Deep Learning with Imbalanced Data Identifies Early Mild Cognitive Impairment Amongst Typical Older Adults

Ruben Dario Zapata, Alejandro Albizu, Cameron Perrin, Andrew O'Shea, Cheshire Hardcastle, Jessica Kraft, Gene E. Alexander, Michael Marsiske, Ronald A. Cohen, Adam J. Woods, Joseph M. Gullett

Submitted to: JMIR AI
on: August 14, 2024

Table of Contents

Deep Learning with Imbalanced Data Identifies Early Mild Cognitive Impairment Amongst Typical Older Adults

Ruben Dario Zapata[1]; Alejandro Albizu[2] PhD; Cameron Perrin[3] BSc; Andrew O'Shea[2] MS; Cheshire Hardcastle[2] PhD; Jessica Kraft[4] PhD; Gene E. Alexander[5] PhD; Michael Marsiske[2] PhD; Ronald A. Cohen[2] PhD; Adam J. Woods[2] PhD; Joseph M. Gullett[2] PhD

[1]Department of Health Outcomes and Biomedical Informatics University of Florida College of Medicine Gainesville US
[2]Department of Clinical & Health Psychology University of Florida Gainesville US
[3]Department of Clinical & Health Psychology University of Florida College of Medicine Gainesville US
[4]Department of Psychology University of Texas at Dallas School of Behavioral and Brain Sciences Dallas US
[5]Department of Psychiatry, Neuroscience and Physiological Sciences Graduate Interdisciplinary Programs, and BIO5 Institute, and Arizona Alzheimer's Disease Consortium University of Arizona Tucson US

Corresponding Author:
Joseph M. Gullett PhD
Department of Clinical & Health Psychology
University of Florida
COMMUNICORE BLDG RM
C2-018A 1249 CENTER DR GAINESVILLE FL 32610
Gainesville
US

Abstract

Background: Mild cognitive impairment is a transitional stage between normal cognitive aging and dementia. Identifying individuals at the preMCI stage, prior to the onset of mild cognitive decline, can be pivotal for early interventions aimed to reduce the progression neurodegeneration.

Objective: The objective of this study is to develop convolutional neural networks trained on fluid attenuated inversion recovery magnetic resonance imaging for classification of preMCI in an imbalanced cohort of 350 participants.

Methods: A DenseNet 264 convolutional neural network was trained on an imbalanced dataset of 350 participants with a dataset split into 64%, training, 16% validation, and 20% testing sets. Training was conducted with a batch size=70, epoch=200, processing images resized to a uniform dimension of 128×128×128 voxels, and optimizer=Adam. The optimization of our network was conducted using the Adam Optimizer with a learning rate of 10-3 and a weight decay of 5-4. Data augmentation strategy included Random affine transformation, random rotation across an axis, and random Gaussian noise on each image during training.

Results: Mean age of the participants was 71.6 (SD 5.14); the average educational attainment was 16.3 years (SD 2.39). The mean MoCA score was 26.7 (SD 1.93). Our DenseNet mode achieved R2=0.146 and RMSE=0.569.

Conclusions: These findings underscore the potential of brain imaging and DL to identify this at-risk population, offering a promising tool for early detection and potential personalized preventative strategies against cognitive decline. Further research is warranted to improve upon the results to validate these findings in a larger, more diverse population.

(JMIR Preprints 14/08/2024:65411)
DOI: https://doi.org/10.2196/preprints.65411

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).

✓ **Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will**

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Deep Learning with Imbalanced Data Identifies Early Mild Cognitive Impairment Amongst Typical Older Adults

Authors: Ruben D. Zapata, M.S[1]., Alejandro Albizu, Ph.D. [2], Cameron Perrin, B.S. [2], Andrew O'Shea, M.S. [2], Cheshire Hardcastle, Ph.D. [2], Jessica N. Kraft, Ph.D.[4], Gene E. Alexander, Ph.D.[3], Michael Marsiske, Ph.D. [2], Ronald A. Cohen, Ph.D. [2], Adam J. Woods, Ph.D.[2], & Joseph M. Gullett, Ph.D. [2]*
*Corresponding Author: Joseph M. Gullett, Ph.D. gullettj@ufl.edu
[1]University of Florida, Department of Health Outcomes and Biomedical Informatics, College of Medicine, Gainesville, FL, USA
[2]University of Florida, Department of Clinical & Health Psychology, Gainesville, FL, USA
[3]University of Arizona, Department of Psychiatry, Neuroscience and Physiological Sciences Graduate Interdisciplinary Programs, and BIO5 Institute, and Arizona Alzheimer's Disease Consortium, Tucson, AZ, USA
[4]University of Texas at Dallas, Center for Vital Longevity, Department of Psychology, School of Behavioral and Brain Sciences, Dallas, Texas

Abstract
Background
Mild cognitive impairment is a transitional stage between normal cognitive aging and dementia. Identifying individuals at the preMCI stage, prior to the onset of mild cognitive decline, can be pivotal for early interventions aimed to reduce the progression neurodegeneration.
Objective
The objective of this study is to develop convolutional neural networks trained on fluid attenuated inversion recovery magnetic resonance imaging for classification of preMCI in an imbalanced cohort of 350 participants.
Methods
A DenseNet 264 convolutional neural network was trained on an imbalanced dataset of 350 participants with a dataset split into 64%, training, 16% validation, and 20% testing sets. Training was conducted with a batch size=70, epoch=200, processing images resized to a uniform dimension of 128×128×128 voxels, and optimizer=Adam. The optimization of our network was conducted using the Adam Optimizer with a learning rate of 10^{-3} and a weight decay of 5^{-4}. Data augmentation strategy included Random affine transformation, random rotation across an axis, and random Gaussian noise on each image during training.
Results
Mean age of the participants was 71.6 (SD 5.14); the average educational attainment was 16.3 years (SD 2.39). The mean MoCA score was 26.7 (SD 1.93). Our DenseNet mode achieved R^2=0.146 and RMSE=0.569.
Conclusion
These findings underscore the potential of brain imaging and DL to identify this at-risk population, offering a promising tool for early detection and potential personalized preventative strategies against cognitive decline. Further research is warranted to improve upon the results to validate these findings in a larger, more diverse population.

