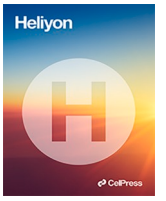
Heliyon 10 (2024) e36728

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Prediction of future dementia among patients with mild cognitive impairment (MCI) by integrating multimodal clinical data

Andrew Cirincione [a], Kirsten Lynch [b], Jamie Bennett [a], Jeiran Choupan [b,c], Bino Varghese [d], Nasim Sheikh-Bahaei [d], Gaurav Pandey [a,*], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] *Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY, 10029, USA*
[b] *Laboratory of Neuro Imaging (LONI), USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, 1975 Zonal Ave, Los Angeles, CA, 90033, USA*
[c] *NeuroScope Inc., Scarsdale, NY, 10583, USA*
[d] *Department of Radiology, Keck School of Medicine, University of Southern California, 1975 Zonal Ave, Los Angeles, CA, 90033, USA*

A B S T R A C T

Efficiently and objectively analyzing the complex, diverse multimodal data collected from patients at risk for dementia can be difficult in the clinical setting, contributing to high rates of underdiagnosis or misdiagnosis of this serious disorder. Patients with mild cognitive impairment (MCI) are especially at risk of developing dementia in the future. This study evaluated the ability of multi-modal machine learning (ML) methods, especially the Ensemble Integration (EI) framework, to predict future dementia development among patients with MCI. EI is a machine learning framework designed to leverage complementarity and consensus in multimodal data, which may not be adequately captured by methods used by prior dementia-related prediction studies. We tested EI's ability to predict future dementia development among MCI patients using multimodal clinical and imaging data, such as neuroanatomical measurements from structural magnetic resonance imaging (MRI) and positron emission tomography (PET) scans, from The Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) challenge. For predicting future dementia development among MCI patients, on a held out test set, the EI-based model performed better (AUC = 0.81, F-measure = 0.68) than the more commonly used XGBoost (AUC = 0.68, F-measure = 0.57) and deep learning (AUC = 0.79, F-measure = 0.61) approaches. This EI-based model also suggested MRI-derived volumes of regions in the middle temporal gyrus, posterior cingulate gyrus and inferior lateral ventricle brain regions to be predictive of progression to dementia.

1. Introduction

Dementia affects millions of people worldwide, incurring a substantial burden on patients, families, healthcare professionals, and the economy [1]. In 2022, there were 55 million people worldwide living with dementia, a figure that is expected to increase to 78 million by 2030 [2]. The annual global cost of dementia in 2022 was $1.3 trillion, and is projected to increase to $2.8 trillion by 2030 [2].

Physiologically, dementia has adverse impacts on cognition, quality of life, and the ability to live independently [3]. This condition

* Corresponding author.
E-mail address: gaurav.pandey@mssm.edu (G. Pandey).

[1] 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: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.heliyon.2024.e36728
Received 6 November 2023; Received in revised form 20 August 2024; Accepted 21 August 2024
Available online 22 August 2024

is generally preceded by mild cognitive impairment (MCI), a stage of moderate memory and cognition loss beyond what is expected for a patient's age [4]. In the United States, around 15 % of people aged 60 or older live with MCI [5], and an estimated 12–20 % of these individuals advance to dementia each year [6]. For this reason, early detection of dementia for MCI patients is crucial, since these patients can be treated to slow cognitive decline [4,7]. However, if these patients progress to dementia, treatment options are limited [8]. Therefore, a major emphasis in dementia research and treatment is to effectively and efficiently identify if a patient with MCI will develop dementia in the future [5,9].

Currently the detection of dementia utilizes multimodal data like cognitive tests, MRI and PET scans, and other biomarkers [3]. However, despite the rich information in these data, it is estimated that over 18 % of dementia patients in the US are treated for the wrong related illness [10], and almost two-thirds of dementia cases go undetected globally [11]. This can be partly attributed to the difficulty in objectively and efficiently analyzing these complex, diverse data, even for experts. To address these issues and help standardize the study of dementia and other neurodegenerative disorders, several research efforts have collected and organized these multimodal data from a large number of at-risk patients. Prominent examples of such efforts are the Alzheimer's Disease Neuroimaging Initiative (ADNI) [12], the National Alzheimer's Coordinating Center (NACC) database [13] and the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging [14]. Since manual analysis of these large multimodal datasets is challenging, computational algorithms, especially from machine learning (ML), have been applied to these data for the study of dementia and other neurodegenerative disorders, often with encouraging results [15,16].

Prior dementia-related prediction studies based on these multimodal data, especially from ADNI, have used various data integration and ML methods [16,17]. For instance, Beltrán et al. applied Gradient Boosting and other traditional ML methods to a concatenated set of features from multimodal ADNI data to predict Alzheimer's disease (AD) progression [18]. Venugopalan et al. used a set of stacked autoencoders and a convolutional neural network to combine features derived from similar multimodal data for the early detection of AD [19]. Although these representative studies reported encouraging results, their strategy of deriving and concatenating/combining features into a single set before model training may obfuscate local signals exclusive to the individual data modalities and hinder model performance [20–22]. This represents a major challenge for the automated prediction of future dementia for MCI patients from clinically standard multimodal data.

An alternative, potentially more effective strategy is to derive specialized local predictive models from the individual modalities before aggregating them into a global predictor. This approach leverages both the consensus and complementarity among the modalities to enhance prediction performance [23]. Our recent Ensemble integration (EI) framework systematically implements this approach to develop effective predictive models from heterogeneous multimodal data [20]. The effectiveness of EI arises from its ability to aggregate an unrestricted number and types of local models derived from each of the data modalities, thus incorporating their inherent consensus and complementarity. In addition, unlike most prior data integration methods, EI also enables the interpretation of the final aggregated model to identify the most predictive features over all data modalities.

In the current work, we evaluated EI's potential for the early prediction of dementia for MCI patients. In particular, we systematically developed and evaluated a predictive model for the development of future dementia among MCI patients using multimodal data derived from the ADNI cohort. In addition to developing this model, we also leveraged EI's interpretation capabilities to identify neuroanatomical features that may be associated with the progression of MCI to dementia.

