-Learning = 2.47) and Parkinson's disease (hazard ratio for MCI/dementia = 2.59).

DISCUSSION: Our work shows the potential of employing blood-derived DNA methylation data in the assessment of dementia risk.

KEYWORDS
aging, Alzheimer's disease, dementia, DNA methylation, epigenetics, machine learning, mild cognitive impairments, Parkinson's disease, risk prediction

Highlights

- We used whole blood DNA methylation as a surrogate for 14 dementia risk factors.
- Created a multivariate methylation risk score for predicting cognitive impairment.
- Emphasized the role of machine learning and omics data in predicting dementia.
- The score predicts cognitive impairment development at the population level.

1 | BACKGROUND

Over the past decades, the most precise models for predicting dementia and Alzheimer's disease (AD) have relied on molecular information derived from cerebrospinal fluid (CSF) and neuroimaging modalities.[1] More recently, advancements in the development of blood-based biomarker assays for neurodegenerative disorders, including amyloid beta (Aβ) isoforms, phosphorylated tau (p-tau) proteins, neurofilament light (NfL), and glial fibrillary acidic protein (GFAP), have prompted a notable transition toward emphasizing the importance of blood-based biomarkers in AD research.[2] The adoption of a reliable blood-based assay offers potential solutions to longstanding challenges, such as the relative invasiveness of CSF sampling, and the high costs and limited accessibility of specialized neuroimaging facilities.

DNA methylation is a key epigenetic mechanism involved in the molecular pathology of dementia.[3] This molecular mechanism is known to mediate the impact of lifestyle and environmental factors on the genome, by regulating gene expression.[4] Notably, prior research has demonstrated associations between peripheral DNA methylation patterns and risk factors for dementia, such as smoking,[5] obesity,[6] and blood pressure.[7] Moreover, recent comprehensive DNA methylation Quantitative Trait Locus (mQTL) analyses have confirmed a substantial genetic influence on methylation patterns.[8] This collective evidence positions DNA methylation as an intriguing molecular biomarker with the potential to capture both genetic and environmental information at the individual level. However, previous endeavors to establish blood-derived DNA methylation-based predictions for AD have encountered challenges in external validation, likely attributed to the heterogeneity of the disease among different cohorts.[9]

Therefore, in this study, instead of utilizing DNA methylation for predicting defined diagnostic labels for dementia and cognitive decline, we sought to initially develop objective and precise DNA methylation-based prediction models for the "LIfestyle for BRAin Health" (LIBRA)[10] and "Cardiovascular Risk Factors, Aging, and Dementia" (CAIDE)[11] scores. LIBRA and CAIDE are two established frameworks that

underscore the significance of modifiable (lifestyle-related) and non-modifiable risk factors in the context of cognitive decline and dementia prevention. In addition to developing the epigenetic score for CAIDE and LIBRA total scores (i.e., epi-CAIDE and epi-LIBRA), we established methylation profile scores (MPSs) for the individual dementia risk factors involved in these two scores such as age, sex, apolipoprotein E (APOE) ε4 genotype, smoking,[12] alcohol intake,[13] plasma cholesterol levels,[14] physical activity,[15] education,[16] and diet[17] in a midlife general population cohort. Next, we employed these individual models to generate a multivariate methylation risk score (MMRS) for prediction of AD dementia and mild cognitive impairment (MCI) in a cross-sectional pre-dementia and AD cohort. Last, we validated the performance of the generated scores in predicting the cross-sectional status and the prospective development of cognitive impairments and dementia in three independent cohorts of aging, AD, and Parkinson's disease (PD).

RESEARCH IN CONTEXT

1. **Systematic review**: Although DNA methylation data is regarded as a molecular mediator that links lifestyle to health and disease, after a careful evaluation of the literature, we did not find any study that successfully utilized blood-derived DNA methylation data to identify people at risk of developing dementia.
2. **Interpretation**: Constructing a blood-derived DNA methylation-based model with cross-sectional cognitive impairment status as the target variable, we associated the risk score with both the current status and the future onset of cognitive impairments in independent dementia cohorts.
3. **Future directions**: Our work establishes a framework for future studies aiming to enhance predictive performance by (a) assessing a broader array of machine learning algorithms, (b) modeling the trajectory of cognitive decline rather than relying solely on the cross-sectional state of cognitive function, (c) utilizing larger risk factor-specific datasets to improve prediction, and (d) incorporating additional omics layers for further enhancement of predictive performance.

2 | METHODS

2.1 | Study design

The applied methodology consists of four main steps (Figure 1): (1) The model generation using the DNA methylation data obtained from the Exeter 10,000 project (EXTEND) cohort[18] and the European Medical Information Framework for Alzheimer's Disease multimodal biomarker discovery (EMIF-AD MBD) study,[19] (2) Model validation in the independent test set of the EMIF-AD MBD study, Parkinson's Progression Markers Initiative (PPMI),[20] the Alzheimer's Disease Neuroimaging Initiative (ADNI),[21] and Berlin Aging Study-II (BASE-II) cohorts,[22] (3) Model interpretation in terms of variable importance, gene ontology (GO) overrepresentation analysis, and the influence of genetic variation, and (4) Model extension by adding genetic and CSF biomarkers to the model.

2.2 | Study cohorts

An overview, including a description, sample size, and sex and age distribution, of the five cohorts used in the current research is provided in Table S1.

2.2.1 | The EXTEND cohort

The EXTEND study is a UK-based National Institute for Health and Care Research (NIHR) funded project aiming to collect blood samples along with extensive health information from people with and without health conditions.[18] In the current study, we used a subset of the EXTEND cohort, comprising 1076 individuals with available genotyping and blood-derived DNA methylation data. This subset exclusively comprised phenotype data from individuals aged 40–75, denoted as the midlife age group.

2.2.2 | The EMIF-AD MBD study

The EMIF-AD MBD study is a European research initiative focused on collecting and collating comprehensive medical and health-related information from individuals affected by AD and related cognitive disorders.[19] In the present study, we excluded the individuals outside the midlife age range (i.e., age < 40 or age > 75). We also excluded individuals who were initially cognitively healthy but later developed MCI or AD (mean follow-up time $\pm$ standard deviation [SD] $\approx 2.3 \pm 1.2$ years). This exclusion helps in limiting ambiguous cases in the model training process. The outcome of this process was a final dataset comprising 110 individuals with AD, 293 individuals with MCI, and 220 healthy controls with available blood-derived DNA methylation data, genotyping data, and/or CSF protein markers.

2.2.3 | The ADNI cohort

The ADNI cohort[21] contains genetic and blood-derived DNA methylation data from cognitively normal individuals, as well as individuals diagnosed with MCI and AD dementia recruited by institutions from the United States and Canada. In addition to the DNA methylation data, ADNI also has extensive individual-level information on various psychometric biomarkers measured at multiple time points. To validate the models, we used the baseline DNA methylation data of 251 midlife individuals ($40 \leq$ age ≤ 75).

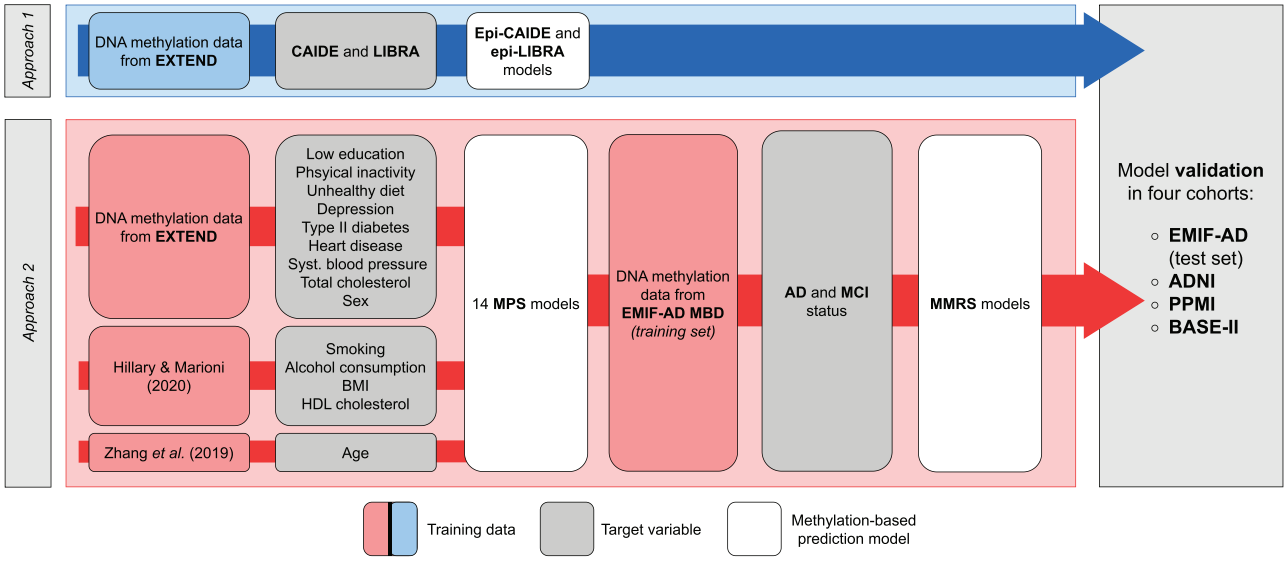

FIGURE 1 Overview of the model generation pipeline. The model generation workflow consists of model training in the EXTEND and EMIF-AD cohorts using two different approaches. In the first approach, models for the prediction of CAIDE and LIBRA were trained in the EXTEND cohort ($n = 1076$). Furthermore, in the second approach, the DNA methylation data obtained from the EXTEND cohort was used to predict 14 known dementia risk factors. The 14 predicted risk scores (i.e., methylation profile scores; MPSs) were next used as variables for the prediction of AD and MCI status in the training set of the EMIF-AD cohort ($n = 436$). The resulting multivariate methylation risk scores for AD (MMRS-AD) and MCI (MMRS-MCI) were evaluated in terms of AD and MCI classification performance in the independent test set in the EMIF-AD cohort ($n = 187$). The model with the best performance was also validated in the ADNI ($n = 223$), PPMI ($n = 129$), and BASE-II cohorts ($n = 1017$).