INTRODUCTION

Memory decline in healthy aging is common and occurs more rapidly for episodic and working memory functions specifically (Luo & Craik, 2008; W. C. Wang et al., 2016). As we age, increasing memory difficulties often become concerning due to the threat of potential Alzheimer's disease (AD). Further, should age-related cognitive decline reach the level of mild cognitive impairment (MCI), those MCI-diagnosed patients who demonstrate impaired memory (amnestic MCI) are much

more likely to develop dementia when compared to the 1-2% yearly rate observed in the general population (Petersen et al., 1999) or even individuals with general cognitive complaints (Visser et al., 2006). For this reason, recent studies have sought to identify older adults at risk for dementia development prior to the occurrence of true normative impairment; when cognitive changes were subtle and before they reached the established clinical threshold. Studies examining this so called "pre-MCI" construct find it to be a viable predictor of future decline to dementia, especially in those with amnestic weaknesses who appear to progress to dementia at a much higher rate than those without cognitive impairment (Duara et al., 2011; Loewenstein et al., 2012).

Pre-MCI is considered an intermediate stage between normal cognition and MCI, which may indicate a greater risk of future cognitive decline (Duara et al., 2011). Pre-MCI criteria is variable and heterogeneous within the literature, but generally seeks to identify individuals who do not meet full criteria for a mild cognitive impairment but show subtle cognitive weakness via objective assessment (Chipi et al., 2019). Prior studies suggest the pre-MCI stage exists when cognitive performance is 1 to 1.5 SD below expected levels, with no evidence of cognitive impairment upon clinical evaluation (Loewenstein et al., 2012; Storandt et al., 2006). Additionally, the presence of subjective cognitive complaints increases risk of progression to dementia, despite normal cognitive functioning (L. Wang et al., 2004). Therefore, the occurrence of subjective cognitive complaints is often incorporated into the pre-MCI criteria (Chipi et al., 2019; Crocco et al., 2018; Loewenstein et al., 2016). Pre-MCI individuals have higher rates of progression to MCI, show greater declines in memory performance and hippocampal volume over time, and have reduced cerebral metabolism in AD signature regions (i.e. posterior cingulate cortex) that is associated with poorer verbal memory (Caselli et al., 2008; Crocco et al., 2018; Duara et al., 2011; Loewenstein et al., 2012). Therefore, it is important to identify those in our sample who may be at a higher probability of future cognitive impairment.

Artificial intelligence approaches are increasingly used as a means to identify patients at-risk for dementia. Approaches such as Support Vector Machine (SVM) and Convolutional Neural Networks (CNN) provide the opportunity to predict group membership with high accuracy using objective biomarker data (Bernal et al., 2019; Hojjati et al., 2018; Payan & Montana, 2015; Wen et al., 2020). These types of approaches separate healthy individuals from those with potential MCI or dementia as well as identify individuals who are at-risk for potential decline using longitudinal data (Gullett et al., 2021; Jo et al., 2019; Moradi et al., 2015). However, these approaches have traditionally required a large amount of varying data to yield accurate group predictions, such as integration of cognitive performance and demographic data with multiple modalities of neuroimaging data. Further, lower predictive accuracy is typically achieved when the severity of participants' clinical diagnoses is closer together along the diagnostic spectrum. In other words, stronger predictive accuracy is achieved when distinguishing cognitively intact older adults from those with dementia and weaker predictive accuracy is achieved when attempting to separate cognitively intact older adults from MCI or MCI from dementia. As such, attempting to differentiate cognitively intact individuals from other cognitively intact individuals who scored at the lower end of the normal range is important as it may be an early indicator of future decline, but creates a tremendous hurdle for achieving accurate results with objective biomarker data; in this case, magnetic resonance imaging (MRI). Given that early identification is of the utmost importance for early intervention to potentially delay progression to dementia, developing a model that can accurately distinguish fine gradations of cognitive functioning differences in otherwise healthy older adults with objective biomarker data alone is a critical challenge.

The use of machine learning models incorporating multiple MRI modalities to predict group membership or disease progression has become increasingly common, although mixed results have been found when testing whether the addition of other modalities increases the accuracy of overall model (Gullett et al., 2021; Hojjati et al., 2018; Moradi et al., 2015). As such, incorporating MRI modalities with increased depth and variety of data may yield stronger predictive value; one such

modality being fluid attenuated inversion recovery (FLAIR). FLAIR imaging is sensitive to white matter ischemia, white matter disease and lesions, and contains much of the same structural feature information as commonly used T1 images. Research has found data gathered via FLAIR imaging correlates with cognitive impairment, aging, and dementia conversion, as well as other validated CSF-derived biomarkers of AD (Bahsoun et al., 2022; Chan et al., 2023; Crystal et al., 2023; DiGregorio et al., 2022). Furthermore, white matter disease is frequently present in the brains of aged individuals and increases in prevalence as thinking changes occur, even within cognitively intact older adults (Dadar et al., 2022). The use of deep learning with FLAIR imaging has been found to be highly predictive of diseases of white matter, such as multiple sclerosis (Yılmaz Acar et al., 2022), or when classifying subtypes of vascular cognitive impairment among a sample of normal aging and affected older adults (Q. Chen et al., 2020). Thus, the FLAIR sequence provides a unique opportunity to incorporate more data and more complex information into machine learning models to yield potentially higher predictive accuracy.