2. Materials and methods

Below, we describe the data and methods used in our study. All the code used to process these data and implement the methods described is available at https://github.com/GauravPandeyLab/TADPOLE_EI.

2.1. Data description and problem definition

The multimodal data used in our study were from the TADPOLE Challenge [24], which were derived in turn from the ADNI 1, 2, and GO studies [12,25,26]. These data modalities included cognitive test scores, demographic information and neuroanatomical measurements from structural magnetic resonance imaging (MRI) and positron emission tomography (PET) scans. For specific details of the data preparation and preprocessing steps taken in this Challenge, please refer to https://tadpole.grand-challenge.org/.

Several of the TADPOLE data modalities consisted of features that were missing for at least 30 % of the MCI patients (Supplementary Fig. 1), which are difficult to impute reliably [27], and are likely to adversely affect downstream analyses [28]. These modalities were excluded from our analyses. Among the remaining modalities, the MRI measurements derived from fine-grained segmented brain regions (MRI ROI) (processed using FreeSurfer [29] software) dominated the full set of features (313/341), which may artificially dominate predictive modeling. To address this issue, we split the TADPOLE MRI ROI modality into five sub-modalities based on the semantics of the constituent features. These sub-modalities were volume (white matter parcellation), volume (cortical parcellation), cortical thickness standard deviation, cortical thickness average and surface area, each of which was then treated as its own separate modality.

Missing values of features in the resultant modalities were imputed using K-Nearest Neighbor imputation (KNNImpute) [30] with $K = 5$ within each modality. Categorical features were treated as continuous ones using their original values in TADPOLE to reduce dimensionality, with the exception of APOE4, which was one-hot encoded [31] due to its close association with dementia risk [32,33]. The resultant continuous features were linearly normalized between 0 and 1. The resultant data modalities considered in our study are detailed in Table 1, and the full set of features are provided in the first sheet of **Supplementary File.**

Since our focus was on identifying which patients currently diagnosed with MCI may progress to dementia, our study cohort

derived from TADPOLE consisted of 841 patients diagnosed with (early (E) or late (L)) MCI at their first (baseline) visit (Fig. 1). Accordingly, our target prediction problem was *to predict which of these MCI patients' diagnosis changed to dementia, and who stayed as MCI, based only on data collected at baseline when they were diagnosed with MCI.* A randomly sampled 80 % subset of this cohort was used to train all the predictive models for the target problem. These models were then evaluated on the test set, which consisted of the remaining 20 % of the cohort.

2.2. Identification of individual features associated with future dementia among MCI patients

As an initial examination of the relevance of our processed TADPOLE data to the target problem, we assessed which of the individual features in the various modalities were associated with the future development of dementia among MCI patients. For this, the Wilcoxon Rank-Sum test was used for continuous features and chi-squared test for categorical ones. The resultant p-values from these tests were adjusted for multiple hypothesis testing using the Benjamini-Hochberg method [34] to yield the corresponding false discovery rates (FDRs).

2.3. Ensemble integration (EI)

EI is a machine learning framework to build predictive models from multimodal data [20]. The effectiveness of EI arises from its integration of local predictors trained on the individual data modalities into heterogeneous ensemble models. This allows EI to leverage both the consensus and complementarity among the modalities more effectively than other data integration techniques. In the current work, we applied EI to the baseline patient data from multiple TADPOLE modalities described above to build an effective predictor of future dementia for MCI patients (Fig. 2). Specifically, each set of local predictors derived from the individual modalities output probabilities of whether or not a patient with MCI will develop dementia in the future. EI models for this outcome were then built on these probabilities using heterogeneous ensemble algorithms like mean aggregation, stacking and ensemble selection (ES). Stacking learns a meta-predictor over the local models using any applicable classification algorithm (e.g., SVM and random forest), while ES uses an iterative strategy to select a subset of the base predictors into the final ensemble. The EI model that performed the best was then used to identify the most predictive features using the framework's interpretation algorithm. To address the imbalance between patients who progressed to dementia and those who remained at MCI, undersampling [35] was used for training all EI models. Further technical details of the version of EI used in this work may be found in the original paper [20], Supplementary Tables 1 and 2 and the GitHub repository for this paper mentioned above.

2.4. Benchmark prediction approaches

To assess if EI produced more accurate predictions of future dementia for MCI patients than other methods, we considered two benchmark prediction algorithms that were adaptations of methods used for similar problems and applied to ADNI data [18,19].

The first benchmark algorithm was XGBoost, an implementation of a gradient boosted decision tree algorithm [36], which is considered to be one of the best methods for making predictions from tabular data [37]. We adopted this benchmark since it had been used in a prior representative study to predict AD progression from multimodal ADNI data [18]. The second approach was inspired by the Stacked Denoising Autoencoders (SDAE) method [19], where a series of autoencoders are each trained on the output of the previous in an unsupervised manner, after which a final classification layer is added to produce a prediction. This deep learning-based algorithm was adopted as a benchmark since it had been used for automated AD staging in a representative previous study [19]. In this work, we concatenated all the TADPOLE modalities considered for use with EI to make the resultant data compatible with both of these benchmark methods. The XGBoost algorithm was run using the xgboost Python library [36], and the SDAE algorithm was

Table 1
Modalities in TADPOLE data used in this study. For each modality, a brief description and the number of features included is provided.