2.2.4 | The PPMI cohort

The PPMI cohort[20] includes blood-derived DNA methylation data from individuals of multiple nationalities who have recently been diagnosed with PD. For validation, only samples from individuals aged 40 to 75 were included, resulting in 129 samples with available baseline DNA methylation data and cognitive impairment outcome information. Publically available data used in the preparation of this article were obtained on May 17, 2023 from the PPMI database (https://www.ppmi-info.org/access-data-specimens/download-data), RRID:SCR_006431. For up-to-date information on the study, visit http://www.ppmi-info.org.

2.2.5 | The BASE-II cohort

The BASE-II cohort is a German multi-institutional longitudinal study characterizing individual changes in the aging process (described in detail elsewhere).[22] Briefly, the BASE-II core cohort (n ≈ 2200) with multi-dimensional data available comprises a group of elderly (n ≈ 1600, age ≥ 60) and younger participants (n ≈ 600, age = 20-35). For this study, we included the group of elderly participants only. The effective dataset investigated here comprised 1,017 elderly adults for whom whole blood DNA methylation data at baseline, and cross-sectional cognitive phenotypic data at baseline were available. For 984 individuals in this baseline sample, longitudinal cognitive follow-up was available (follow-up time ≈ 7.5 years).

2.3 | Data pre-processing

2.3.1 | Clinical outcomes

The cognitive status of individuals in the EMIF-AD MBD study was defined as described previously.[19] In summary, cognitively healthy individuals were defined by a normal neuropsychological assessment score. The MCI diagnosis in nine of the EMIF-AD MBD subcohorts was based on the criteria of Petersen,[23] while for two subcohorts, the Winblad et al. criteria[24] were used. Furthermore, the diagnosis of AD dementia was defined based on the criteria of the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA).[25]

Due to the reported significant number of false positives MCI diagnosis in ADNI,[26] our analysis in the ADNI cohort shifted its focus to examine baseline and longitudinal measures of various cognitive domains assessed over four years. These measures included Alzheimer's Disease Assessment Scale (ADAS), Rey's Auditory Verbal Learning Test (RAVLT), Trail Making Test Part B Time (TMT), and Mini-Mental State Examination (MMSE) scores. For the MMSE score, cognitive impairments were defined as MMSE < 24. For the other cognitive outcomes, impairment status was determined by a score of either 2 SD below or above the mean of the control group, depending on the direction of the score (Table S2).

In the PPMI cohort, dementia and MCI diagnoses were based on the Movement Disorders Society (MDS) recommended criteria.[27] MCI and PD dementia individuals recorded as having reverted to a cognitively

TABLE 1 Risk factor definitions in the EXTEND cohort.

Risk factor	Definition	Risk factor model[a]	CAIDE[b]	LIBRA[b]
Low education	A person is defined to be lowly educated if the individual has none of the following educational achievements: 1. College or university degree. 2. O level, GCSEs, or equivalent. 3. NVQ, HND, HNC, or equivalent. 4. A-level, AS-level, or equivalent. 5. CSEs or equivalent. 6. Other professional qualifications	Training in EXTEND	X	
Physical inactivity	A person is defined to be physically inactive if self-reported to do exercise with an increased pulse of more than 2.5 h per week.	Training in EXTEND	X	X
Unhealthy diet	Self-reported consumption of three or less fruits/vegetables per day	Training in EXTEND		X
Depression	Self-reported depression status	Training in EXTEND		X
Type II diabetes	Self-reported type II diabetes status	Training in EXTEND		X
Heart disease	Self-reported heart disease status	Training in EXTEND		X
Sex	Sex	Training in EXTEND	X	
Systolic blood pressure	Mean systolic blood pressure (mmHg)	Training in EXTEND		X
Total cholesterol	Log-transformed total serum cholesterol levels (mmol/L)	Training in EXTEND	X	
Age	Chronological age	Zhang et al. (2019) model	X	
BMI	Body mass index	Hillary & Marioni (2020) model	X	X
Smoking	Self-reported current smoking status	Hillary & Marioni (2020) model		X
HDL cholesterol	Log-transformed serum HDL cholesterol levels (mmol/L)	Hillary & Marioni (2020) model		X
Alcohol intake	High alcohol intake is defined by the following criteria: 1. Once a month, more than 10 alcoholic drinks per day. 2. 2-4 a month, 5 or more alcoholic drinks per day 3. 2-3 a week, 5 or more alcoholic drinks per day 4. 4 or more a week, 3 or more alcoholic drinks per day	Hillary & Marioni (2020) model		X
Kidney disease[c]	Self-reported chronic kidney disease status	NA[c]		X

Note: The risk factors in the EXTEND cohort were used for the calculation of the CAIDE and LIBRA scores and as target variables for the training or validation of the DNA methylation-based risk factor models. The risk factor weights and cutoffs used for the calculation of the LIBRA and CAIDE scores.
[a]This column indicates the DNA methylation-based model that was used for the prediction of the corresponding dementia risk factor (approach 2).
[b]This column indicates whether the corresponding risk factor was used in the calculation of the LIBRA and CAIDE score (approach 1). See Tables S3 and S4 for the exact risk factor weights.
[c]No risk factor model was trained for kidney disease status due to the limited number of cases in the EXTEND cohort (i.e., 17 cases and 1059 controls).

normal status (absence of MCI and PD dementia) were excluded from the analysis.[28]

Finally, the quantitative assessment of episodic memory, working memory performance, spatial memory, and fluid intelligence of BASE-II participants was conducted using seven neuropsychological tests (Face-Profession Task, Verbal Learning and Memory Test, Number-N-Back task, Spatial Updating Task, Letter Updating Task, Object Location Task, and Figural Analogies).[29]

2.3.2 | Dementia risk factors

Dementia risk factors generated in the EXTEND cohort are listed in Table 1. Specifically, 15 dementia risk factors were used for the calculation of the CAIDE and LIBRA scores based on the previously established weights[10,11] (Tables S3 and S4). It should be noted that information about a person's cognitive activity was not available in the EXTEND cohort and therefore was not used in the calculation of the LIBRA score. Additionally, due to the small number of (chronic) kidney disease cases, only 14 dementia risk factors were used as target variables for the construction and/or validation of the DNA methylation-based risk factor models.

2.3.3 | Cerebral spinal fluid biomarkers

Z-scores for three CSF biomarkers, amyloid-β (Aβ), phosphorylated tau181p (p-tau), and total tau (t-tau), were defined in the EMIF-AD MBD study as previously described.[19] In summary, the t-tau and p-tau z-scores were defined by their local p-tau levels as measured by

center-specific enzyme-linked immunosorbent assays (ELISAs), standardized within assay according to the mean and standard deviation of the healthy controls. Furthermore, the Aβ z-scores were specified to be standardized scores for amyloid pathology. Specifically, this score was standardized within-assay according to the mean and standard deviation of the control group and was based on the CSF Aβ42/40 ratio from central analyses, local CSF-Aβ42, and the standardized uptake value ratio (SUVR) on an amyloid PET scan. In ADNI, the t-tau, p-tau, and Aβ scores were determined as described by Shaw et al.[30]

2.3.4 | Whole blood DNA methylation profiling

DNA methylation for all samples in the four studied cohorts was quantified using the Illumina Infinium Human Methylation EPIC BeadChip Array (EPIC array) (Illumina), which interrogates over 850,000 methylation sites throughout the genome. Sample filtering was performed as a first step of the DNA methylation data pre-processing pipeline and includes the removal of samples with a median bisulfite conversion rate below 80%, incorrect sex labels, and a low median (un)methylated intensity according to the minfi package's guidelines (i.e., median $\log_2$ unmethylated intensity + median $\log_2$ methylated intensity ≤ 21).[31]

In the normalization procedure, the combination of Noob (minfi package (v1.46.0)[31]) and BMIQ (wateRmelon package (v2.6.0)[32]) normalization was applied. This pipeline has previously been shown to be a high-performing method for reducing type I/type II bias and enhancing reproducibility.[33] Furthermore, both Noob and BMIQ are within-sample normalization methods, which avoid information leakage from the training to the test set.

Before model training in the EXTEND cohort, previously reported cross-reactive probes,[34] probes with a detection $P > 0.01$ in at least one sample, sex-chromosome probes, and probes with single nucleotide polymorphisms (SNPs) at the single base extension and/or CpG interrogation site were removed. Next, density plots and principal component analysis (PCA) score plots were constructed to assess the quality of the pre-processing and to identify possible outliers. Last, β-values were converted to M-values to account for the inherently heteroscedastic nature of methylation data. The M-values were used for the subsequent feature selection and machine learning pipeline.

It should be noted that not all features that are incorporated in the generation of the risk factor models passed the described probe filtering steps in the other cohorts. Therefore, when applying the risk factor models in a different cohort, low-quality probes (i.e., detection $P > 0.1$) were imputed using the imputePCA function from the missMDA package (v1.8).[35] This function uses a regularized iterative PCA algorithm to impute missing values. Specifically, this algorithm first imputes all missing values with the feature's mean after which it iteratively performs PCA and imputes each missing value using the low-rank representation until convergence is reached (i.e., a difference of less than 1×10^{-6} between iterations). The optimal number of principal components (PCs) for the low-rank representation was found by removing and imputing 100 random values and choosing the number of PCs that yields the lowest mean absolute error (MAE). This iterative PCA imputation method has previously been shown to be among the best-performing and computationally efficient algorithms for missing value imputation of DNA methylation data.[36]

DNA methylation data in BASE-II were processed as previously described.[37] For this study, sex discrepancies were assessed by comparing the genetically determined sex with the reported sex.