Comparing cognitively intact older adults to pre-MCI samples also leads to an issue of imbalanced data, as there are many more cognitively intact adults than pre-MCI. Machine learning models have difficulty achieving a high specificity score when evaluating imbalanced data; the overall deep learning model may perform well on the majority cases but perform poorly on the minority cases of interest. In the present study, we sought to leverage multimodal MRI techniques in a homogenous sample of typically aging older adults to predict the presence of pre-MCI using deep learning approaches (Huang et al., 2017; Wen et al., 2020). Specifically, we sought to determine whether use of deep learning methods with T2 fluid attenuated inversion recovery (FLAIR) imaging data would provide sufficient accuracy for the separation of typically aging older adults from those with mild thinking changes classified as pre-MCI. Lastly, we compare the performance of the utilized convolutional neural network (CNN) to our previously established support vector machine (SVM) model.

METHODS

Participant Selection

Data were collected at baseline from participants recruited through the Augmenting Cognitive Training in Older Adults (ACT, R01AG054077) study (Woods et al., 2018). The 350 participants from the present study were selected from a larger pool of 396 potential participants if they met the following criteria: a) high quality T2 FLAIR, and b) complete neurocognitive testing to allow for computation of the pre-MCI variable.

Participant Screening and Classification

The inclusion and exclusion criteria for the larger study are detailed in a prior work by Woods et al. (Woods et al., 2018). Briefly, participants were between the ages of 65 and 89 with no history of brain or head injury resulting in loss of consciousness greater than 20 minutes, no history of major psychiatric illness, and no formal diagnosis or evidence of MCI, dementia, or neurological brain disease. All participants were right-handed and had no contraindications for MRI. Prior to beginning study procedures, participants signed a consent form approved by the Institutional Review Boards at the University of Florida and at the University of Arizona.

To screen for individuals with possible MCI or dementia, the study used the Uniform Data Set (UDS) from the National Alzheimer's Coordinating Center (NACC) (Weintraub et al., 2018). Individuals were excluded from the parent study with possible MCI if they scored 1.5 standard deviations below the normative mean in any of the following cognitive domains: general cognition, memory, visuospatial, executive functioning/working memory, or language. Further, older adults who performed above the 80th percentile on a composite measure of cognitive training were excluded to prevent ceiling effects in the larger clinical trial, which focused on a cognitive training intervention. For the purposes of determining pre-MCI status, participants were considered pre-MCI

if they scored at least 0.5 standard deviations below the normative mean on a combination of delayed verbal (Craft Story) and non-verbal (Benson Figure) recall memory (see formula in next section). As such, all participants classified as pre-MCI had combined delayed verbal recall normative scores between -0.5 and -1.5 standard deviations below the mean.

Neuropsychological Battery

Participants completed a comprehensive neuropsychological evaluation which assessed various cognitive domains. Within the larger overall battery of tests, verbal learning and memory was measured using Craft 21 Story Recall (Craft et al., 1996); confrontation naming was assessed with the MINT (Gollan et al., 2012); visuospatial cognitive functioning was evaluated with the Benson Figure Drawing (Possin et al., 2011) and Block Design (Wechsler et al., 2008); executive function was assessed with the Stroop Test Color-Word trial (Stroop, 1935; Tobergte & Curtis, 2013; Trenerry et al., 2012) and Trail Making Test - B (Arango-Lasprilla, Rivera, Aguayo, et al., 2015; Reitan, 1958); verbal fluency was assessed using category (Benton, 1968; Ostrosky-Solis et al., 2007) and phonemic fluency (Ruff et al., 1996); and non-verbal memory was measured using the delayed recall of the aforementioned Benson Figure Drawing. For the purposes of this study as a means of identifying those particularly at risk for developing MCI, participants who scored between -0.5 and -1.5 standard deviations below the mean on a Memory Composite were classified as pre-MCI according to the formula:

Memory Composite = (Craft Story Delay Z-Score + Benson Figure Delay Z-score)/2

Magnetic Resonance Imaging

Participants completed a 1-hour MRI acquisition on a Siemens Prisma 3T MRI scanner (Siemens Medical Solutions, Erlangen, Germany) with 64-channel head coil at the University of Florida and on a 3T Siemens Skyra with a 32-channel head coil at University of Arizona. Both sites implemented identical scanning procedures and sequences. A 3D T1 weighted volumetric magnetization-prepared rapid gradient-echo sequence (MPRAGE) sequence was utilized for visualization of potential exclusionary findings. This MPRAGE sequence was three minutes in duration with a protocol as follows: TR = 1,800 ms, TE = 2.26 ms, flip angle = 8°, voxel size = 1.0 mm^3, 176 slices; FOV = 256 × 256 mm. The three-minute T2-weighted fluid-attenuated inversion recovery (FLAIR) imaging protocol was as follows: TR = 7,000 ms; TE = 101 ms; 1.0 × 1.0 × 2.5 mm3 voxels; 55 slices; FOV = 256 × 256 mm; FA = 120. The T2-weighted FLAIR sequence, which provides high tissue contrast and high spatial resolution with whole brain coverage, was also used for visualization of potential exclusionary incidental findings.

Structural MRI Pre-processing

Individual MPRAGE and FLAIR images were converted from DICOM to NIFTI using dcm2niix (Li et al., 2016). Raw FLAIR images for each participant were included in the model as opposed to warping the MRI images because 1) it decreases the preprocessing on the front end, allowing the deployed model to make inferences on raw MRI data with little latency, 2) the variance in head placement allows us to use data augmentation by injecting further head motion to force the model to “learn” the brain, and 3) it creates a more accessible model for clinicians that might not have access to preprocessing software.