Data modality	Description	# Features
Main cognitive tests	Scores from neuropsychological tests, such as ADAS 13, CDR-SB, and MMSE, administered by a clinical expert.	9
MRI volumes	Volumes of seven large/general brain regions, such as the ventricles, medial temporal lobe, and the cranial cavity.	7
MRI ROI (regions of interest): volume (white matter parcellation)	Volumes from white matter parcellation of specific brain regions, such as left and right sections of the thalamus, from MRI scans.	40
MRI ROI: volume (cortical parcellation)	Volumes from cortical parcellation of specific brain regions from MRI scans.	69
MRI ROI: cortical thickness standard deviation	Standard deviations of cortical thicknesses of specific brain regions from MRI scans.	68
MRI ROI: cortical thickness average	Averages of cortical thicknesses of specific brain regions from MRI scans.	68
MRI ROI: surface area	Surface areas of specific brain regions from MRI scans.	68
CSF biomarkers	Measures of amyloid, tau and ptau proteins from cerebral spinal fluid (CSF).	3
Demographics, APOE4 allele and others	Demographic variables, such age, ethnicity and marital status, as well as other relevant measurements, such as the value of the APOE4 allele and the baseline diagnosis of early (E) or late (L) MCI.	9

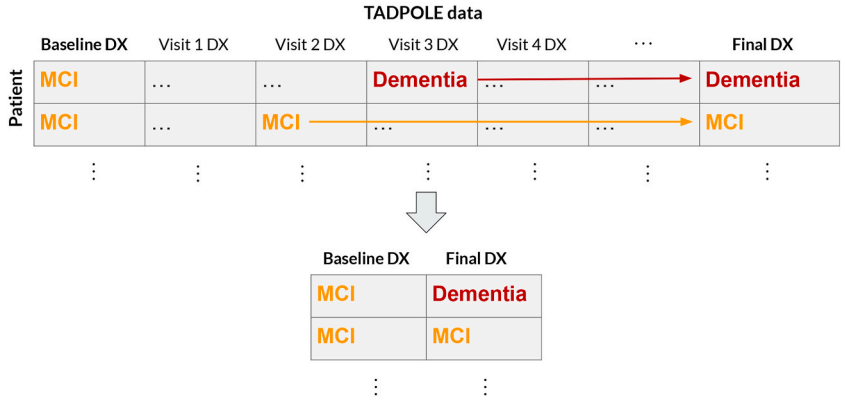


Fig. 1. Details of the patient cohort from TADPOLE analyzed in this study. Patients with an initial (baseline) diagnosis (DX) of MCI were identified, along with the multimodal data recorded at this visit (see Table 1). The most recent diagnosis given to each patient (Final DX) was the outcome predicted (stable MCI or progressed to dementia).

reimplemented in Python based on the description in the original study [19]. Then, we trained models using these algorithms to predict future dementia in MCI patients.

Finally, to assess if, and how well, EI was performing its core task of integrating information from multiple data modalities, we also compared EI's performance to those of heterogeneous ensembles derived from the individual modalities using exactly the same underlying algorithms as EI.

2.5. Evaluation methodology

All heterogeneous ensembles in EI applied to all and the individual data modalities were trained and evaluated on the training set consisting of baseline data of 672 patients using a five-fold nested cross-validation (CV) setup [38] (Supplementary Fig. 2). The XGBoost and autoencoder benchmark methods were trained and evaluated on the same training set in a standard five-fold CV setup. The heterogeneous ensemble algorithm in EI that performed the best in this CV-based evaluation was then used to develop the final EI model for evaluation on the test sets and interpretation. For comparison, final XGBoost and autoencoder models were also trained on the whole training set and evaluated on the test set.

In our TADPOLE MCI cohort, only a minority of patients progressed to dementia, which made this outcome more challenging to predict than the other class (patient staying at MCI). Thus, to assess the effectiveness of all the methods tested for predicting the outcome of future dementia, we used F_{max} [39,40], the maximum value of F-measure for all calculated precision and recall [41] scores for predicting the minority outcome, as the representative evaluation measure. To further assess the quality of these predicted outcomes, we also evaluated the methods using area under the ROC curve (AUC) [42] score, which summarizes the true positive and false positive rates of predictions over all classification thresholds.

3. Results

3.1. Cohort characteristics

Our cohort derived from TADPOLE consisted of 841 patients diagnosed with MCI at their first visit (baseline). The diagnosis of 288 of these patients changed to dementia at a later time point, while 553's diagnosis stayed as MCI. The distribution of basic clinical/demographic characteristics of this cohort are provided in Table 2.

3.2. Individual features associated with future dementia among MCI patients

Univariate differential analysis revealed several individual features already known to be important for the detection of dementia (full results in the third sheet of the **Supplementary File**). As expected based on clinical practice, several cognitive tests, such as ADAS 13 (Alzheimer's disease assessment scale), FAQ (functional activities questionnaire) and ADAS 11, were the most significantly associated [43,44]. Several measurements of the hippocampus were also highly significant, which was also expected given the association between hippocampal atrophy and dementia [45–47].

3.3. Performance of EI and benchmark methods for early detection of dementia

Fig. 3 shows the cross-validation-derived performance on the training set of EI applied to all modalities, EI applied to each individual modality and both benchmark methods (XGBoost and autoencoder). Specifically, this figure shows the performance of the