2.3.5 | Polygenic score generation

The genotyping data from EMIF-AD MBD was pre-processed as described previously.[38] In short, this pre-processing pipeline includes filtering of strand-ambiguous SNPs, aligning alleles to the human genome assembly GRCh37/hg19, phasing, imputation based on the Haplotype Reference Consortium (HRC) reference panel, pre- and post-imputation quality control (QC), and filtering of SNPs with a minor allele frequency (MAF) < 0.01.

The genotyping data of the EXTEND cohort underwent a similar pre-processing pipeline including SNP filtering and alignment using the HRC/1000Genomes imputation preparation and checking pipeline (v4.2.7),[39] imputation using the Michigan Imputation Server with the 1000Genomes reference panel (phase 3 v5 hg19, Population: EUR, Phasing: Eagle, R-squared filter: 0.3),[40] and post-imputation SNP filtering (MAF < 0.01, Hardy-Weinberg equilibrium (HWE) $P < 10^{-4}$, missing call rate > 90%).

Subsequently, the LDAK tool (v5.2)[41] was applied to calculate the polygenic (risk) scores (PGSs) for 11 of the 14 dementia risk factors as well as AD status (including the APOE region) using the HapMap reference panel and the summary statistics of 12 genome-wide association studies (GWASs) listed in Table S5. In short, LDAK splits the summary statistics in pseudo training and test summary statistics and then uses a variational Bayes approach to estimate the regression coefficients of the SNPs.[41] For the hyperparameter optimization, multiple models are trained and evaluated on the test summary statistics for different combinations of prior distribution parameters. This methodology is available for six model types (i.e., bayesR, bayesR-shrink, lasso, lasso-sparse, ridge, and bolt regression models) that differ in the form of the prior distribution for the SNP effect sizes. As the bayesR is the recommended method for PGS generation by the developers of LDAK, all PGSs were generated using the bayesR approach.

2.4 | Model generation

The model generation consists of two approaches; the prediction of the CAIDE and LIBRA scores in the EXTEND cohort as well as the prediction of MCI and AD status in EMIF-AD MBD by 14 MPSs of dementia risk factors (Figure 1).

TABLE 2 The applied feature selection and machine learning methods.

Feature selection method	Machine learning method	CAIDE/LIBRA	Dementia risk factors
None	ElasticNet	x	x
Correlation-based[a]	Random Forest	x	x
Correlation-based[a]	ElasticNet	x	x
Literature-based[b]	Random Forest		x
Literature-based[b]	ElasticNet		x

Note: For the prediction of nine dementia risk factors and the CAIDE and LIBRA scores in the EXTEND cohort, different combinations of feature selection and machine learning methods were used.
[a]In the *correlation-based feature selection* method, the 10,000 CpGs that have the highest absolute Spearman correlation coefficient with the target variable (i.e., predicted dementia risk factor) in the training set were selected for model training (Text S1).
[b]In the *literature-based feature selection* method, the CpGs that reached genome-wide significance in a previously performed epigenome-wide association study (Table S7) for the corresponding risk factor were selected for model training.

2.4.1 | Approach 1: Generation of methylation risk scores for CAIDE and LIBRA

In the first approach, we aimed to construct an epigenetic model for the prediction of the LIBRA score (i.e., epi-LIBRA model) and CAIDE score (i.e., epi-CAIDE model). Before constructing these models, we evaluated multiple supervised and unsupervised feature selection methods to investigate which method is suitable for reducing the large dimensionality of the DNA methylation data and found superior performance of the "correlation-based feature selection method" (see Supplementary Information for a more detailed description of the applied methodology and results). In short, in the correlation-based feature selection method, the 10,000 CpGs that have the highest absolute Spearman correlation coefficient with the target variable in the training set were selected for model training.

Hence, for predicting both the CAIDE and LIBRA scores, we trained an ElasticNet and Random Forest model on the 10,000 features selected by correlation-based feature selection, as well as an ElasticNet model trained on all CpGs that passed QC (Table 2). Accordingly, we applied five-repeated five-fold cross-validation to find the optimal hyperparameter values that yield the minimal MAE (the searched hyperparameter space is shown in Table S6).

2.4.2 | Approach 2: Generation of MMRS

For the second approach, we first aimed at predicting 14 known dementia risk factors included in the CAIDE and/or LIBRA scoring systems (Table 1). Specifically, for the prediction of smoking, alcohol consumption, high-density lipoprotein (HDL) cholesterol, and body mass index (BMI), the corresponding DNA methylation-based models from Hillary and Marioni (2020)[42] were used. The epigenetic clock model was adopted from Zhang et al.[43] for the prediction of age and for each of the remaining risk factors (i.e., low education, physical inactivity, unhealthy diet, depression, type II diabetes, heart disease, sex, systolic blood pressure, and total cholesterol), five models were trained in the EXTEND cohort corresponding to different combinations of feature selection and machine learning algorithms (Table 2). These include an ElasticNet model without prior feature selection, an ElasticNet and Random Forest model with correlation-based feature selection, as well as an ElasticNet and random Forest model trained on the CpGs that reached genome-wide significance in previously performed epigenome-wide association studies (EWAS) (Table S6).

Accordingly, for each of these five models, five-repeated five-fold cross-validation was applied to find the optimal hyperparameter values that yield the maximal area under the receiver operating characteristic curve (AUROC) (for discrete risk factors) or minimal MAE (for continuous risk factors) (the searched hyperparameter space is shown in Table S7). From the five models per risk factor, the model that achieved the highest average AUROC or R^2 over all folds was considered the best-performing risk factor model.

Subsequently, the predicted risk scores of each risk factor model (for binary variables this is defined as log(1/1-p), where p is the estimated class probability), referred to as MPSs, were used as variables for the construction of a MMRS for the prediction of "MCI versus control" (i.e., MMRS-MCI model) and "AD versus control" (i.e., MMRS-AD model) in the training set of the EMIF-AD MBD study. For this, the Kennard-Stone algorithm[44] was first applied to the MPSs to split the data into a training (n = 436) and an independent test set (n = 187) (Table S8). Accordingly, an ElasticNet (EN), sparse partial least squares-discriminant analysis (sPLS-DA), and Random Forest model with recursive feature elimination (RF-RFE) models were trained by five-repeated five-fold cross-validation to find the optimal hyperparameter combination that yields the highest AUROC (the searched hyperparameter space is shown in Table S9).

2.5 | Model validation

Following the model performance estimation on the independent test set of the EMIF-AD MBD study, we applied our top-performing model to generate methylation-based risk scores in the ADNI, PPMI, and BASE-II cohorts.

2.5.1 | Validation in ADNI

In ADNI, a linear regression model with age and sex as covariates was fitted to assess the baseline association between the risk scores calculated by our best-performing model and the cognitive outcomes and CSF biomarkers.

Besides the cross-sectional validation, the predictive capability of risk scores generated by our best-performing model for longitudinal cognitive functioning was assessed using survival analysis. Specifically, based on the risk scores calculated by our best-performing model, the individuals of the ADNI cohort were divided into three equally sized risk categories; low- (n = 74), intermediate- (n = 75), and high-risk (n = 74). We accordingly assessed the statistical significance of the difference in time-dependent conversion to cognitive impairments by comparing the low- and high-risk groups using the log-rank test and a Cox regression model with age and sex as covariates (survival package (v3.5.5)[45]). Furthermore, a Kaplan-Meier curve was constructed to visualize the probability of cognitive impairments over time for each of the three risk categories. It should be noted that not all cognitive measures were available for each individual; hence, the number of samples per risk category was different per cognitive outcome (Table S2).

To assess whether the expected direction of effect (i.e., hazard ratio (HR) > 1 for the high-risk group) for all the nine cognitive outcomes is observed more frequently than would be expected by chance, a 10,000-permutation analysis was performed. For each permutation, the samples were randomly divided into three risk categories, and the HR from the Cox regression model was recorded.

2.5.2 | Validation in PPMI

The PPMI cohort was used to assess the longitudinal prediction of our best-performing model among the epi-LIBRA, epi-CAIDE, and MMRS models. As for the ADNI cohort, the individuals of the PPMI cohort were divided into three equally sized risk categories; low- (n = 43), intermediate- (n = 43), and high-risk (n = 43). We accordingly assessed the statistical significance of the difference in time-dependent conversion to MCI or PD dementia by comparing the low- and high-risk groups using the log-rank test and a Cox regression model with age and sex as covariates (survival package (v3.5.5)[45]).

2.5.3 | Validation in BASE-II

We incorporated our best-performing model into a publicly available R Shiny app (https://github.com/Dementia-Systems-Biology/DementiaRiskPrediction) and used this app to calculate the methylation model's risk scores for the pre-processed DNA methylation data of the BASE-II cohort. For the validation, we calculated linear regression models of the available cognitive measures on the risk scores. We adjusted for age, sex, and the first six genetic PCs. All cognitive measures were tested cross-sectionally and longitudinally. The longitudinal scores were calculated using the following formula: score = $(\text{test}_{\text{Follow-Up}} - \text{test}_{\text{Baseline}})$/time interval. The P-values were false discovery rate (FDR) -adjusted to account for multiple testing.

2.6 | Model interpretation

2.6.1 | Variable importance

The importance of variables contributing to the best-performing model was evaluated using mean absolute SHapley Additive exPlanations (SHAP) values of the test set samples in the EMIF-AD MBD study as calculated with the DALEX package (v2.4.3).[46] To make the SHAP values comparable between models, the values were normalized such that the absolute sum equals one. The scaled mean absolute SHAP values can therefore be interpreted as the average proportional contribution to the predicted score.