Supervised Machine-Learning

Convolutional Neural Network

In the development of our deep learning framework for prognostic memory performance, we employed the DenseNet264 convolutional neural network (Huang et al., 2017) of FLAIR images, utilizing a significant growth rate (k=32) across the network to optimize feature representation. Our training/validation/testing data sets were 64%/16%/20%, respectively. Our parameters for training include a batch size=70, epoch=200, processing images resized to a uniform dimension of 128×128×128 voxels, and optimizer=Adam. Optimizer algorithms were compared based on performance between Adam and SGD between 2 full training cycles, and the Adam optimizer was

chosen based on performance of the model. The optimization of our network was conducted using the Adam Optimizer with a learning rate of 10^{-3} and a weight decay of 5^{-4}. A data augmentation strategy was used to enhance model robustness and generalizability. A total of three augmentations were applied to each image in the training dataset: Random affine transformation, random rotation across an axis, and random Gaussian noise. Each image was subjected to random translations of two voxels in any axial direction to account for slight variations in brain positioning within the MRI scanner. Images were also randomly rotated around the x-axis by approximately ± 5 degrees, which introduced anatomically plausible head rotations into the images. Finally, random Gaussian noise was added to mimic the electronic noise inherent in the MRI data acquisition. The augmentation process resulted in a final training set of 1115 images. These augmentations were carefully chosen to simulate the variability and imperfections present in real-world medical imaging data, thereby preparing the model to handle a wide range of imaging scenarios (Huang et al., 2017).

Support Vector Machine

To compare our CNN model, we also used Support Vector Machine (SVM); a machine learning algorithm to search for the optimal hyperplane that predicts the target with less than ε error, which is satisfied in this study. Specifically, Scikit-Learn (Abraham et al., 2014; Pedregosa et al., 2011) was used to optimize the objective function:

$$\max_{w,b} \frac{1}{2} w^T w + c \sum_{i=1}^{n} max\left(\left|y_i - \left(w^T x_i + b\right)\right| - \varepsilon, 0\right)$$

SVM performance was evaluated with ten-fold stratified cross-validation.

RESULTS

A total of 350 participants met study criteria and were utilized for this secondary data analysis. Mean age of the participants was 71.6 (SD 5.14); the average educational attainment was 16.3 years (SD 2.39). The mean MoCA score was 26.7 (SD 1.93) (Table 1). Seventy participants were classified as meeting criteria for pre-MCI while 280 participants were classified as having normal cognitive function for their age. Participants meeting criteria for pre-MCI were more likely to score lower on the MoCA, Memory Composite, and Language Composite compared to normal individuals (Table 1). Otherwise, groups did not differ significantly on any of the examined demographic or cognitive factors.

Table 1. Demographics and cognitive performance for total sample, typically aging, and pre-MCI groups

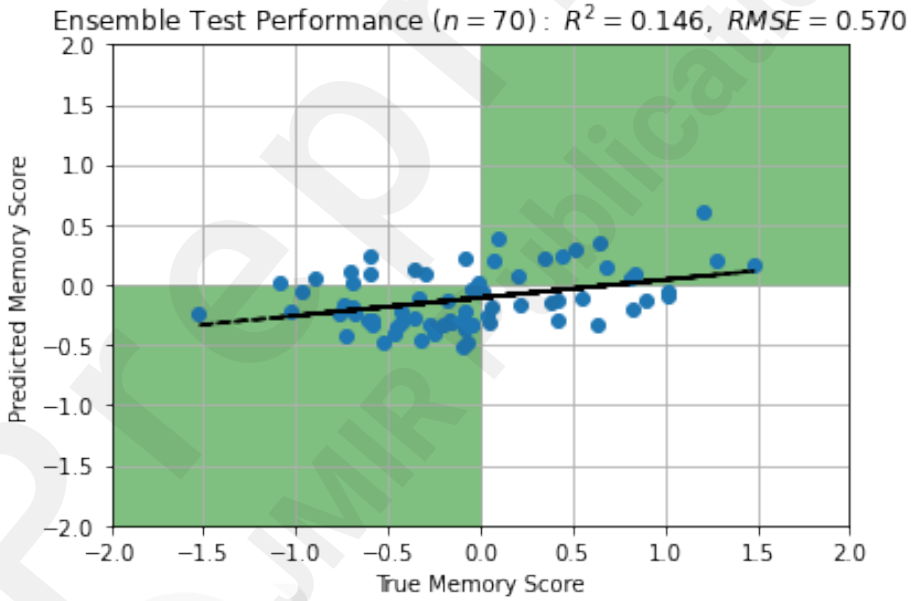
	Total (N=350)	Typically-Aging (N=280)	Pre-MCI [a] (N=70)	χ^2/ t	p-val
Age	71.58 (5.14)	71.41 (4.89)	72.26 (6.04)	-1.08	.282
Education	16.25 (2.40)	16.24 (2.35)	16.30 (2.58)	-0.18	.859
Gender (% Female)	63.4%	63.9%	61.4%	0.15	.698
Race				10.43	.108
Amer. Indian/Alaskan Native	2.6%	3.2%	0%	-	-
Asian	1.1%	1.4%	0%	-	-
Black	5.7%	5.0%	8.6%	-	-
White	87.1%	87.9%	84.3%	-	-
Native Hawaiian/Pacific Islander	0.3%	0.4%	0%	-	-
Multi-racial	1.7%	1.1%	4.3%	-	-
Unknown	1.4%	1.1%	2.9%	-	-
Ethnicity (% Hispanic)	6.0%	6.4%	4.3%	4.41	.110
		Cognitive Performance			
MoCA Total Score	26.69 (1.93)	26.90 (1.83)	25.86 (2.14)	3.75	<.001

Memory Composite[b]	0.05 (0.66)	0.27 (0.52)	-0.86 (0.29)	24.43	<.001
Attention Composite[c]	-0.11 (0.79)	-0.10 (0.77)	-0.16 (0.87)	0.65	.516
Executive Function Composite[d]	-0.01 (0.57)	-0.01 (0.58)	0.13 (0.52)	-0.33	.742
Language Composite[e]	0.14 (0.62)	0.20 (0.61)	-0.06 (0.59)	3.15	.002

[a] Based on Memory Composite Z-score ≤-0.5; [b]((Craft Story Delay Z-Score + Benson Figure Delay Z-score)/2); [c] ((Digit Span Forward Z-score + Digit Span Backward Z-score)/2); [d]((Trails B Z-score + Letter Fluency Z-score)/2); [e]((MINT Naming Z-score + Animal Fluency Z-score)/2).