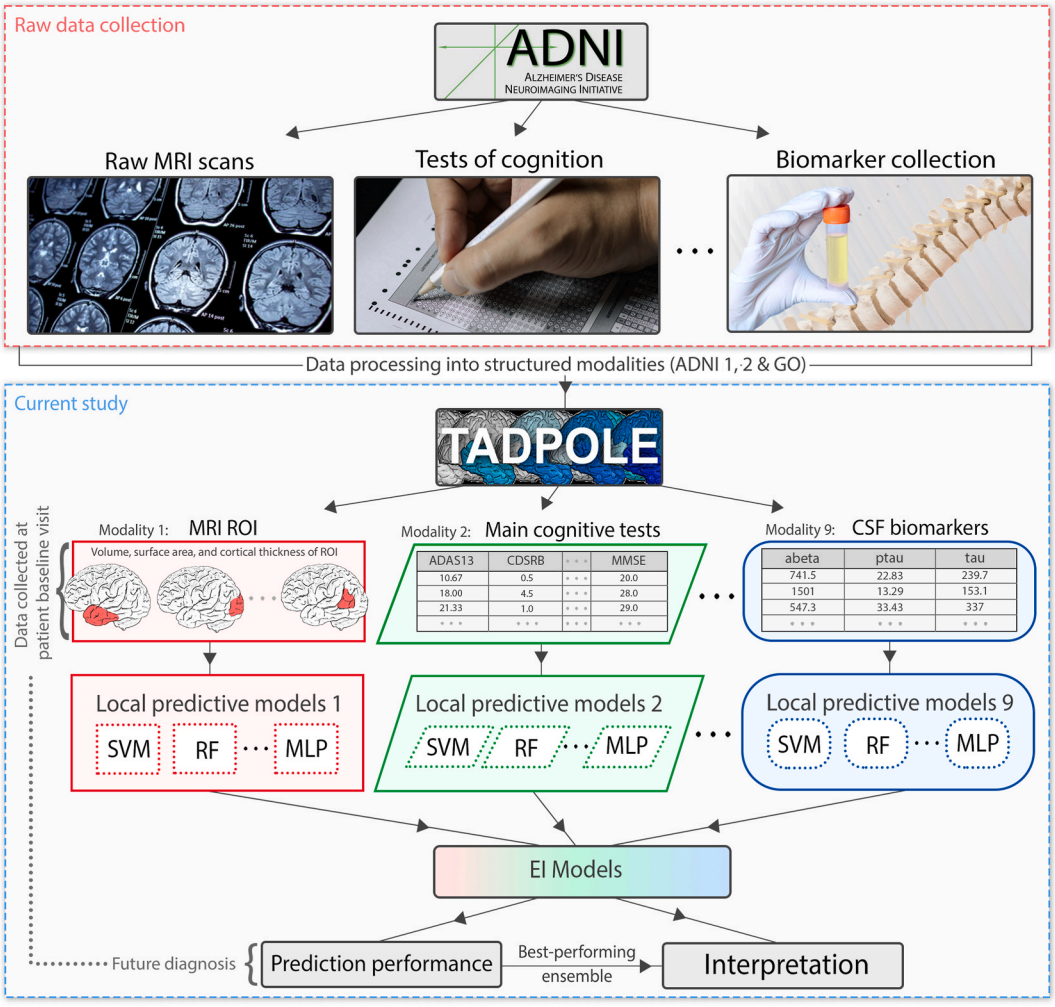


Fig. 2. Overview of the EI workflow for this study. Patient data from ADNI 1, 2 and GO were processed by TADPOLE Challenge organizers into the structured data modalities used in this study. Within EI, sets of local predictive models were trained on each modality separately using algorithms like SVM, random forest (RF) and multi-layer perceptron (MLP; traditional neural network). EI models were then trained on the outputs of all the local models using heterogeneous ensemble algorithms like mean aggregation, stacking and ensemble selection to predict future dementia for MCI patients. The most predictive features were identified by interpreting the best-performing EI model. Note that while MRI regions of interest (ROI) is shown as a single modality in this figure for simplicity and clarity, this study divided it into five modalities based on the brain region measurements included to balance the number of features across all resultant modalities (Table 1). Subimage credits: ADNI (logo), TADPOLE (logo), Nur Ceren Demir for iStock (Raw MRI scans), Nguyen Dang Hoang for Unsplash (Tests of cognition) and Derks24 for Pixabay (Biomarker collection).

various implementations of EI developed using the stacking, mean and ensemble selection algorithms used in the second stage of the framework (Fig. 2). EI was applied to the individual modalities, as well as all of them taken together (All modalities). The resultant performance was assessed in terms of both the (Fig. 3A) F_{max} for the positive (MCI progressing to dementia) class, our primary evaluation measure, as well as the commonly used (Fig. 3B) AUC measure. The benchmark algorithms, namely autoencoder and XGBoost, were also applied to the individual and all modalities, and their performance assessed in terms of the same measures.

In terms of median F_{max} measure, EI implementations applied to all the modalities in Table 1 outperformed those tested on the individual modalities. Furthermore, four of the 11 EI implementations utilizing all modalities performed better than both benchmark methods. In terms of AUC as well, EI implementations applied to all the modalities in Table 1 outperformed those tested on the individual modalities, and four of the 11 all-modality EI implementations performed better than both benchmark methods. These results showed that EI indeed integrated and leveraged the information in all the modalities to yield a promising method to predict future dementia among MCI patients. In particular, since EI based on stacking with random forest produced the highest cross-validated F_{max} and AUC, this implementation was used to develop the final EI predictive model from the whole training set. Fig. 4 shows the performance of this model, as well as those trained using the XGBoost and autoencoder benchmark methods, on the test set in terms of the F-measure (Fig. 4A) and AUC (Fig. 4B) measures. As on the training set, the EI model maintained relatively high performance on the test set (F-measure = 0.68 and AUC = 0.81), and better performance than both benchmark methods (F-measure = 0.61 & 0.57 and

Table 2
Distribution of clinical/demographic features from the demographics, APOE4 allele and others modality across the whole TADPOLE MCI cohort. Distributions of these features in the training and test cohorts are in the second sheet of the Supplementary Excel File.

Features	Overall (whole cohort)	Stayed at MCI (whole cohort)	Progressed to dementia (whole cohort)
	***n** (%) or **median** (interquartile range)*		
All patients	**841**	**555**	**286**
Age at baseline	**73.5** (68.0, 78.8)	**72.7** (66.6, 78.25)	**74.3** (69.5, 79.1)
Years of education	**16.0** (14.0, 18.0)	**16.0** (14.0, 18.0)	**16.0** (14.0, 18.0)
Diagnosis at baseline visit			
LMCI	**543** (64.6 %)	**291** (52.4 %)	**252** (88.1 %)
EMCI	**298** (35.4 %)	**264** (47.6 %)	**34** (11.9 %)
Sex			
Male	**499** (59.3 %)	**326** (58.7 %)	**173** (60.5 %)
Female	**342** (40.7 %)	**229** (41.3 %)	**113** (39.5 %)
Ethnicity			
Hispanic/Latino	**28** (3.3 %)	**19** (3.4 %)	**9** (3.1 %)
Not Hispanic/Latino	**809** (96.2 %)	**533** (96.0 %)	**276** (96.5 %)
Unknown	**6** (0.7 %)	**6** (1.1 %)	**0** (0.0 %)
Race			
White	**786** (93.5 %)	**511** (92.1 %)	**275** (96.2 %)
Asian	**14** (1.7 %)	**9** (1.6 %)	**5** (1.7 %)
Black	**26** (3.1 %)	**20** (3.6 %)	**6** (2.1 %)
More than one	**8** (1.0 %)	**8** (1.4 %)	**0** (0.0 %)
Hawaiian or other Pacific Islander	**2** (0.2 %)	**2** (0.4 %)	**0** (0.0 %)
Marital status			
Married	**650** (77.3 %)	**414** (74.6 %)	**236** (82.5 %)
Never married	**19** (2.3 %)	**13** (2.3 %)	**6** (2.1 %)
Divorced	**77** (9.2 %)	**64** (11.5 %)	**13** (4.5 %)
Widowed	**89** (10.6 %)	**58** (10.5 %)	**31** (10.8 %)
APOE4 allele frequency			
0	**414** (49.2 %)	**317** (57.1 %)	**97** (33.9 %)
1	**334** (39.7 %)	**193** (34.8 %)	**141** (49.3 %)
2	**93** (11.1 %)	**45** (8.1 %)	**48** (16.8 %)