2.6.2 | GO overrepresentation analysis

The missMethyl package (v1.34.0)[47] was used to perform GO overrepresentation analysis on the union of the most important features of the methylation-based risk factor models (i.e., MPS models) that were used for the prediction by the best-performing model. For the Elastic-Net models, the most important features are the CpGs with a non-zero coefficient, while for the Random Forest models, the 1000 CpGs with the largest Gini index were considered as the most important variables.

2.6.3 | Influence of genetic variation

To evaluate the genetic contribution to variation explained by DNA methylation within the best-performing model, we quantified the joint variation between the genetic determinants of the model's CpGs and the CpG beta values using the Joint and Individual Variation Explained (JIVE) method.[48] We obtained the genetic determinants of the model's CpGs from the Genetics of DNA Methylation Consortium (goDMC) using the following parameters: clumped cis- and trans-mQTLs, P < 1e-5[8] and the joint variation quantified using the r.jive package (v2.4).

In addition, a Bayesian colocalization method (coloc package (v5.2.2)[49]) was performed to test for common genetic variants associated with DNA methylation and the genomic loci reported in the latest AD GWAS.[50] For this analysis, the latest GWAS summary statistics were downloaded from the European Bioinformatics Institute GWAS Catalog (https://www.ebi.ac.uk/gwas/) under accession no. GCST90027158. We defined AD-associated loci as the physical region containing correlated SNPs (r^2 > 0.6) within 250K bp up and downstream of the 88 index SNPs reported by Bellenguez *et al.*[51] Non-clumped cis-mQTLs with P < 1e-5 associated with the contributing MPSs to the best-performing MMRS model, were extracted from the GoDMC database.[8]

2.7 | Model extension

To compare the performance of models based on MPSs, PGSs, and CSF biomarkers individually or in combination, we re-trained the EMIF-AD MBD data using the 14 MPSs along with 12 PGSs (Table S5) and/or three CSF biomarkers (i.e., Aβ1-42, t-tau, and p-tau z-scores) as additional variables. For this, the same machine learning strategy was applied as described earlier.

3 | RESULTS

3.1 | Generation and validation of the epi-CAIDE and epi-LIBRA scores

In the EXTEND cohort, the Random Forest model demonstrated a relatively effective prediction of the CAIDE score (cross-validation R^2 = 0.47). However, the prediction of the LIBRA score proved challenging, with a maximal cross-validation R^2 of 0.04 (Table S10).

Within the EMIF-AD MBD study, a robust correlation was observed between the epi-CAIDE score and chronological (R^2 = 0.45) as well as epigenetic (R^2 = 0.55) age, underscoring the predominant influence of age on the epi-CAIDE score. Nevertheless, with an AUROC ≤ 0.61, both the epi-CAIDE and epi-LIBRA scores as predicted by the best-performing random Forest models were shown to be poor estimators of both MCI and AD status in the independent test set of the EMIF-AD MBD study (Figure 2 and Table S11).

3.2 | Generation and validation of MMRS models

Besides age (R^2 = 0.92) and sex (AUROC = 1), the best-predicted dementia risk factors by blood-derived DNA methylation data in the EXTEND cohorts included notable performance for smoking (AUROC = 0.91), type II diabetes (AUROC = 0.89), and heart disease status (AUROC ≈ 0.80) (Tables S12 and S13). Nine out of the 14 risk factors were also recorded in EMIF-AD MBD. Performance of the nine corresponding MPSs demonstrated weaker yet statistically significant predictive capability. Particularly, similar to the findings in the EXTEND cohort, age (R^2 = 0.87), sex (AUROC = 1), smoking (AUROC = 0.80), and heart disease status (AUROC = 0.67) were identified as the best-predicted risk factors by the methylation data of the EMIF-AD MBD study (Table S12).

As shown in Figure 2 and outlined in Table S11, our MMRS-AD models, incorporating the 14 MPSs as variables, did not show significant predictive capability for AD status in the independent EMIF-AD MBD test set (AUROC ≤ 0.60 However, our MMRS-MCI model (RF-RFE) demonstrated a notable ability to predict MCI status, achieving an AUROC of 0.68 ($P = 2.0 \times 10^{-3}$ for a linear regression model adjusted for age and sex). This performance significantly surpassed MCI prediction based solely on epigenetic age (DeLong's $P = 8.0 \times 10^{-3}$).

3.3 | Validation in independent dementia cohorts

3.3.1 | Validation in the PPMI cohort

To validate our MMRS-MCI (RF-RFE) model in the PPMI cohort, we performed a survival analysis. Specifically, we stratified the population into three equally sized risk categories based on the baseline risk score predicted by the MMRS-MCI model and monitored their conversion to cognitive impairments (i.e., MCI or PD dementia) over time. Notably, we observed a higher incidence of conversion to cognitive impairments within the high-risk group compared to the low-risk category (HR = 2.59, Cox regression P = 0.036). These findings suggest that the baseline risk score predicted by the MMRS-MCI model in this PD cohort is indicative of the likelihood of developing cognitive impairments, as illustrated in Figure 3A.

3.3.2 | Validation in the ADNI cohort

To validate our MMRS-MCI (RF-RFE) model in the ADNI cohort, our initial assessment focused on examining the association between the MMRSs and various cognitive outcomes at baseline. Notably, at baseline, the risk score of our MMRS-MCI (RF-RFE) model demonstrated associations with all cognitive outcomes and CSF biomarkers in the anticipated direction (Table S14). Following FDR adjustment, significant associations were observed for ADAS-Q4, TMT, tau, and p-tau (FDR-adjusted P = 0.03, 0.03, 0.02, and 0.02, respectively). These findings underscore the robustness of the MMRS-MCI model in capturing relevant cognitive and biomarker variations in the ADNI cohort.

To further validate our best-performing MMRS model in the ADNI cohort, we conducted a survival analysis to assess the conversion to cognitive impairments across three risk categories based on the baseline MMRS-MCI (RF-RFE) scores. Notably, within the high-risk group, we observed a greater incidence of conversion to cognitive impairments, as assessed by various cognitive tests. This reached nominal statistical significance for RAVLT – learning (HR = 2.27, P = 0.01) and RAVLT – percent forgetting (HR = 1.73, P = 0.045) (Figure 3B, Table S15, and Figure S1). While not all cognitive tests in the ADNI cohort reached statistical significance, the consistent direction of effect (HR > 1 for the high-risk group, adjusted for age and sex) was observed across all nine cognitive tests, surpassing what would be expected by chance (permutation P = 0.048). These results strengthen the evidence supporting the predictive capability of our MMRS-MCI model in forecasting cognitive impairment in the ADNI cohort.

3.3.3 | Validation in the BASE-II cohort

In the BASE-II cohort, our association analyses of MMRS-MCI (RF-RFE) with the seven cognitive test scores, both cross-sectional and longitudinal, revealed a statistically significant association specifically with the longitudinal Face-Profession Task score. This particular score

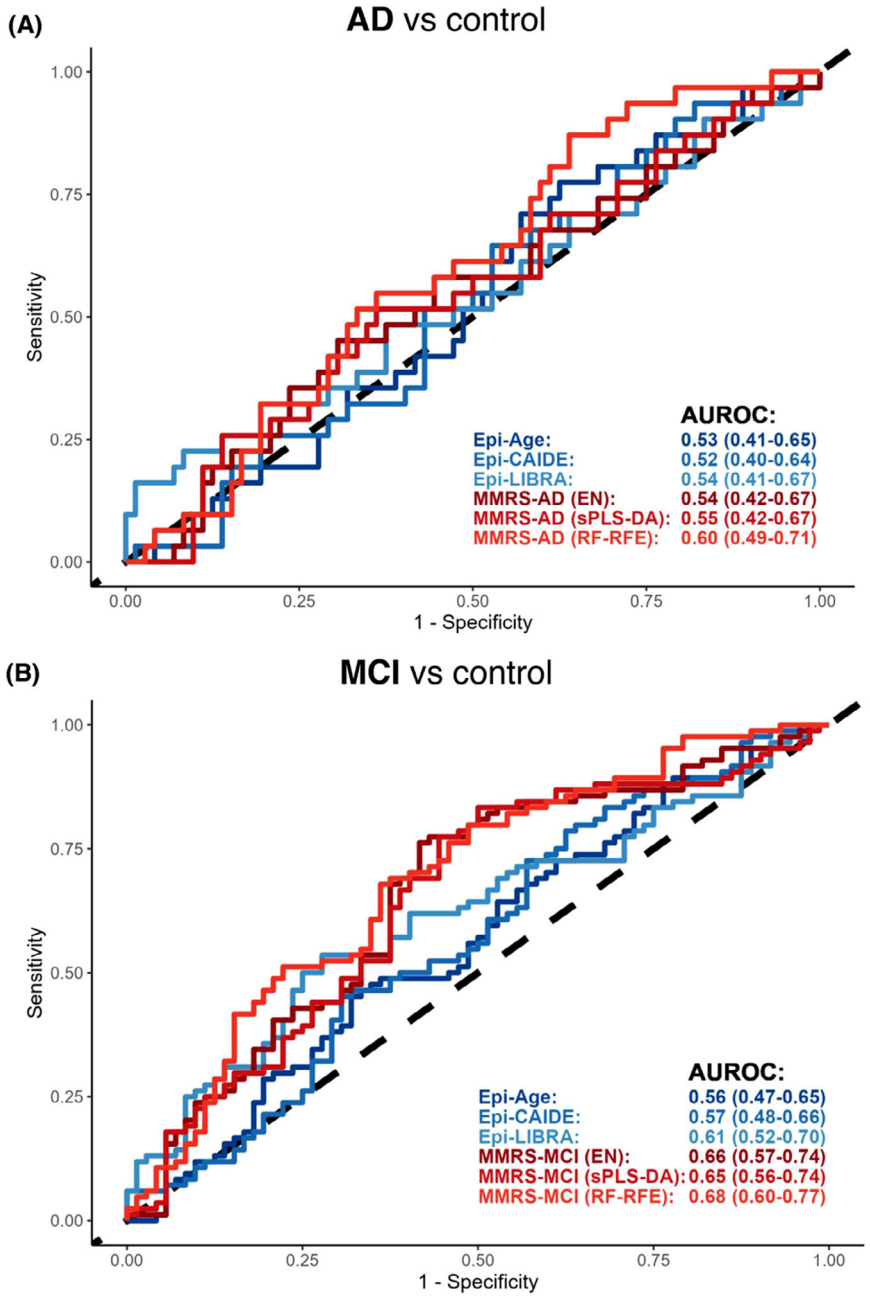

FIGURE 2 ROC curves of cross-sectional AD and MCI status prediction in the independent test set of the EMIF-AD MBD study. The MMRS models (red) are trained on the 14 MPSs for the prediction of AD **(A)** and MCI **(B)**. The epi-LIBRA and epi-CAIDE scores (blue) are both predicted by the Random Forest model with a correlation-based feature selection method (i.e., the best-performing model) from the EXTEND data. The 95% confidence intervals of the AUROC values are indicated between brackets. EN, ElasticNet; sPLS-DA, sparse partial least squares-discriminant analysis; RF-RFE, random Forest with recursive feature elimination.

serves as a measure of episodic (associative) memory (FDR-adjusted $P = 2.0e\text{-}3$). However, the MMRS score did not exhibit evidence for association with any of the other cognitive scores analyzed (Table S16). These findings highlight a notable and specific link between our MMRS-MCI model and episodic memory performance in the BASE-II cohort.