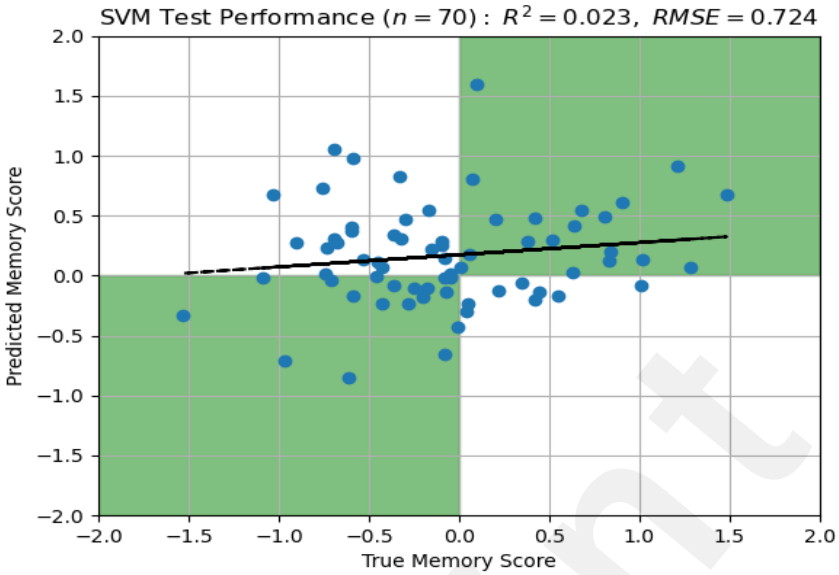
Convolutional Neural Network

The DenseNet264 architecture was evaluated for its effectiveness in leveraging FLAIR data for prediction of NACC memory composite scores, a quantitative measure reflecting an individual's memory performance, which is crucial for early detection and monitoring of cognitive decline. The model demonstrated a coefficient of determination (R^2) of 0.146, indicating approximately 14.6% of the variance in the NACC memory composite scores could be explained by the model's predictions. This is an indication that structural changes in the brain capture a meaningful portion of the variance in memory performance scores of older adults. Moreover, the model also achieved a root mean square error (RMSE) of 0.569, demonstrating the model's capacity to offer reasonably close estimates of an individual's memory performance. In the context of early detection and monitoring of cognitive changes, this can be incredibly valuable.



Support Vector Machine

In evaluating the efficacy of our newly developed model relative to a conventional benchmark, we employed a support vector machine (SVM) for comparative analysis. The performance metrics for the SVM were primarily gauged using the R^2 score, which was recorded at 0.023. This metric suggests that the SVM's predictive capabilities could account for merely ~2% of the variability observed in the NACC memory composite scores when utilizing FLAIR imaging data. Furthermore, the SVM model reported an RMSE of 0.724, highlighting a higher discrepancy in its predictive accuracy in contrast to our CNN model. This comparative assessment underscores the enhanced performance of our CNN model. While the SVM is recognized for its robustness across a variety of regression challenges, it was outperformed by our CNN in the context of predicting cognitive performance metrics.



DISCUSSION

The ability to determine whether a person with normal cognitive function will go on to develop a neurodegenerative disorder is a tremendous challenge in clinical translational research. Use of multimodal machine learning models provides a unique opportunity to take advantage of unseen patterns within neuroanatomical data to overcome this disease prediction challenge. However, recent work demonstrates that the ability to discern MCI, a potential precursor to AD, from normal controls remains relatively low when compared to the ability to differentiate MCI from AD (Odusami et al., 2023). A potential diagnostic method for bridging the gap between normal (healthy) controls and those with MCI is classification of a third, pre-clinical MCI (pre-MCI) group, which has been fruitful at predicting future development of MCI and AD (Duara et al., 2011; Loewenstein et al., 2012, 2018). The present study demonstrates the first deep learning model for early detection of pre-MCI using a multimodal deep learning model. This finding has considerable clinical relevance as it demonstrates the ability for a model to predict pre-MCI based off of FLAIR data at a point in time that interventions or plan of care can be implemented years ahead of potential neurodegeneration and subsequent functional impairment. Our previous study demonstrated that machine learning algorithms have the ability to precisely predict the decline from MCI to AD over a year later with over 90% accuracy (Gullett et al., 2021) . Using deep learning models to examine the difference between pre-MCI and healthy older adults requires an accurate model with large amounts of data in each class to identify subtle differences in MRI data and aid in the detection of the early stages of cognitive decline. Given that the prevalence of pre-MCI is ~20% of the population (Loewenstein et al., 2012), our study had a moderately imbalanced dataset of n=350 , with 70 positive samples pertaining to the pre-MCI group. To our knowledge, this represents one of the largest datasets utilized for pre-MCI classification. Other studies that utilize imaging data have focused on clinical presentations of MCI and AD (P. Chen et al., 2024). Despite this imbalance, our model out performed a state-of-the-art machine learning technique in making this subtle distinction when compared to the SVM model which tends to work well with class imbalance and a small number of training samples.