AUC = 0.79 & 0.68 for Autoencoder and XGBoost respectively).

3.4. Most predictive features in final EI model

The EI framework enables the interpretation of its models, specifically in the form of a ranked list of features from all the input modalities in terms of their contributions [20]. In the case of the final EI model to predict future dementia among MCI patients, several features found to be the highest contributors were not found to be significantly associated in the univariate analyses. Table 3 provides a list of the ten such top-ranked features in our processed TADPOLE data modalities, while the **Supplementary File** provides the EI interpretation ranking for all the features, as well as the subset not individually associated with the outcome.

4. Discussion

This work investigated the potential of the Ensemble Integration (EI) framework [20], when applied to multimodal clinical data from the ADNI-derived TADPOLE [24] cohort, to predict if a patient with MCI may develop dementia in the future, a pressing healthcare need [4–6]. The results showed EI's ability to effectively integrate data from a diverse set of clinical and imaging modalities to predict future dementia among MCI patients. EI also performed better than other approaches that have been used for dementia detection approaches on a test set consisting of MCI patients whose data were not used during the model training process.

EI showed improved performance at early dementia detection over the benchmark methods previously used in the area, namely XGBoost and SDAE. The main reason behind this performance difference was the ability of EI to capture the signals specific to the TADPOLE individual modalities into local models. The integration of these models allowed EI to leverage both the complementarity and consensus across the modalities, a validation of the core hypothesis underlying the framework. In contrast, XGBoost and SDAE were applied to the original TADPOLE modalities concatenated together, which did not allow them to benefit from the same critical factors as effectively.

In addition, EI's interpretation capability [20] revealed several predictive features that were not found to be individually associated with this outcome by traditional statistical methods. Among these were MRI-derived volumes of the left middle temporal gyrus, right posterior cingulate and left inferior lateral ventricle brain regions (Table 3). Changes in the morphologies of these brain regions have been observed in cases of Alzheimer's disease (AD), the most commonly observed form of dementia [48]. In a study of conversion from MCI to AD at various time points using MRI derived features, Wei et al. found the volume of the left middle temporal gyrus to be a consistent predictor of AD conversion at each future time point [49]. Specifically, Guo et al. studied multiple cortical features of MRI and found that the volumes of both the middle temporal gyrus and the right posterior cingulate gyrus were significantly associated

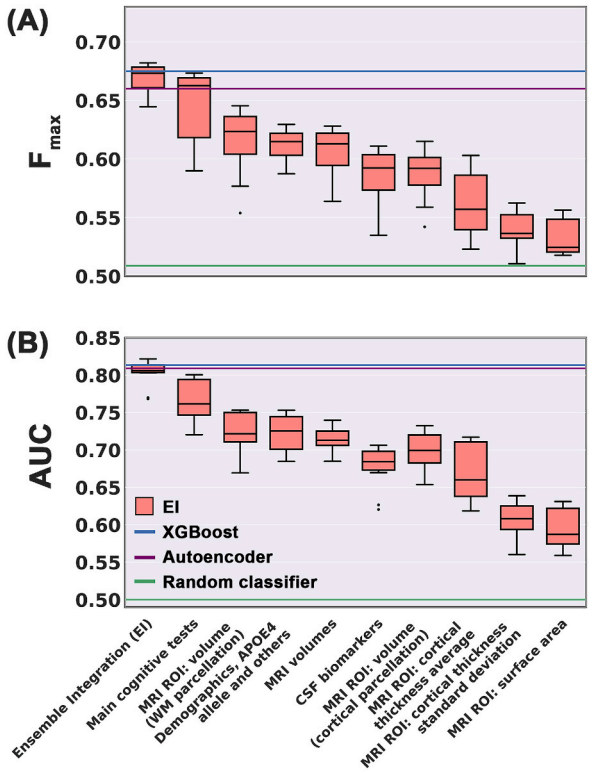


Fig. 3. Cross-validation-derived performance of all prediction methods on the training set evaluated in terms of the (A) F_{max} and (B) AUC evaluation measures. Note that the values for EI on all modalities and the individual ones are distributions, as the scores from each heterogeneous ensemble algorithm used within the framework are displayed. In contrast, the XGBoost and autoencoder benchmark methods are single algorithms represented by horizontal lines. The performance of a random classifier is shown for reference.

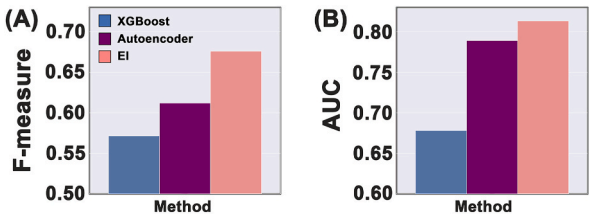


Fig. 4. Performance of the training set-derived EI, XGBoost and autoencoder models on the test set in terms of the (A) F-measure and (B) AUC scores. The EI model was trained using the stacking algorithm implemented with random forest. The F-measure (Y-axis of (A)) was evaluated for the predictions generated by the models for the class of MCI patients that progressed to dementia using the corresponding classification thresholds from the F_{max} calculated for the same class during cross-validation (Fig. 3).

with progression from MCI to AD [50]. In a longitudinal analysis of cognitive decline in MCI and AD patients, Zhao et al. found a significant difference in the volume of the left inferior lateral ventricle between MCI to AD converters and stable MCI patients at both baseline and future measurements [51]. Some studies have also reported the association of intracranial volume (ICV) to the likelihood of AD or dementia, although the extent of this association is unclear [52–54]. Such prior evidence suggests EI's ability to recapture known knowledge about progression to dementia.Furthermore, the other predictive features identified only by EI could serve as novel hypotheses about dementia progression that would need to be (in)validated in future studies.