3.4 | Model interpretation

As shown in Figure 4, the best-performing MMRS-MCI (RF-RFE) model relies on 10 out of the 14 MPSs for its prediction. Moreover, the mean absolute SHAP values indicate that the MMRS is predominantly driven by depression, HDL cholesterol, physical inactivity, and low education MPSs (Figure 4). The distribution of the MPSs among the diagnostic groups is shown in Figure S2.

"AMPA glutamate receptor clustering (GO:0097113)" stands out as the most overrepresented GO term by the union of the CpGs used in these 10 risk factor models (7,571 CpGs, unadjusted $P = 4.2e\text{-}4$) (Table S17). Notably, the model's CpGs that are associated with this GO term reside within the untranslated region (UTR) or gene body of the APOE, DLG, NLGN1, SHANK3, SHISA6, SHISA7, SLC7A11, and SSH1 genes (Figure S3). However, the JIVE analysis indicated that only minimal joint information is captured by the model's CpGs and their associated genetic variants. Particularly, for PRS, less than 3% of the variance in DNA methylation data is explained by genetic variation (Figure S4). Finally, the colocalization analysis indicated that only 11 CpGs out of 7,571 CpGs, forming the 10 contributing MPSs to the best-performing

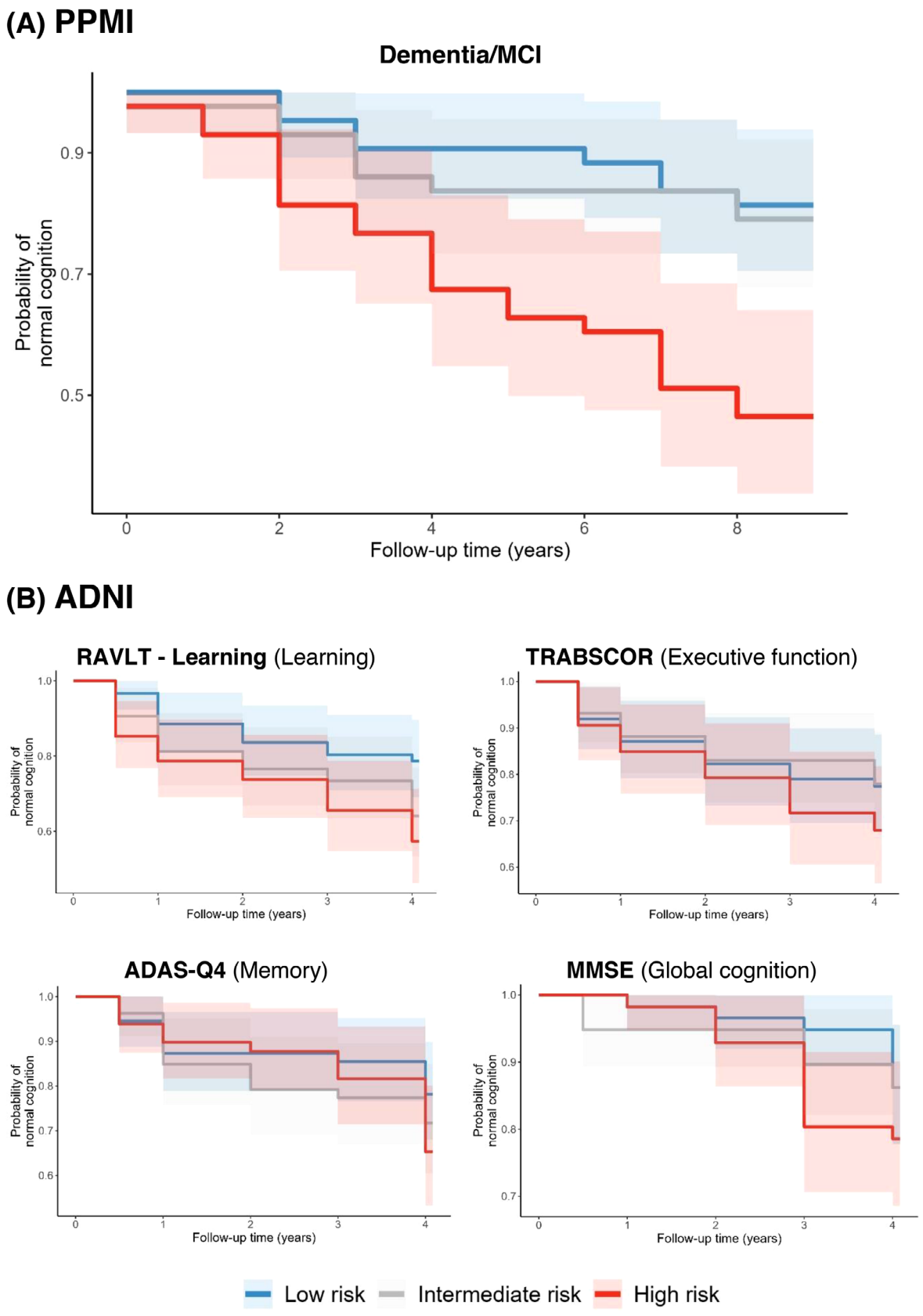

FIGURE 3 Kaplan-Meier curves of cognitive impairments in the ADNI and PPMI cohorts. The risk categories were defined based on the baseline score predicted by the MMRS-MCI (RF-RFE) model. The shaded area around the line indicates the 95% confidence interval. ADAS, Alzheimer's Disease Assessment Scale; RAVLT, Rey's Auditory Verbal Learning Test; TMT, Trail Making Test Part B Time; MMSE, Mini-Mental State Examination.

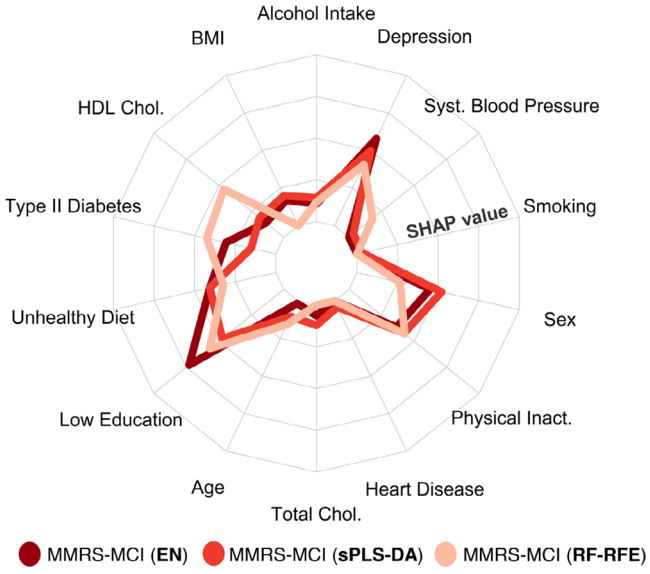

FIGURE 4 Radar chart of the scaled mean absolute SHAP values. The scaled mean absolute SHAP values indicate the variable importance of the 14 MPSs in the three MMRS-MCI models. The values for each of the models are scaled such that the sum equals one. EN, ElasticNet; sPLS-DA, sparse partial least squares-discriminant analysis; RF-RFE, random Forest with recursive feature elimination.

MMRS model, are likely to share a common genetic basis with AD risk loci, with a posterior probability larger than 0.99. (Table S18). Collectively, these results suggest that our best-performing model incorporates distinctive information from DNA methylation data, with a limited influence from genetic factors. This underscores the unique contribution of epigenetic information in the predictive capacity of the model, providing valuable insights into the distinct molecular features associated with AD risk.

3.5 | Model extension

Despite the significant associations observed between most PGSs and their corresponding dementia risk factors in the EXTEND and EMIF-AD MBD cohorts, their predictive performance remains poor, characterized by R^2 values of 0.1 or lower (Table S18). Furthermore, we found low correlations, as indicated by Pearson correlation coefficients consistently below 0.1, between most PGSs and their corresponding MPSs (Figure S5).

The addition of the 14 MPSs as additional variables alongside the initial 12 polygenic PGSs resulted in a significant increase in the AUROC for predicting MCI in the EMIF-AD MBD test set. Specifically, the AUROC increased from 0.64 (when only PGSs were utilized as variables) to 0.69, showcasing the added predictive value brought about by the inclusion of MPSs into the model. Similarly, incorporating MPSs into the CSF biomarkers further enhanced the MCI predictive performance, raising the AUROC from 0.76 to 0.88 (Figure 5).