Despite identifying a substantial number of patients with pre-MCI, the overall dataset was skewed towards participants with normal cognitive function. We applied several techniques to develop the deep learning model to consider the class distribution in the dataset such as oversampling the minority class and under sampling the majority class during the training phase of model development. Each technique was applied individually, to assess the performance of the model once the change was made and then one attempt of applying both techniques together which

concluded with a decrease in overall performance of the model. Furthermore, we implored the use of binary cross entropy with inverse class probability weighting. Predicting rare events is difficult because they occur in scenarios with low probability, leading to imbalanced data. This causes models to prioritize the more frequent class, overlooking important patterns in the minority class (Díez-Pastor et al., 2015). Given the low propensity of capturing pre-MCI, to our knowledge we have developed the first deep learning classifier that is able to classify pre-MCI through an imbalanced dataset. Since pre-MCI occurs in roughly 20% of the population, it is imperative that models be developed on imbalanced data to achieve performances that can generalize and identify this small population. This insight will guide future personalized medicine research and targeted interventions to changes the trajectory of cognitive aging.

Prior work utilizing brain MRI to predict group membership has included structural T1 and resting-state functional images, which when combined often prove to be high performing predictors of disease progression or group membership (Gullett et al., 2021; Hojjati et al., 2018; Moradi et al., 2015). T1 (structural) MRI was used during initial data exploration to determine its usefulness in our classification task for pre-MCI but fell below the chance level. The present study investigated MRI data as features in the model attempting to separate pre-MCI from normal cognition and found that T1 (structural) MRI performance was inadequate for inclusion in the final model (below chance). Of the many reasons for this potential discrepancy, the strongest is likely that structural differences between normal aging and pre-MCI are not yet apparent on T1 MRI at this early stage of prodromal disease progression. FLAIR imaging, however, was found to be a strong predictor of pre-MCI group membership, likely because it contains a combination of structural features and markers of cerebrovascular health and disease (Bahsoun et al., 2022; Chan et al., 2023; Q. Chen et al., 2020; Crystal et al., 2023; DiGregorio et al., 2022) which may make it a stronger early progression biomarker than T1 alone. Additionally, it is likely that despite numerous methods implemented to counteract the negative effects of imbalanced data, the present study likely still suffers from accuracy loss secondary to the low base-rate of pre-MCI in this sample.

CONCLUSION

Machine learning algorithms have achieved good performance in their classification abilities based on a variety of data types. Within a sample of typically aging older adults, it is important to determine the onset of neurodegenerative diseases as early as possible given that these insidious processes are often several years underway before a formal diagnosis is made. Given the prevalence of individuals that would be classified as having pre-MCI is relatively low, our model showcases the ability of machine learning to learn from imbalanced data and to lay the foundation for machine learning and deep learning to continue to be explored in this domain.

Funding: This work was supported by the National Institute on Aging (Grant to AW, R01-AG054077; Grant to JG, K23-AG080127) and the National Science Foundation (Grant to AA, NSF 1842473)

Conflict(s) of Interest: None

Data Availability: Data can be made available by request to the corresponding author given that a formal data sharing agreement is signed by the requesting agency. All software and code used in the present manuscript is freely available to the public upon email request to the corresponding author.

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*(FEB). https://doi.org/10.3389/fninf.2014.00014

Arango-Lasprilla, J. C., Rivera, D., Aguayo, A., Rodríguez, W., Garza, M. T., Saracho, C. P., Rodríguez-Agudelo, Y., Aliaga, A., Weiler, G., Luna, M., Longoni, M., Ocampo-Barba, N., Galarza-Del-Angel, J., Panyavin, I., Guerra, A., Esenarro, L., García De La Cadena, P., Martínez, C., & Perrin, P. B. (2015). Trail Making Test: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, *37*(4), 639–661. https://doi.org/10.3233/NRE-151284

Arango-Lasprilla, J. C., Rivera, D., Garza, M. T., Saracho, C. P., Rodríguez, W., Rodríguez-Agudelo, Y., Aguayo, A., Schebela, S., Luna, M., Longoni, M., Martínez, C., Doyle, S., Ocampo-Barba, N., Galarza-Del-Angel, J., Aliaga, A., Bringas, M., Esenarro, L., García-Egan, P., & Perrin, P. B. (2015). Hopkins Verbal Learning Test- Revised: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, *37*(4), 699–718. https://doi.org/10.3233/NRE-151286

Bahsoun, M. A., Khan, M. U., Mitha, S., Ghazvanchahi, A., Khosravani, H., Jabehdar Maralani, P., Tardif, J. C., Moody, A. R., Tyrrell, P. N., & Khademi, A. (2022). FLAIR MRI biomarkers of the normal appearing brain matter are related to cognition. *NeuroImage: Clinical*, *34*. https://doi.org/10.1016/j.nicl.2022.102955

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia*, *6*(1), 53–60. https://doi.org/10.1016/0028-3932(68)90038-9

Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., & Lladó, X. (2019). Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial Intelligence in Medicine*, *95*. https://doi.org/10.1016/j.artmed.2018.08.008

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*, *5*(2), 125–142. https://doi.org/10.1080/13854049108403297

Caselli, R. J., Chen, K., Lee, W., Alexander, G. E., & Reiman, E. M. (2008). Correlating cerebral hypometabolism with future memory decline in subsequent converters to amnestic pre-mild cognitive impairment. *Archives of Neurology*, *65*(9). https://doi.org/10.1001/archneurol.2008.1

Chan, K., Fischer, C. E., Khosravani, H., Black, S. E., Tyrrell, P., Maralani, P. J., Moody, A. R., & Khademi, A. (2023). FLAIR MRI biomarkers of the Normal-Appearing Brain Matter (NABM) are related to APOE-4 status and CSF markers Aβ42 and Tau. *Alzheimer's & Dementia*, *19*(S3). https://doi.org/10.1002/alz.063352

Chen, P., Zhang, S., Zhao, K., Kang, X., Rittman, T., & Liu, Y. (2024). Robustly uncovering the heterogeneity of neurodegenerative disease by using data-driven subtyping in neuroimaging: A review. *Brain Research*, *1823*, 148675. https://doi.org/10.1016/j.brainres.2023.148675