Finally, although this study yielded a predictor of, and potentially useful information about the future development of dementia among MCI patients, it also had some limitations. During our processing of the TADPOLE data, any feature with missing values for over 30 % of baseline MCI patients was eliminated. This resulted in several imaging modalities, namely FDG PET, AV45 PET, AV1451 PET and DTI, being removed, since the missingness of all their constituent features was over the above threshold. This could cause a loss of information and may have hindered the performance of EI and the benchmark prediction methods. Furthermore, the TADPOLE Challenge data were aggregated from multiple editions of ADNI (1, 2 and GO), which may have been processed in distinct ways, such as using different versions of the evolving Freesurfer software. This may have made different subsets of the TADPOLE Challenge data

Table 3
Ten features found to be the most predictive by the interpretation of the final EI model, but not significantly individually associated with future dementia among MCI patients through our differential analysis. The full feature ranking from the EI interpretation and the subset of these features not found to be individually associated with the outcome can be found in the fourth and fifth sheets of the Supplementary File respectively.

Feature	Data modality
Volume (Cortical Parcellation) of LeftMiddleTemporal	MRI ROI: Volume (Cortical Parcellation)
Surface Area of RightPostcentral	MRI ROI: Surface Area
Intracranial Volume (Cortical Parcellation)	MRI ROI: Volume (Cortical Parcellation)
Volume (Cortical Parcellation) of RightPosteriorCingulate	MRI ROI: Volume (Cortical Parcellation)
Volume (WM Parcellation) of RightInferiorLateralVentricle	MRI ROI: Volume (WM Parcellation)
Surface Area of RightBankssts	MRI ROI: Surface Area
Surface Area of LeftBankssts	MRI ROI: Surface Area
Surface Area of RightEntorhinal	MRI ROI: Surface Area
Volume (WM Parcellation) of LeftInferiorLateralVentricle	MRI ROI: Volume (WM Parcellation)
Surface Area of LeftInferiorParietal	MRI ROI: Surface Area

slightly inconsistent with each other, which may have affected our prediction and interpretation results. In addition, since we were concerned with detecting dementia in a MCI patient as early as possible, only baseline patient data from TADPOLE were considered, and did not utilize the available longitudinal data. Examination of other missing data handling/imputational strategies and the use of longitudinal data, although complex tasks, may further improve prediction performance and allow for additional insights into dementia progression.

5. Conclusion

The application of the Ensemble Integration (EI) framework to multimodal clinical and imaging data derived from ADNI enabled this study to develop a promising model to detect dementia early in MCI patients. Early detection of MCI patients may help healthcare providers to plan early interventions and treatments for these patients more efficiently and objectively than current clinical practice. The algorithmic interpretation of this model also revealed several clinical features to be particularly predictive of this outcome, which may shed light on disease pathophysiology. Further enhancements to EI and/or the data used may enable further progress on this pressing healthcare problem. These enhancements to EI include extending the framework to accommodate the longitudinal data that are usually collected for MCI patients, enabling the modeling of their dementia status over time and potentially improving performance. Similarly, although discrete features derived from MRI scans in the TADPOLE data were useful for our prediction task, the effectiveness of this data modality can be further enhanced by utilizing the full scans and processing them through deep learning methods within EI.

Data and code availability

The original TADPOLE data used in this study can be obtained from https://tadpole.grand-challenge.org/Data/. The code for preprocessing these data for this study, as well as downstream analyses, e.g., EI, is available from https://github.com/GauravPandeyLab/TADPOLE_EI.

CRediT authorship contribution statement

Andrew Cirincione: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Kirsten Lynch:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Jamie Bennett:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Jeiran Choupan:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Bino Varghese:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Nasim Sheikh-Bahaei:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Gaurav Pandey:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gaurav Pandey reports financial support was provided by National Institutes of Health. Andrew Cirincione reports financial support was provided by National Institutes of Health. Kirsten Lynch reports financial support was provided by National Institutes of Health. Jamie Bennett reports financial support was provided by National Institutes of Health. Jeiran Choupan reports financial support was provided by National Institutes of Health. Bino Varghese reports financial support was provided by National Institutes of

Health. Nasim Sheikh-Bahaei reports financial support was provided by National Institutes of Health. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. All the authors report that financial support was provided by the National Institutes of Health.

Acknowledgments

This work was supported by National Institutes of Health grants R01HG11407, R01AG070825 and R01NS128486. It was also supported in part by Oracle Cloud credits and related resources provided by the Oracle for Research program. We also thank Yan Chak Li and Catherine McDonough for their technical help.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e36728.