4 | DISCUSSION

In this study, we leveraged whole-blood-derived DNA methylation data obtained from a midlife general population to construct molecular scores serving as a surrogate for modifiable and non-modifiable risk factors for dementia. We used the LIBRA and CAIDE total scores, along with the 14 individual risk factors contributing to these scores, as outcomes for training of methylation-based risk score models (i.e., epi-LIBRA, epi-CAIDE, and 14 MPS models). We further trained multivariate models in the EMIF-AD MBD training set using the MPSs as variables for the cross-sectional prediction of AD and MCI status (i.e., MMRS-AD and MMRS-MCI, respectively) (Figure 1). Our findings indicated that despite the poor predictive performance of the epi-CAIDE and epi-LIBRA models for AD and MCI statuses within the EMIF-AD MBD cohort, the MMRS model, leveraging individual MPSs, demonstrated a notable predictive capacity for MCI status, achieving an AUROC score of 0.68 (Figure 2). This predictive ability was further validated through its application in the prospective and/or cross-sectional prediction of cognitive impairments at a population level, as evidenced by its effectiveness across three independent cohorts: PPMI, ADNI, and BASE-II (Figure 3). Moreover, our analysis revealed that incorporating the 14 MPSs as supplementary variables enhanced the precision of cross-sectional MCI predictions, when compared to traditional genetic and/or CSF biomarkers in the EMIF-AD MBD study (Figure 5).

The poor performance of the epi-CAIDE and epi-LIBRA models for MCI and AD prediction might be partly attributed to the fact that the CAIDE and LIBRA scores do not consider the weights of risk factors estimated by DNA methylation data, whereas an ideal epigenetic model would give more weight to the factors best predicted by the DNA methylation data. Furthermore, it is possible for individuals to share the same CAIDE or LIBRA score while exhibiting different contributing risk factors. For instance, one person may have a high CAIDE score due to a combination of high BMI and low physical activity, while another individual may attain the same CAIDE score solely due to advanced age. The discrepancies in the contributions to the total scores might not be adequately captured by our molecular-based epi-CAIDE and epi-LIBRA models, which could potentially account for their poor performance. Lastly, the reported poor performance of the CAIDE and LIBRA scores in individual-level dementia risk prediction may be partially reflected in our findings.[10,11]

The MMRS models generated from 14 methylation-based dementia risk factor models, could not be used to significantly predict AD status in the independent test set of the EMIF-AD MBD study, though it demonstrated better performance in predicting MCI status (Figure 2). This might be attributed to the prevalence of cardiovascular-related risk factors in our analysis such as BMI, systolic blood pressure, heart disease, type II diabetes, physical activity, diet, smoking, as well as HDL and total cholesterol. These factors might exhibit stronger associations with non-AD types of dementia, like vascular dementia, as compared to AD.[52] This was further confirmed when the MMRS model for MCI showed promising predictive capabilities for cognitive impairment, both cross-sectionally and prospectively in newly diagnosed PD (Figure 3).

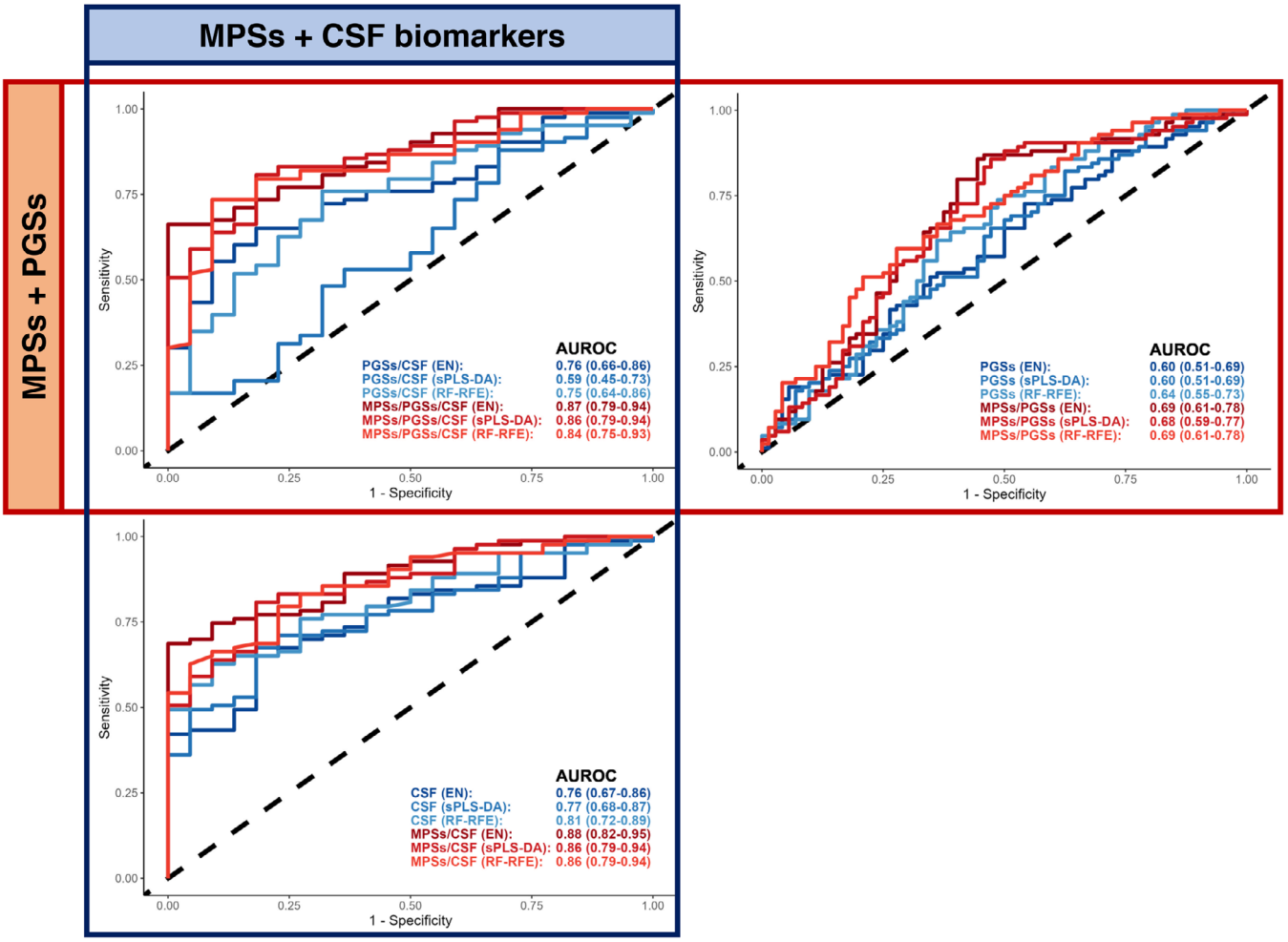

FIGURE 5 ROC curves of MCI prediction. MCI status was predicted in the independent test set of the EMIF-AD MBD study using the PGSs (top right), CSF biomarkers (bottom left), and both (top left) with and without the MPSs as additional variables. EN, ElasticNet; sPLS-DA, sparse partial least squares-discriminant analysis; RF-RFE, random Forest with recursive feature elimination.

The enhanced prediction of MCI status by adding MPSs to the CSF and genetic risk factors for AD dementia in comparison to using only CSF or genetic biomarkers as variables (Figure 5) indicates that the MPSs provide unique information beyond the already established genetic and CSF biomarkers. This was further confirmed with low correlation between corresponding PGSs and MPSs (Figure S6) and the outcomes of the JIVE analysis and colocalization analyses, which collectively underscored the limited impact of genetic variation on the predictive capacity of the model (Figure S4 and Table S18).

The genes associated with the CpGs contributing to our best-performing MMRS model were significantly overrepresented in the "AMPA glutamate receptor clustering (GO:0097113)" GO term. Interestingly, AMPA glutamate receptors are known to play an important role in synaptic transmission. It is worth mentioning that previous studies have demonstrated an association between AD pathology and increased removal of these receptors from the post-synaptic membrane.[53] Although the methylation changes in these processes were measured in the blood, they may resemble the alterations that occur in AD-affected brain.[54] Specifically, the CpGs in our model that are associated with this GO term are located within known dementia-related genes. Notably these genes include APOE,[55] DLG1,[56] NLGN1,[57] SHANK3,[58] SHISA6,[59] SHISA7,[60] and SSH1.[61]

One notable strength of the present study is reducing the dimensionality of the DNA methylation data into 14 interpretable latent features (i.e., MPSs). This approach enabled the construction of a robust and replicable model and overcoming the lack of replication of CpG-level models in assessing the risk of dementia and cognitive impairment as described previously.[9] Importantly, the MPSs have a clear biological meaning, which contrasts with the more difficult-to-interpret latent features generated by other commonly used dimensionality reduction methods such as (s)PLS, PCA, and autoencoders. This way, our model can provide direct information about which risk factors contribute to an (elevated) dementia risk.

In assessing the findings and methodologies presented in this study, it is imperative to consider certain limitations that may influence the interpretation and potential clinical applications of the results in the future. First, an AUROC of 0.68 for our MMRS model indicates limited accuracy for individual-level predictions. This is comparable

to the AUROC of 0.65 reported by Decker et al. for the predictive performance of the midlife LIBRA score.[62] However, in contrast to questionnaire-based risk scores (e.g., CAIDE and LIBRA), the MMRS model allows for a more objective assessment of dementia risk across multiple independent cohorts using only blood samples. Second, although MCI is a well-established dementia risk factor, it is not a perfect predictor of the future onset of dementia. Specifically, a significant proportion of MCI individuals revert to a cognitively healthy status.[63] Therefore, modeling the prospective cognitive outcome (e.g., the trajectory of cognitive decline) may result in a better model for the prediction of the future development of cognitive impairments instead of using the cross-sectional MCI status as the dependent variable. The lack of (baseline) DNA methylation data in large-scale prospective studies, however, makes this approach currently not feasible, and future initiatives should aim to collect such data, allowing for more sophisticated analyses. Last, it should be noted that in the current study, all the models have only been trained and validated in a predominantly Caucasian population and, hence, the reported performance might be different for other ethnicities.