Chen, Q., Wang, Y., Qiu, Y., Wu, X., Zhou, Y., & Zhai, G. (2020). A Deep Learning-Based Model for Classification of Different Subtypes of Subcortical Vascular Cognitive Impairment With FLAIR. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00557

Chipi, E., Salvadori, N., Farotti, L., & Parnetti, L. (2019). Biomarker-based signature of alzheimer's disease in pre-MCI individuals. *Brain Sciences*, *9*(9). https://doi.org/10.3390/brainsci9090213

Craft, S., Newcomer, J., Kanne, S., Dagogo-Jack, S., Cryer, P., Sheline, Y., Luby, J., Dagogo-Jack, A., & Alderson, A. (1996). Memory improvement following induced hyperinsulinemia in Alzheimer's disease. *Neurobiology of Aging*, *17*(1), 123–130. https://doi.org/10.1016/0197-4580(95)02002-0

Crocco, E. A., Loewenstein, D. A., Curiel, R. E., Alperin, N., Czaja, S. J., Harvey, P. D., Sun, X., Lenchus,

J., Raffo, A., Peñate, A., Melo, J., Sang, L., Valdivia, R., & Cardenas, K. (2018). A novel cognitive assessment paradigm to detect Pre-mild cognitive impairment (PreMCI) and the relationship to biological markers of Alzheimer's disease. *Journal of Psychiatric Research*, *96*. https://doi.org/10.1016/j.jpsychires.2017.08.015

Crystal, O., Maralani, P. J., Black, S., Fischer, C., Moody, A. R., & Khademi, A. (2023). Detecting conversion from mild cognitive impairment to Alzheimer's disease using FLAIR MRI biomarkers. *NeuroImage: Clinical*, *40*. https://doi.org/10.1016/j.nicl.2023.103533

Dadar, M., Mahmoud, S., Zhernovaia, M., Camicioli, R., Maranzano, J., & Duchesne, S. (2022). White matter hyperintensity distribution differences in aging and neurodegenerative disease cohorts. *NeuroImage: Clinical*, *36*. https://doi.org/10.1016/j.nicl.2022.103204

Díez-Pastor, J. F., Rodríguez, J. J., García-Osorio, C. I., & Kuncheva, L. I. (2015). Diversity techniques improve the performance of the best imbalance learning ensembles. *Information Sciences*, *325*. https://doi.org/10.1016/j.ins.2015.07.025

DiGregorio, J., Gibicar, A., Khosravani, H., Maralani, P. J., Tardif, J. C., Tyrrell, P. N., Moody, A. R., & Khademi, A. (2022). Cross-sectional and longitudinal Biomarker extraction and analysis for multicentre FLAIR brain MRI. *Neuroimage: Reports*, *2*(2). https://doi.org/10.1016/j.ynirp.2022.100091

Duara, R., Loewenstein, D. A., Greig, M. T., Potter, E., Barker, W., Raj, A., Schinka, J., Borenstein, A., Schoenberg, M., Wu, Y., Banko, J., & Potter, H. (2011). Pre-MCI and MCI: Neuropsychological, clinical, and imaging features and progression rates. *American Journal of Geriatric Psychiatry*, *19*(11), 951–960. https://doi.org/10.1097/JGP.0b013e3182107c69

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism*, *15*(3), 594–615. https://doi.org/10.1017/S1366728911000332

Gullett, J. M., Albizu, A., Fang, R., Loewenstein, D. A., Duara, R., Rosselli, M., Armstrong, M. J., Rundek, T., Hausman, H. K., Dekosky, S. T., Woods, A. J., & Cohen, R. A. (2021). Baseline Neuroimaging Predicts Decline to Dementia From Amnestic Mild Cognitive Impairment. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.758298

Hausman, H. K., O'Shea, A., Kraft, J. N., Boutzoukas, E. M., Evangelista, N. D., Van Etten, E. J., Bharadwaj, P. K., Smith, S. G., Porges, E., Hishaw, G. A., Wu, S., DeKosky, S., Alexander, G. E., Marsiske, M., Cohen, R., & Woods, A. J. (2020). The Role of Resting-State Network Functional Connectivity in Cognitive Aging. *Frontiers in Aging Neuroscience*, *12*, 177. https://doi.org/10.3389/fnagi.2020.00177

Hojjati, S. H., Ebrahimzadeh, A., Khazaee, A., & Babajani-Feremi, A. (2018). Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI. *Computers in Biology and Medicine*, *102*, 30–39. https://doi.org/10.1016/j.compbiomed.2018.09.004

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*. https://doi.org/10.1109/CVPR.2017.243

Ibrahim, B., Suppiah, S., Ibrahim, N., Mohamad, M., Hassan, H. A., Nasser, N. S., & Saripan, M. I. (2021). Diagnostic power of resting-state fMRI for detection of network connectivity in Alzheimer's disease and mild cognitive impairment: A systematic review. In *Human Brain Mapping* (Vol. 42, Issue 9). https://doi.org/10.1002/hbm.25369

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data - Recommendations for the use of performance metrics. *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*. https://doi.org/10.1109/ACII.2013.47

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00220

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001

Loewenstein, D. A., Curiel, R. E., Duara, R., & Buschke, H. (2018). Novel Cognitive Paradigms for the Detection of Memory Impairment in Preclinical Alzheimer's Disease. *Assessment*, *25*(3), 348–359. https://doi.org/10.1177/1073191117691608

Loewenstein, D. A., Curiel, R. E., Greig, M. T., Bauer, R. M., Rosado, M., Bowers, D., Wicklund, M., Crocco, E., Pontecorvo, M., Joshi, A. D., Rodriguez, R., Barker, W. W., Hidalgo, J., & Duara, R. (2016). A Novel Cognitive Stress Test for the Detection of Preclinical Alzheimer Disease: Discriminative Properties and Relation to Amyloid Load. *American Journal of Geriatric Psychiatry*, *24*(10), 804–813. https://doi.org/10.1016/j.jagp.2016.02.056