References

[1] Y.-H. Hsiao, C.-H. Chang, P.-W. Gean, Impact of social relationships on Alzheimer's memory impairment: mechanistic studies, J. Biomed. Sci. 25 (Jan. 2018) 3, https://doi.org/10.1186/s12929-018-0404-x.
[2] J.-H. Shin, Dementia epidemiology fact sheet 2022, Ann. Rehabil. Med. 46 (2) (Apr. 2022) 53–59, https://doi.org/10.5535/arm.22027.
[3] Z. Arvanitakis, R.C. Shah, D.A. Bennett, Diagnosis and management of dementia: a review, JAMA 322 (16) (Oct. 2019) 1589–1599, https://doi.org/10.1001/jama.2019.4782.
[4] Y.-X. Chen, N. Liang, X.-L. Li, S.-H. Yang, Y.-P. Wang, N.-N. Shi, Diagnosis and treatment for mild cognitive impairment: a systematic review of clinical practice guidelines and consensus statements, Front. Neurol. 12 (Oct. 2021) 719849, https://doi.org/10.3389/fneur.2021.719849.
[5] R.C. Petersen, et al., Practice guideline update summary: mild cognitive impairment, Neurology 90 (3) (Jan. 2018) 126–135, https://doi.org/10.1212/WNL.0000000000004826.
[6] B.L. Plassman, et al., Prevalence of cognitive impairment without dementia in the United States, Ann. Intern. Med. 148 (6) (Mar. 2008) 427–434, https://doi.org/10.7326/0003-4819-148-6-200803180-00005.
[7] D.E. Bredesen, Reversal of cognitive decline: a novel therapeutic program, Aging 6 (9) (Sep. 2014) 707–717.
[8] K.S. Shaji, P.T. Sivakumar, G.P. Rao, N. Paul, Clinical practice guidelines for management of dementia, Indian J. Psychiatr. 60 (Suppl 3) (Feb. 2018) S312–S328, https://doi.org/10.4103/0019-5545.224472.
[9] O.L. Lopez, Mild cognitive impairment, Contin. Lifelong Learn. Neurol. 19 (2 Dementia) (Apr. 2013) 411–424, https://doi.org/10.1212/01.CON.0000429175.29601.97.
[10] J.E. Gaugler, H. Ascher-Svanum, D.L. Roth, T. Fafowora, A. Siderowf, T.G. Beach, Characteristics of patients misdiagnosed with Alzheimer's disease and their medication use: an analysis of the NACC-UDS database, BMC Geriatr. 13 (1) (Dec. 2013) 137, https://doi.org/10.1186/1471-2318-13-137.
[11] L. Lang, et al., Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis, BMJ Open 7 (2) (Feb. 2017) e011146, https://doi.org/10.1136/bmjopen-2016-011146.
[12] R.C. Petersen, et al., Alzheimer's disease neuroimaging initiative (ADNI), Neurology 74 (3) (Jan. 2010) 201–209, https://doi.org/10.1212/WNL.0b013e3181cb3e25.
[13] D.L. Beekly, et al., The National Alzheimer's Coordinating Center (NACC) database: an alzheimer disease database, Alzheimer Dis. Assoc. Disord. 18 (4) (2004) 270–277.
[14] K.A. Ellis, et al., The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease, Int. Psychogeriatr. 21 (4) (Aug. 2009) 672–687, https://doi.org/10.1017/S1041610209009405.
[15] S. Grueso, R. Viejo-Sobera, Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review, Alzheimer's Res. Ther. 13 (1) (Sep. 2021) 162, https://doi.org/10.1186/s13195-021-00900-w.
[16] A. Kline, et al., Multimodal machine learning in precision health: a scoping review, Npj Digit. Med. 5 (1) (Nov. 2022) 1, https://doi.org/10.1038/s41746-022-00712-8.
[17] R. Logan, et al., Deep convolutional neural networks with ensemble learning and generative adversarial networks for Alzheimer's disease image data classification, Front. Aging Neurosci. 13 (2021) [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnagi.2021.720226. (Accessed 6 March 2023).
[18] J.F. Beltrán, B.M. Wahba, N. Hose, D. Shasha, R.P. Kline, A.D.N. Initiative, Inexpensive, non-invasive biomarkers predict Alzheimer transition using machine learning analysis of the Alzheimer's Disease Neuroimaging (ADNI) database, PLoS One 15 (7) (Jul. 2020) e0235663, https://doi.org/10.1371/journal.pone.0235663.
[19] J. Venugopalan, L. Tong, H.R. Hassanzadeh, M.D. Wang, Multimodal deep learning models for early detection of Alzheimer's disease stage, Sci. Rep. 11 (1) (Feb. 2021) 1, https://doi.org/10.1038/s41598-020-74399-w.

[20] Y.C. Li, L. Wang, J.N. Law, T.M. Murali, G. Pandey, Integrating multimodal data through interpretable heterogeneous ensembles, Bioinforma. Adv. 2 (1) (Jan. 2022) vbac065, https://doi.org/10.1093/bioadv/vbac065.

[21] D. Greene, P. Cunningham, A matrix factorization approach for integrating multiple data views, in: W. Buntine, M. Grobelnik, D. Mladenić, J. Shawe-Taylor (Eds.), Machine Learning and Knowledge Discovery in Databases, Springer, Berlin, Heidelberg, 2009, pp. 423–438, https://doi.org/10.1007/978-3-642-04180-8_45. Lecture Notes in Computer Science.

[22] M.W. Libbrecht, W.S. Noble, Machine learning applications in genetics and genomics, Nat. Rev. Genet. 16 (6) (Jun. 2015) 6, https://doi.org/10.1038/nrg3920.

[23] V. Gligorijević, N. Pržulj, Methods for biological data integration: perspectives and challenges, J. R. Soc. Interface 12 (112) (Nov. 2015) 20150571, https://doi.org/10.1098/rsif.2015.0571.

[24] R.V. Marinescu, et al., TADPOLE Challenge: accurate Alzheimer's disease prediction through crowdsourced forecasting of future data, Predict. Intell. Med. PRIME Workshop 11843 (Oct. 2019) 1–10, https://doi.org/10.1007/978-3-030-32281-6_1.

[25] L.A. Beckett, M.C. Donohue, C. Wang, P. Aisen, D.J. Harvey, N. Saito, The Alzheimer's disease neuroimaging initiative 2: increasing the length, breadth, and depth of our understanding, Alzheimers Dement. J. Alzheimers Assoc. 11 (7) (Jul. 2015) 823–831, https://doi.org/10.1016/j.jalz.2015.05.004.

[26] A.W. Toga, K.L. Crawford, The Alzheimer's disease neuroimaging initiative informatics core: a decade in review, Alzheimers Dement. J. Alzheimers Assoc. 11 (7) (Jul. 2015) 832–839, https://doi.org/10.1016/j.jalz.2015.04.004.

[27] J.D. Dziura, L.A. Post, Q. Zhao, Z. Fu, P. Peduzzi, Strategies for dealing with missing data in clinical trials: from design to analysis, Yale J. Biol. Med. 86 (3) (Sep. 2013) 343–358.

[28] S. Nijman, et al., Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review, J. Clin. Epidemiol. 142 (Feb. 2022) 218–229, https://doi.org/10.1016/j.jclinepi.2021.11.023.

[29] B. Fischl, FreeSurfer, Neuroimage 62 (2) (Aug. 2012) 774–781, https://doi.org/10.1016/j.neuroimage.2012.01.021.

[30] O. Troyanskaya, et al., Missing value estimation methods for DNA microarrays, Bioinformatics 17 (6) (Jun. 2001) 520–525, https://doi.org/10.1093/bioinformatics/17.6.520.

[31] T. Al-Shehari, R.A. Alsowail, An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques, Entropy 23 (10) (Sep. 2021) 1258, https://doi.org/10.3390/e23101258.

[32] M. Kivipelto, et al., Apolipoprotein E ε4 magnifies lifestyle risks for dementia: a population-based study, J. Cell Mol. Med. 12 (6b) (Dec. 2008) 2762–2771, https://doi.org/10.1111/j.1582-4934.2008.00296.x.

[33] M. Safieh, A.D. Korczyn, D.M. Michaelson, ApoE4: an emerging therapeutic target for Alzheimer's disease, BMC Med. 17 (Mar. 2019) 64, https://doi.org/10.1186/s12916-019-1299-4.

[34] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Methodol. 57 (1) (1995) 289–300.

[35] N. Japkowicz, S. Stephen, The class imbalance problem: a systematic study, Intell. Data Anal. 6 (5) (Oct. 2002) 429–449.

[36] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16, Association for Computing Machinery, New York, NY, USA, Aug. 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[37] R. Shwartz-Ziv, A. Armon, "Tabular data: deep learning is not all You need.", arXiv, Nov. 23 (2021) https://doi.org/10.48550/arXiv.2106.03253.

[38] S. Whalen, O.P. Pandey, G. Pandey, Predicting protein function and other biomedical characteristics with heterogeneous ensembles, Methods San Diego Calif 93 (Jan. 2016) 92–102, https://doi.org/10.1016/j.ymeth.2015.08.016.

[39] P. Radivojac, et al., A large-scale evaluation of computational protein function prediction, Nat. Methods 10 (3) (Mar. 2013) 221–227, https://doi.org/10.1038/nmeth.2340.

[40] N. Zhou, et al., The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens, Genome Biol. 20 (1) (Nov. 2019) 244, https://doi.org/10.1186/s13059-019-1835-8.

[41] S.A. Hicks, et al., On evaluation metrics for medical applications of artificial intelligence, Sci. Rep. 12 (Apr. 2022) 5979, https://doi.org/10.1038/s41598-022-09954-8.

[42] N. Altman, M. Krzywinski, Graphical assessment of tests and classifiers, Nat. Methods 18 (8) (Aug. 2021) 8, https://doi.org/10.1038/s41592-021-01232-1.

[43] M.N. Samtani, et al., Alzheimer's disease assessment scale-cognitive 11-item progression model in mild-to-moderate Alzheimer's disease trials of bapineuzumab, Alzheimers Dement. Transl. Res. Clin. Interv. 1 (3) (Oct. 2015) 157–169, https://doi.org/10.1016/j.trci.2015.09.001.

[44] J. K. Kueper, M. Speechley, and M. Montero-Odasso, "The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. A narrative review," J. Alzheimers Dis., vol. 63, no. 2, pp. 423–444, doi: 10.3233/JAD-170991.

[45] M.P. Laakso, et al., Hippocampal volumes in Alzheimer's disease, Parkinson's disease with and without dementia, and in vascular dementia: an MRI study, Neurology 46 (3) (Mar. 1996) 678–681, https://doi.org/10.1212/WNL.46.3.678.

[46] T.E. Evans, et al., Subregional volumes of the hippocampus in relation to cognitive function and risk of dementia, Neuroimage 178 (Sep. 2018) 129–135, https://doi.org/10.1016/j.neuroimage.2018.05.041.

[47] J.A. Kaye, et al., Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia, Neurology 48 (5) (May 1997) 1297–1304, https://doi.org/10.1212/WNL.48.5.1297.

[48] R. Mayeux, Y. Stern, Epidemiology of Alzheimer disease, Cold Spring Harb. Perspect. Med. 2 (8) (Aug. 2012) a006239, https://doi.org/10.1101/cshperspect.a006239.

[49] R. Wei, C. Li, N. Fogelson, L. Li, Prediction of conversion from mild cognitive impairment to Alzheimer's disease using MRI and structural network features, Front. Aging Neurosci. 8 (Apr. 2016), https://doi.org/10.3389/fnagi.2016.00076.

[50] S. Guo, B. Xiao, C. Wu, Identifying subtypes of mild cognitive impairment from healthy aging based on multiple cortical features combined with volumetric measurements of the hippocampal subfields, Quant. Imag. Med. Surg. 10 (7) (Jul. 2020) 1477–1489, https://doi.org/10.21037/qims-19-872.

[51] L. Zhao, et al., Risk estimation before progression to mild cognitive impairment and Alzheimer's disease: an AD resemblance atrophy index, Aging 11 (16) (Aug. 2019) 6217, https://doi.org/10.18632/aging.102184.

[52] A.C. van Loenhoud, et al., Association of education and intracranial volume with cognitive trajectories and mortality rates across the Alzheimer disease continuum, Neurology 98 (16) (Apr. 2022) e1679–e1691, https://doi.org/10.1212/WNL.0000000000200116.

[53] H. An, et al., Large intracranial volume accelerates conversion to dementia in males and APOE4 non-carriers with mild cognitive impairment, Int. Psychogeriatr. 28 (5) (Dec. 2015) 769–778, https://doi.org/10.1017/S104161021500229X.

[54] D.F. Tate, et al., Intracranial volume and dementia: some evidence in support of the cerebral reserve hypothesis, Brain Res. 1385 (Dec. 2010) 151–162, https://doi.org/10.1016/j.brainres.2010.12.038.