In conclusion, our established MMRS model demonstrates utility in the population-based prediction of cognitive impairment and dementia. This model serves as a foundation for future studies with the potential to enhance predictive performance through the exploration of novel feature selection and machine learning methods, integration of additional omics layers, and training on larger (prospective) datasets. Such endeavors could significantly contribute to improving the accuracy of existing blood-based models for the early identification of individuals at risk of developing dementia, a crucial step for the implementation of effective intervention strategies. As DNA methylation profiles have previously been shown to be modifiable through lifestyle changes, the information provided by our model possibly allows for targeted intervention strategies, aimed at maximally reducing the patient-specific risk scores. Future studies should investigate how, and to what extent, these MPSs can be best modified by, for example, lifestyle interventions.

AFFILIATIONS

[1]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNs), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, Maastricht, The Netherlands

[2]Department of Biochemistry, Cardiovascular Research Institute Maastricht (CARIM), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, Maastricht, The Netherlands

[3]Department of Advanced Computing Sciences (DACS), Faculty of Science and Engineering (FSE), Maastricht University, Maastricht, The Netherlands

[4]Medical School, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK

[5]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany

[6]Department of Bioinformatics - BiGCaT, Research Institute of Nutrition and Translational Research in Metabolism (NUTRIM), Faculty of Health, Medicine and Life Sciences (FHML), Maastricht University, Maastricht, The Netherlands

[7]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Dr. Manuel Nagel, Würzburg, Germany

[8]Department of Endocrinology and Metabolic Diseases (including Division of Lipid Metabolism), Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[9]BCRT - Berlin Institute of Health Center for Regenerative Therapies, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany

[10]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

[11]Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck, Germany

[12]Danish Dementia Research Centre, Rigshospitalet, Copenhagen, Denmark

[13]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

[14]Laboratory for Cognitive Neurology, KU Leuven, Leuven, Belgium

[15]Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

[16]School of Medical Sciences, Örebro University, Örebro, Sweden

[17]Department of Geriatrics, Södertälje Hospital, Södertälje, Sweden

[18]Department of Geriatric Psychiatry, Central Institute of Mental Health; Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany

[19]Department of Neurology, Alzheimer Center Amsterdam, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

[20]Neurochemistry Laboratory, Department of Laboratory Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

[21]Memory Center, Geneva University and University Hospitals; on behalf of the AMYPAD Consortium, Genève, Switzerland

[22]Aix-Marseille University-CNRS, Marseille, France

[23]Neuroscience Therapeutic Area, GlaxoSmithKline R&D, Stevenage, Hertfordshire, UK

[24]Université de Lille, Lille Cedex, France

[25]Department of Biomedical Sciences, University of Antwerp, Antwerpen, Belgium

[26]Neuroprotection & Neuromodulation (NEUR) Research Group, Center for Neurosciences (C4N), Vrije Universiteit Brussel (VUB), Jette, Brussels, Belgium

[27]Center for Research and Advanced Therapies, CITA-Alzheimer Foundation, Gipuzkoa, Spain

[28]Neurology Department, Centro de Investigación en Red en enfermedades neurodegenerativas (CIBERNED), Hospital Sant Pau, Sant Antoni Maria Claret, Barcelona, Spain

[29]University Hospital of Psychiatry Zürich, University of Zürich, Zürich, Switzerland

[30]Department of Psychiatry, University Hospital of Lausanne (CHUV), Lausanne, Switzerland

[31]1st Department of Neurology, School of Medicine, 'Laboratory of Neurodegenerative Diseases, Center for Interdisciplinary Research and Innovation, Aristotle University of Thessaloniki, and Alzheimer Hellas, Macedonia, Balkan Center, Thessaloniki, Greece

[32]Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, The Sahlgrenska Academy at University of Gothenburg, Göteborg, Sweden

[33]Clinical Neurochemistry Lab, Sahlgrenska University Hospital, Göteborg, Sweden

[34]Paris Brain Institute, ICM, Pitié-Salpêtrière Hospital, Sorbonne University, Paris, France

[35]Neurodegenerative Disorder Research Center, Division of Life Sciences and Medicine, and Department of Neurology, Institute on Aging and Brain Disorders,

University of Science and Technology of China and First Affiliated Hospital of USTC, Hefei, P.R. China

[36]Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, UK

[37]UK Dementia Research Institute at UCL, Maple House, London, UK

[38]Hong Kong Center for Neurodegenerative Diseases, Clear Water Bay, Shatin, N.T., Hong Kong, China

[39]Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, USA

[40]AC Immune SA, formerly Janssen R&D, LLC. Beerse, Belgium at the time of study conduct, Lausanne, Switzerland

[41]University of Oxford, Oxford, United Kingdom; Currently at Johnson & Johnson Innovative Medicines, Beerse, Belgium

[42]Ageing Epidemiology Research Unit, School of Public Health, Imperial College, South Kensington Campus, London, UK

ACKNOWLEDGMENTS

E.P. is supported for this work by a ZonMw Memorabel/Alzheimer Nederland Grant (733050516). The EMIF-AD MBD study is part of the EMIF-AD project. The EMIF-AD project has received support from the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement n° 115372, the resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007-2013) and EFPIA companies'' in kind contribution. For this study JM was supported by the National Institute for Health and Care Research (NIHR) Exeter Biomedical research Centre (BRC). The view expressed are those of the author(s) and not necessarily those of the NIHR of the Department of Health and Social Care. CML was supported by the Heisenberg grant of the German Research Foundation (DFG; LI 2654/4–1). C.M.L. acknowledges funding by the ''EU Joint Programme – Neurodegenerative Disease Research 2021″(JPND2021, EPIC4ND project) and the Cure Alzheimer's Fund (CAF, EPIC4AD project). HZ is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2022-01018 and #2019-02397), the European Union's Horizon Europe research and innovation programme under grant agreement No 101053962, Swedish State Support for Clinical Research (#ALFGBG-71320), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C, and #ADSF-21-831377-C), the Bluefield Project, the Olav Thon Foundation, the Familjen Erling-Perssons Stiftelse, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2022-0270), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), the European Union Joint Programme – Neurodegenerative Disease Research (JPND2021-00694), the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre, and the UK Dementia Research Institute at UCL (UKDRI-1003). KB is supported by the Swedish Research Council (#2017-00915 #2022-00732), the Swedish Alzheimer Foundation (#AF-930351, #AF-939721 and #AF-968270), Hjärnfonden, Sweden (#FO2017-0243 and #ALZ2022-0006), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-715986 and #ALFGBG-965240), the European Union Joint Program for Neurodegenerative Disorders (JPND2019-466-236), the Alzheimer's Association 2021 Zenith Award (ZEN-21-848495), and the Alzheimer's Association 2022-2025 Grant (SG-23-1038904 QC). KL was supported for the work in this project by a major project grant from the Alzheimer's Society UK (AS-PG-14-038), a Medical Research Council (MRC) grant (MR/S011625/1) and a National Institute of Aging (NIA) of the National Institutes of Health (NIH) grant (R01AG067015). We thank all participants and teams who contributed data to PPMI. PPMI, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners including 4D Pharma, AbbVie, AcureX Therapeutics, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's (ASAP), Avid Radiopharmaceuticals, Bial Biotech, Biogen, BioLegend, Bristol Myers Squibb, Calico Life Sciences LLC, Celgene Corporation, DaCapo Brainscience, Denali Therapeutics, The Edmond J. Safra Foundation, Eli Lilly and Company, GE Healthcare, GlaxoSmithKline, Golub Capital, Handl Therapeutics, Insitro, Janssen Pharmaceuticals, Lundbeck, Merck & Co., Meso Scale Diagnostics LLC, Neurocrine Biosciences, Pfizer, Piramal Imaging, Prevail Therapeutics, F. Hoffmann-La Roche and its affiliated company Genentech, Sanofi Genzyme, Servier, Takeda Pharmaceutical Company, Teva Neuroscience, UCB, Vanqua Bio, Verily Life Sciences, Voyager Therapeutics and Yumanity Therapeutics. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST STATEMENT

H.Z. has served at scientific advisory boards and/or as a consultant for Abbvie, Acumen, Alector, Alzinova, ALZPath, Annexon,

Apellis, Artery Therapeutics, AZTherapies, Cognito Therapeutics, CogRx, Denali, Eisai, Merry Life, Nervgen, Novo Nordisk, Optoceutics, Passage Bio, Pinteon Therapeutics, Prothena, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Alzecure, Biogen, Cellectricon, Fujirebio, Lilly, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program (outside submitted work). K.B. has served as a consultant and at advisory boards for Acumen, ALZPath, BioArctic, Biogen, Eisai, Lilly, Moleac Pte. Ltd, Novartis, Ono Pharma, Prothena, Roche Diagnostics, and Siemens Healthineers; has served at data monitoring committees for Julius Clinical and Novartis; has given lectures, produced educational materials and participated in educational programs for AC Immune, Biogen, Celdara Medical, Eisai and Roche Diagnostics; and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program, outside the work presented in this paper. The other authors declare no conflicts of interests. Author disclosures are available in the supporting information.

CONSENT STATEMENT

All participants provided informed consent in their respective cohort studies.