Loewenstein, D. A., Greig, M. T., Schinka, J. A., Barker, W., Shen, Q., Potter, E., Raj, A., Brooks, L., Varon, D., Schoenberg, M., Banko, J., Potter, H., & Duara, R. (2012). An investigation of PreMCI: Subtypes and longitudinal outcomes. *Alzheimer's and Dementia*, *8*(3). https://doi.org/10.1016/j.jalz.2011.03.002

Luo, L., & Craik, F. I. M. (2008). Aging and memory: A cognitive approach. In *Canadian Journal of Psychiatry* (Vol. 53, Issue 6). https://doi.org/10.1177/070674370805300603

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., & Tohka, J. (2015). Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage*, *104*, 398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002

Odusami, M., Maskeliūnas, R., Damaševičius, R., & Misra, S. (2023). Machine learning with multimodal neuroimaging data to classify stages of Alzheimer's disease: a systematic review and meta-analysis. In *Cognitive Neurodynamics*. https://doi.org/10.1007/s11571-023-09993-5

Ostrosky-Solis, F., Gutierrez, A. L., Flores, M. R., & Ardila, A. (2007). Same or different? Semantic verbal fluency across Spanish-speakers from different countries. *Archives of Clinical Neuropsychology*, *22*(3), 367–377. https://doi.org/10.1016/j.acn.2007.01.011

Payan, A., & Montana, G. (2015). Predicting Alzheimer's disease a neuroimaging study with 3D convolutional neural networks. *ICPRAM 2015 - 4th International Conference on Pattern Recognition Applications and Methods, Proceedings*, *2*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*.

Penny, W., Friston, K., Ashburner, J., Kiebel, S., & Nichols, T. (2007). Statistical Parametric Mapping: The Analysis of Functional Brain Images. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier. https://doi.org/10.1016/B978-0-12-372560-8.X5000-1

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3). https://doi.org/10.1001/archneur.56.3.303

Possin, K. L., Laluz, V. R., Alcantar, O. Z., Miller, B. L., & Kramer, J. H. (2011). Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in Alzheimer's disease and behavioral variant frontotemporal dementia. *Neuropsychologia*, *49*(1), 43–48. https://doi.org/10.1016/j.neuropsychologia.2010.10.026

Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, *8*(3), 271–276. https://doi.org/10.2466/pms.1958.8.3.271

Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). Benton controlled Oral Word Association Test: Reliability and updated norms. *Archives of Clinical Neuropsychology*, *11*(4), 329–338. https://doi.org/10.1016/0887-6177(95)00033-X

Shirer, W. R., Jiang, H., Price, C. M., Ng, B., & Greicius, M. D. (2015). Optimization of rs-fMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination.

NeuroImage, *117*, 67–79. https://doi.org/10.1016/j.neuroimage.2015.05.015

Storandt, M., Grant, E. A., Miller, J. P., & Morris, J. C. (2006). Longitudinal course and neuropathologic outcomes in original vs revised MCI and in pre-MCI. *Neurology*, *67*(3). https://doi.org/10.1212/01.wnl.0000228231.26111.6e

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fisch, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Tobergte, D. R., & Curtis, S. (2013). STROOP Test de colores y palabras. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). TEA.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (2012). Stroop Neuropsychological Screening Test (adult). In *SpringerReference*. Psychological Assessment Resources. https://doi.org/10.1007/springerreference_183456

Visser, P. J., Kester, A., Jolles, J., & Verhey, F. (2006). Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology*, *67*(7). https://doi.org/10.1212/01.wnl.0000238517.59286.c5

Wang, L., Van Belle, G., Crane, P. K., Kukull, W. A., Bowen, J. D., McCormick, W. C., & Larson, E. B. (2004). Subjective memory deterioration and future dementia in people aged 65 and older. *Journal of the American Geriatrics Society*, *52*(12). https://doi.org/10.1111/j.1532-5415.2004.52568.x

Wang, W. C., Daselaar, S. M., & Cabeza, R. (2016). Episodic memory decline and healthy aging. In *The Curated Reference Collection in Neuroscience and Biobehavioral Psychology*. https://doi.org/10.1016/B978-0-12-809324-5.21093-6

Wechsler, D., Coalson, D. L., & Raiford, S. E. (2008). WAIS-IV technical and interpretive manual. In *San Antonio*. Pearson.

Weintraub, S., Besser, L., Dodge, H. H., Teylan, M., Ferris, S., Goldstein, F. C., Giordani, B., Kramer, J., Loewenstein, D., Marson, D., Mungas, D., Salmon, D., Welsh-Bohmer, K., Zhou, X. H., Shirk, S. D., Atri, A., Kukull, W. A., Phelps, C., & Morris, J. C. (2018). Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). *Alzheimer Disease and Associated Disorders*, *32*(1), 10–17. https://doi.org/10.1097/WAD.0000000000000223

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., & Colliot, O. (2020). Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis*, *63*. https://doi.org/10.1016/j.media.2020.101694

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, *2*(3), 125–141. https://doi.org/10.1089/brain.2012.0073

Woods, A. J., Cohen, R., Marsiske, M., Alexander, G. E., Czaja, S. J., & Wu, S. (2018). Augmenting cognitive training in older adults (The ACT Study): Design and Methods of a Phase III tDCS and cognitive training trial. *Contemporary Clinical Trials*, *65*, 19–32. https://doi.org/10.1016/j.cct.2017.11.017

Yılmaz Acar, Z., Başçiftçi, F., & Ekmekci, A. H. (2022). A Convolutional Neural Network model for identifying Multiple Sclerosis on brain FLAIR MRI. *Sustainable Computing: Informatics and Systems*, *35*. https://doi.org/10.1016/j.suscom.2022.100706