ORCID

Ehsan Pishva https://orcid.org/0000-0002-8964-0682

REFERENCES

1. Bouwman FH, Frisoni GB, Johnson SC, et al. Clinical application of CSF biomarkers for Alzheimer's disease: from rationale to ratios. *Alzheimers Dement (Amst)*. 2022;14:e12314.
2. Hansson O, Edelmayer RM, Boxer AL, et al. The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. *Alzheimers Dement*. 2022;18:2669-2686.
3. van den Hove DLA, Riemens RJM, Koulousakis P, Pishva E. Epigenome-wide association studies in Alzheimer's disease; achievements and challenges. *Brain Pathol*. 2020;30:978-983.
4. Dhar GA, Saha S, Mitra P. Nag Chaudhuri R. DNA methylation and regulation of gene expression: guardian of our health. *Nucleus (Calcutta)*. 2021;64:259-270.
5. Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9:436-447.
6. Samblas M, Milagro FI, Martínez A. DNA methylation markers in obesity, metabolic syndrome, and weight loss. *Epigenetics*. 2019;14:421-444.
7. Gonzalez-Jaramillo V, Portilla-Fernandez E, Glisic M, et al. The role of DNA methylation and histone modifications in blood pressure: a systematic review. *J Hum Hypertens*. 2019;33:703-715.
8. Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet*. 2021;53:1311-1321.
9. Schäfer Hackenhaar F, Josefsson M, Nordin Adolfsson A, et al. Sixteen-year longitudinal evaluation of blood-based DNA methylation biomarkers for early prediction of Alzheimer's disease. *J Alzheimers Dis*. 2023;94:1443-1464.
10. Schiepers OJ, Köhler S, Deckers K, et al. Lifestyle for Brain Health (LIBRA): a new model for dementia prevention. *Int J Geriatr Psychiatry*. 2018;33:167-175.
11. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol*. 2006;5:735-741.
12. Kim JH, Chang IB, Kim YH, Min CY, Yoo DM, Choi HG. Association between various types or statuses of smoking and subjective cognitive decline based on a community health survey of Korean adults. *Front Neurol*. 2022;13:810830.
13. Schaefer SM, Kaiser A, Behrendt I, Eichner G, Fasshauer M. Association of alcohol types, coffee, and tea intake with risk of dementia: prospective cohort study of UK biobank participants. *Brain Sci*. 2022;12:360.
14. Malik R, Georgakis MK, Neitzel J, et al. Midlife vascular risk factors and risk of incident dementia: longitudinal cohort and Mendelian randomization analyses in the UK Biobank. *Alzheimers Dement*. 2021;17:1422-1431.
15. Iso-Markku P, Waller K, Vuoksimaa E, et al. Midlife physical activity and cognition later in life: a prospective twin study. *J Alzheimers Dis*. 2016;54:1303-1317.
16. Xu W, Tan L, Wang HF, et al. Education and risk of dementia: dose-response meta-analysis of prospective cohort studies. *Mol Neurobiol*. 2016;53:3113-3123.
17. Shannon OM, Ranson JM, Gregory S, et al. Mediterranean diet adherence is associated with lower dementia risk, independent of genetic predisposition: findings from the UK Biobank prospective cohort study. *BMC Med*. 2023;21:81.
18. Hattersley A, The Exeter 10,000 (EXTEND) project. In: Facility NECR, editor.2020.
19. Bos I, Vos S, Vandenberghe R, et al. The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. *Alzheimers Res Ther* . 2018;10:1-9.
20. Marek K, Chowdhury S, Siderowf A, et al. The Parkinson's progression markers initiative (PPMI)–establishing a PD biomarker cohort. *Ann Clin Transl Neurol* . 2018;5:1460-1477.
21. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*. 2010;74:201-209.
22. Bertram L, Böckenhoff A, Demuth I, et al. Cohort profile: the Berlin Aging Study II (BASE-II). *Int J Epidemiol*. 2014;43:703-712.
23. Petersen RC. Mild cognitive impairment as a diagnostic entity. *J Intern Med*. 2004;256:183-194.
24. Winblad B, Palmer K, Kivipelto M, et al. Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med*. 2004;256:240-246.
25. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34:939.
26. Edmonds EC, Delano-Wood L, Jak AJ, et al. "Missed" mild cognitive impairment: high false-negative error rate based on conventional diagnostic criteria. *J Alzheimers Dis*. 2016;52:685-691.
27. Litvan I, Goldman JG, Tröster AI, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: movement disorder society task force guidelines. *Mov Disord*. 2012;27:349-356.
28. Harvey J, Reijnders RA, Cavill R, et al. Machine learning-based prediction of cognitive outcomes in de novo Parkinson's disease. *npj Parkinson's Disease*. 2022;8:150.
29. Demuth I, Banszerus V, Drewelies J, et al. Cohort profile: follow-up of a Berlin Aging Study II (BASE-II) subsample as part of the GendAge study. *BMJ Open*. 2021;11:e045576.
30. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009;65:403-413.

31. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30:1363-1369.
32. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *Bmc Genom [Electronic Resource]*. 2013;14:1-10.
33. Liu J, Siegmund KD. An evaluation of processing methods for HumanMethylation450 BeadChip data. *Bmc Genom [Electronic Resource]*. 2016;17:1-11.
34. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data*. 2016;9:22-24.
35. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. *J Stat Softw*. 2016;70:1-31.
36. Lena PD, Sala C, Prodi A, Nardini C. Methylation data imputation performances under different representations and missingness patterns. *BMC Bioinf*. 2020;21:1-22.
37. Sommerer Y, Dobricic V, Schilling M, et al. Epigenome-wide association study in peripheral tissues highlights DNA methylation profiles associated with episodic memory performance in humans. *Biomedicines*. 2022;10:2798.
38. Hong S, Prokopenko D, Dobricic V, et al. Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD multimodal biomarker discovery dataset. *Transl Psychiatry*. 2020;10:403.
39. McCarthy. HRC or 1000G imputation preparation and checking. 2018.
40. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284-1287.
41. Zhang Q, Privé F, Vilhjálmsson B, Speed D. Improved genetic prediction of complex traits from individual-level data or summary statistics. *Nat Commun*. 2021;12:4192.
42. Hillary RF, Marioni RE. MethylDetectR: a software for methylation-based health profiling. *Wellcome Open Res*. 2020;5:283.
43. Zhang Q, Vallerga CL, Walker RM, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med*. 2019;11:1-11.
44. Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics*. 1969;11:137-148.
45. Therneau TM, Lumley T. Package 'survival'. *R Top Doc*. 2015;128:28-33.
46. Biecek P. DALEX: explainers for complex predictive models in R. *J Mach Learn Res* . 2018;19:3245-3249.
47. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2015;32:286-288.
48. O'Connell MJ, Lock EF. R. JIVE for exploration of multi-source molecular data. *Bioinformatics*. 2016;32:2877-2879.
49. Rasooly D, Peloso GM, Giambartolomei C. Bayesian genetic colocalization test of two traits using coloc. *Curr Protoc*. 2022;2:e627.
50. Bellenguez C, Kucukali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412-436.
51. Marioni RE, Harris SE, Zhang Q, et al. GWAS on family history of Alzheimer's disease. *Transl Psychiatry*. 2018;8:99.
52. Dong C, Zhou C, Fu C, et al. Sex differences in the association between cardiovascular diseases and dementia subtypes: a prospective analysis of 464,616 UK Biobank participants. *Biol Sex Differ*. 2022;13:21.
53. Babaei P. NMDA and AMPA receptors dysregulation in Alzheimer's disease. *Eur J Pharmacol*. 2021;908:174310.
54. Wei X, Zhang L, Zeng Y. DNA methylation in Alzheimer's disease: in brain and peripheral blood. *Mech Ageing Dev*. 2020;191:111319.
55. Martens YA, Zhao N, Liu CC, et al. ApoE Cascade Hypothesis in the pathogenesis of Alzheimer's disease and related dementias. *Neuron*. 2022;110:1304-1317.
56. Taskesen E, Mishra A, van der Sluis S, et al. Susceptible genes and disease mechanisms identified in frontotemporal dementia and frontotemporal dementia with Amyotrophic Lateral Sclerosis by DNA-methylation and GWAS. *Sci Rep*. 2017;7:8899.
57. Arias-Aragón F, Tristán-Clavijo E, Martínez-Gallego I, et al. A Neuroligin-1 mutation associated with Alzheimer's disease produces memory and age-dependent impairments in hippocampal plasticity. *iScience*. 2023;26:106868.
58. Landry O, François A, Oye Mintsa Mi-Mba MF, et al. Postsynaptic protein shank3a deficiency synergizes with Alzheimer's disease neuropathology to impair cognitive performance in the 3xTg-AD murine model. *J Neurosci*. 2023;43:4941-4954.
59. Ramos J, Caywood LJ, Prough MB, et al. Genetic variants in the SHISA6 gene are associated with delayed cognitive impairment in two family datasets. *Alzheimers Dement*. 2023;19:611-620.
60. Sabaie H, Talebi M, Gharesouarn J, et al. Identification and analysis of BCAS4/hsa-miR-185-5p/SHISA7 competing endogenous RNA axis in late-onset alzheimer's disease using bioinformatic and experimental approaches. *Front Aging Neurosci*. 2022;14:812169.
61. Cazzaro S, Woo JA, Wang X, et al. Slingshot homolog-1-mediated Nrf2 sequestration tips the balance from neuroprotection to neurodegeneration in Alzheimer's disease. *Proc Natl Acad Sci USA*. 2023;120:e2217128120.
62. Deckers K, Barbera M, Köhler S, et al. Long-term dementia risk prediction by the LIBRA score: a 30-year follow-up of the CAIDE study. *Int J Geriatr Psychiatry*. 2020;35:195-203.
63. Jongsiriyanyong S, Limpawattana P. Mild cognitive impairment in clinical practice: a review article. *Am J Alzheimers Dis Other Demen*. 2018;33:500-507.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Koetsier J, Cavill R, Reijnders R, et al. Blood-based multivariate methylation risk score for cognitive impairment and dementia. *Alzheimer's Dement*. 2024;20:6682–6698. https://doi.org/10.1002/alz.14061

APPENDIX 1: COLLABORATORS

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf