scientific reports

OPEN

Brain metabolic imaging-based model identifies cognitive stability in prodromal Alzheimer's disease

Jarrad Perron[1,2], Carly Scramstad[3] & Ji Hyun Ko[1,2,4]✉

The recent approval of anti-amyloid pharmaceuticals for the treatment of Alzheimer's disease (AD) has created a pressing need for the ability to accurately identify optimal candidates for anti-amyloid therapy, specifically those with evidence for incipient cognitive decline, since patients with mild cognitive impairment (MCI) may remain stable for several years even with positive AD biomarkers. Using fluorodeoxyglucose PET and biomarker data from 594 ADNI patients, a neural network ensemble was trained to forecast cognition from MCI diagnostic baseline. Training data comprised PET studies of patients with biological AD. The ensemble discriminated between progressive and stable prodromal subjects (MCI with positive amyloid and tau) at baseline with 88.6% area-under-curve, 88.6% (39/44) accuracy, 73.7% (14/19) sensitivity and 100% (25/25) specificity in the test set. It also correctly classified all other test subjects (healthy or AD continuum subjects across the cognitive spectrum) with 86.4% accuracy (206/239), 77.4% sensitivity (33/42) and 88.23% (165/197) specificity. By identifying patients with prodromal AD who will not progress to dementia, our model could significantly reduce overall societal burden and cost if implemented as a screening tool. The model's high positive predictive value in the prodromal test set makes it a practical means for selecting candidates for anti-amyloid therapy and trials.

The combination of modern healthcare practices which extend life along with decreasing birthrates across the globe have created unprecedented changes in age-related demographics. Since age is the primary risk factor for neurodegenerative disease, it is expected that by the middle of the 21st century there will be approximately 150 million people living with dementia worldwide, with a total economic burden measuring in the trillions of dollars[1,2]. Alzheimer's disease (AD) is the most common cause of age-related neurodegenerative dementia[3]. It is characterized by the accumulation of pathological amyloid protein plaques and tau neurofibrillary tangles which eventually cause neuronal injury and lead to a progressive decline in overall cognitive function[4,5]. The continued neurodegeneration is eventually fatal and those with AD dementia will require extensive personal care, much of which often falls on informal caregivers such as friends and family. While symptomatic treatments have existed for many years, disease-modifying therapies are now available which putatively treat the underlying cause of AD[6–8]. Lecanemab was approved in 2023 as a disease-modifying therapy for AD but is estimated to cost $26,500 USD per year, has the risk of serious side-effects and is only indicated in those with mild cognitive impairment (MCI) or early AD dementia[7,9,10]. Donanemab has also been recently approved but is predicted to have similar cost and secondary effects[8,11]. Aducanumab was the first of this pharmaceutical class to be approved for use, however, it has since been discontinued[12–14]. Reactions to the performance and efficacy of these medications have been mixed in the neurological community and outstanding issues of candidacy for prescription remain to be solved[7,8,10,13–16].

Dementia diagnosis is a common task for clinicians; however, differential diagnoses of AD are prone to error, especially early in the disease progression[6,17–19]. Post-mortem studies show that even for experienced specialists, current practices for ante-mortem diagnosis are somewhat more than modestly sensitive (70.9 − 87.3%) but nonspecific (44.3 − 70.8%) and suffer from issues of replication[17–19]. Furthermore, common clinical methods typically do not make use of protein biomarker information, nor any neuroimaging studies. AD diagnostics

[1]Graduate Program in Biomedical Engineering, Price Faculty of Engineering, University of Manitoba, 75 Chancellor's Circle, Winnipeg, MB R3T 5V6, Canada. [2]PrairieNeuro Research Centre, Kleysen Institute for Advanced Medicine, Health Sciences Centre, 710 William Avenue, Winnipeg, MB R3E 3J7, Canada. [3]Section of Neurology, Department of Internal Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada. [4]Department of Human Anatomy and Cell Science, Rady Faculty of Health Sciences, University of Manitoba, 130-745 Bannatyne Avenue, Winnipeg, MB R3E 0J9, Canada. ✉email: ji.ko@umanitoba.ca

are also complicated by the potentially prodromal MCI state, wherein approximately 20% progress (PMCI) to develop dementia over a 3-year period while the majority remain cognitively stable (SMCI) or even revert to healthy cognition[20–23]. Diagnosis at a later stage where the effects of AD dementia are clinically evaluable may not allow sufficient time for the full benefit of anti-AD drugs to be realized. Even if biomarkers were used as definitive criteria for selecting candidates for anti-amyloid therapies, a substantial proportion of these subjects will remain cognitively stable, and prescription of anti-amyloid drugs would result in unnecessary treatment with pharmaceuticals that are expensive and potentially injurious. This necessitates a balance between the ability to identify those with incipient dementia early enough to realize the fullest practical benefit of anti-AD treatment, but also to prevent those without evidence of such from undergoing unnecessary treatment. Thus, owing to the cost and side effects of currently available anti-AD pharmaceuticals, combined with the modest accuracy of AD differential diagnoses, there exists a pressing unmet need to accurately discriminate between PMCI and SMCI neuropathologies in people with biologically confirmed AD for which we hypothesize machine learning is well-suited[24,25].

In a paradigm of early detection for pharmaceutical intervention, the ability to accurately predict which patients with MCI are not likely to develop AD-related dementia (high positive predictive value) would decrease incidences of unnecessary treatment, reduce risk of side effects and reduce societal/governmental costs in prescribing anti-amyloid pharmaceuticals[2,26]. It may also alleviate overall physician burden and workload due to projected shortages of nuclear medicine specialists and increased demand for readings of positron emission tomography studies[27,28]. Such prognostic capability would also be beneficial to older adults and their families for dementia-related planning and care (wills, personal care, advanced medical directives, powers of attorney, etc.) or in providing peace of mind for patients and their families for a meaningful time window.

We present a machine learning-based model to forecast cognition in patients with prodromal AD (pathological amyloid and tau biomarkers with MCI) through resting-state metabolic neuroimaging studies with fluorodeoxyglucose (FDG) PET from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Metabolic neuroimaging (FDG PET) was chosen for its position as a standard for the differential diagnosis of dementia and its relative commonality within modern healthcare systems[29–31]. This model can differentiate subtle variations in glucose metabolism in the brains of those who progress to dementia versus those who remain cognitively stable at MCI. We introduce a stratification method based on combining cognitive and biomarker statuses and demonstrate the use of novel forms of data augmentation based on frame summation and multi-template registration to increase samples sizes. We construct an ensemble of models and evaluate its performance on multiple independent test sets of different cognitive-biomarker status, specifically upon those patients with prodromal AD, and then visualize the network activations for post-hoc interpretation.

Results

Statistical parametric mapping of prodromal AD

Prior to visualization and training the deep learning ensemble network, we first performed a group-wise two-sample t-test in SPM12 with results shown in Fig. 1 on all amyloid and tau positive (A + T+) subjects at MCI stage. The t-maps show hypometabolism of the angular gyrus, precuneus, the posterior cingulate gyrus (all bilateral) and the middle temporal gyrus (left-only). No hypermetabolic activations survive to significance with family-wise error (FWE) adjustment to $p < 0.05$ (Fig. 1, top). Uncorrected results ($p < 0.001$) are shown below for reference. Without FWE adjustment, several hypermetabolic clusters survive threshold within the cerebellar grey matter, posterior brainstem and cortex anterior to the central sulcus (Fig. 1, bottom).
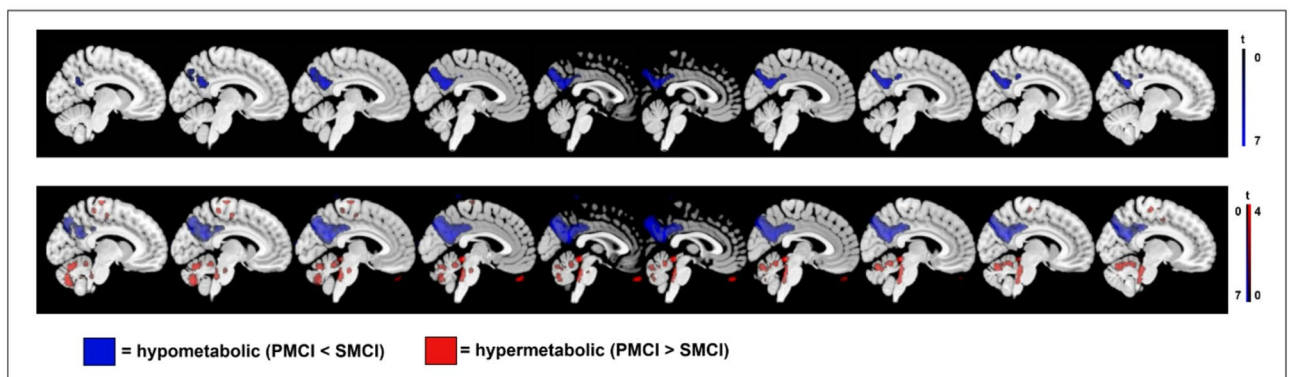

Fig. 1. Results from statistical parametric mapping-based contrast of prodromal MCI subjects. Corrected (above) and uncorrected (below) hypermetabolic (red) and hypometabolic (blue) composite resulting from the SPM12 contrast of all A + T + MCI subjects. Images are in sagittal orientation with equally spaced slices straddling the anatomical midline. Reference group is the pool of A + T + subjects stable at MCI stage for a minimum of 3 years who never develop dementia. An explicit grey matter mask was used. The FWE correction threshold was set with $p < 0.05$. Uncorrected images use $p < 0.001$ as a threshold.

Classification by machine learning models

Both models were compared by area-under-curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE) and both positive and negative predictive values (PPV/NPV). The machine learning model ensemble trained from the first experiment (ML-1) without transfer learning or data augmentation correctly differentiated between the $PMCI_{A+T+}$ and $SMCI_{A+T+}$ subjects at baseline FDG PET study from the test set with 88.2% AUC, 84.1% ACC (38/44), 73.7% SEN (14/19), 92.0% SPE (23/25), 87.5% PPV and 82.1% NPV. The machine learning model trained from the second set (ML-2) with transfer learning and data augmentation did the same with AUC = 88.6%, ACC = 88.6% (40/44), 73.7% SEN (14/19), SPE = 100% (19/19), 100% PPV and 83.3% NPV. Prevalence given by the ratio of the prodromal test set (43.2%) has been assumed.

The models were then applied to non-prodromal subjects, comprising stable healthy controls (SHC), unstable healthy controls (UHC) and subjects with dementia at baseline (DMNT), in addition to SMCI and PMCI subjects who were not positive for both amyloid and tau biomarkers. M-1 correctly classified all SHC_{A-T-} subjects with ACC = 100% (68/68) and 93.3% (28/30) of SHC_{A+T-} subjects while M-2 classified these with 89.7% (61/68) and 93.3% (28/30) accuracy, respectively. M-1 correctly classified 92.3% (16/18) of UHC_{A-T-} subjects while M-2 performed better with 96.2% (17/18). M-1 was able to correctly classify 98.0% (48/49) of $SMCI_{A-T-}$ and 86.7% (26/30) of $SMCI_{A+T-}$ subjects, and M-2 with 93.9% (46/49) and 73.3% (22/30), respectively. Both models correctly identified 64.3% (9/14) of $PMCI_{A+T-}$ subjects, but M-2 correctly labeled 25% (1/4) of $PMCI_{A-T-}$ subjects compared to none for M-1.

Overall, M-1 classified non-prodromal test subjects (all groups except SHC_{A+T+}, UHC_{A+T+}, $SMCI_{A+T+}$, $PMCI_{A+T+}$ and $DMNT_{A+T+}$) with 91.1% ACC (217/239), 59.5% SEN (25/42) and 94.4% SPE (186/197) while M-2 did the same with 86.4% ACC (206/239), 77.4% SEN (33/42) and 88.23% SPE (165/197).

We decided to move forward with a detailed analysis of M-2 for its significantly increased specificity, PPV and task-specific accuracy in the test set over M-1, comparable specificity in the non-prodromal cohorts and greater non-prodromal sensitivity. See Table 1 for a summary of these results and Table 2 for a comparison with other deep learning models on the prodromal test set.

A threshold analysis was performed on M-2. The a priori selected threshold for these results was an ensemble-wise average class probability score of 0.5 (higher = PMCI, lower = SMCI). Using the optimal threshold as defined from cross-validation did not improve performance when changes were made to the testing set.

Effects of cohort and sex on output score

Class probability scores for all subjects output by M-2 underwent 5×3 analysis of variance (ANOVA) followed by post-hoc Bonferroni tests to examine the interaction of cognitive status and biomarker cohort upon machine classification. Cognitive status ($F(4, 593) = 43.95$, $p = 3.17e-32$) and biomarker profile ($F(2, 593) = 17.65$, $p = 3.61e-08$) were significant predictors of output score. Significant interaction between these was found ($F(8, 593) = 4.79$, $p = 1.03e-05$). SHC, UHC and SMCI cognitive cohorts had significantly lower group means than PMCI and DMNT ($p = 4.60e-05$). A-T- biomarker profiles had a significantly lower mean class probability score than A + T- or A + T+ ($p = 8.28e-07$). No significant difference was found between biomarker profiles of SHC, UHC or SMCI subjects ($p = 0.98$). There were significant group differences between A-T- and A + T- or A + T+ for PMCI and DMNT subjects ($p = 1.4e-07$). No significant difference was seen between A + T- and A + T + in PMCI or DMNT groups ($p = 0.97$). Sex was not a significant factor in determining score ($F(2,593) = 0.03$, $p = 0.8675$).

Cognitive profile	Biomarker profile	MI-1	MI-2
SHC	A-T-	**100% (68/68)**	**89.7% (61/68)**
	A + T-	**93.3% (28/30)**	**93.3% (28/30)**
	A + T+	83.9% (26/31)	83.9% (26/31)
UHC	A-T-	**92.3% (16/18)**	**96.2% (17/18)**
	A + T-	**50.0% (1/2)**	**50% (1/2)**
	A + T+	100% (9/9)	100% (9/9)
SMCI	A-T-	**98% (48/49)**	**93.9% (46/49)**
	A + T-	**86.7% (26/30)**	**73.3% (22/30)**
	A + T+	83.1% (64/77)	84.4% (65/77)
PMCI	A-T-	**0% (0/4)**	**25% (1/4)**
	A + T-	**64.3% (9/14)**	**64.3% (9/14)**
	A + T+	80.2% (77/96)	90.6% (87/96)
DMNT	A-T-	**33.3% (2/6)**	**83.3% (5/6)**
	A + T-	**77.8% (14/18)**	**100% (18/18)**
	A + T+	81.7% (116/142)	87.3% (124/142)

Table 1. Performance of M-1 and M-2 on all subjects. Results written in bold script indicate subjects belonged to an independent test subject set. Other results include subjects used in training or split between training and test sets.

Study	PMCI vs. SMCI			
	AUC (%)	ACC (%)	SEN (%)	SPE (%)
Katako (2018)[39]	~	~	67.3	79.6
Popuri (2018)[26]	80.0	~	~	~
Yee (2020)[40]	81.1	74.7	74.0	75.0
Pan (2020)[41]	86.8	83.1	72.1	88.1
Ma (2021)[42]	84.2	72.8	**84.5**	67
Lu (2018)[43]	~	82.5	81.4	82.9
M-1 (ours)	88.2	84.1	73.7	92
M-2 (ours)	**88.6**	**88.6**	73.7	**100**

Table 2. Comparison of M-1 and M-2 performance in holdout test set with other deep learning-based studies. Studies which used any information beyond unimodal FDG PET were excluded from comparison. Metrics in bold script are the highest.
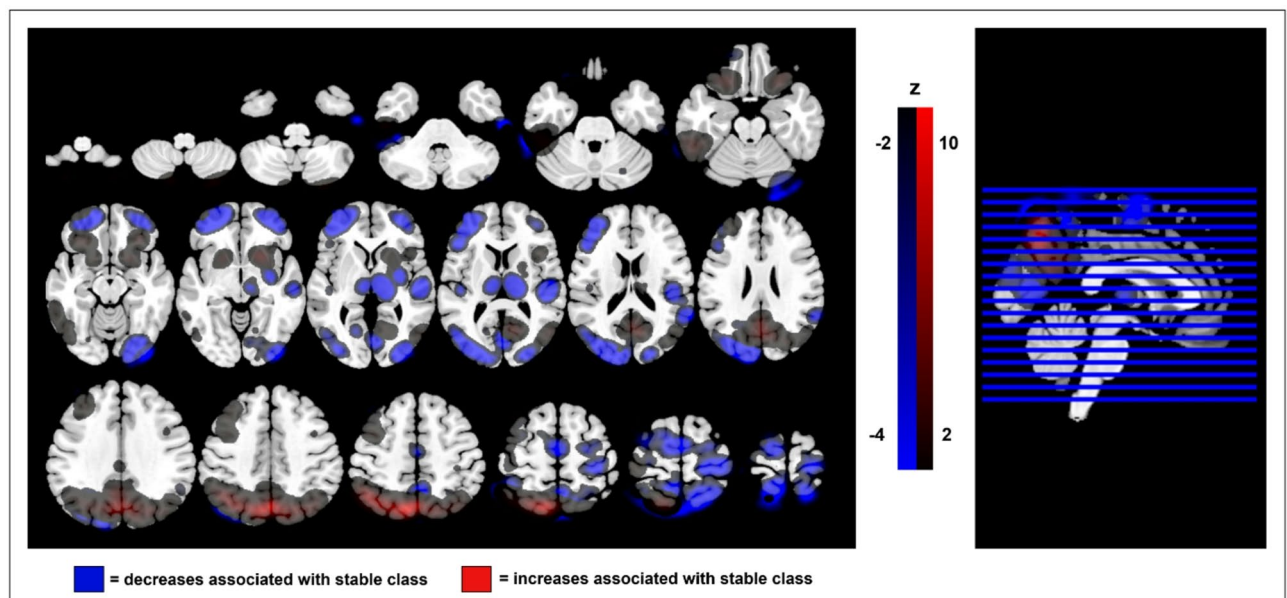

Fig. 2. Grand average GBP heatmaps in prodromal MCI subjects. All images shown are in axial view from slices 16 to 152 every 8 slices. Red clusters are voxels where increasing metabolism positively correlate with increasing probability of SMCI classification (hypometabolically associated with cognitive decline), and blue clusters are voxels where decreasing metabolism correlates with SMCI classification (hypermetabolically associated with cognitive decline).

Visualization of machine class activations

Figure 2 shows the average activations from guided backpropagation (GBP) over all $PMCI_{A+T+}$ and $SMCI_{A+T+}$ subjects overlaid upon the masked T_1-weighted magnetic resonance image provided in CAT12 (http://www.neuro.uni-jena.de/cat). These activations are displayed axially every 8 slices from slices 16 to 152. These maps show how small changes in the voxel intensity (voxel gradient) affect subject classification. The images shown are grand averages over all subnetworks within the ensemble. Regions coloured in blue or red are those where respectively decreasing and increasing voxel intensities (negative and positive voxel gradient, respectively) improve the probability of SMCI classification. Negative and positive values have been separately z-scored for ease of visual interpretation.

Beginning with positive regions in red, we note that the derived clusters overlap broadly with the posterior orbital gyrus, precuneus, superior parietal gyrus, posterior cingulate gyrus (all bilateral) and inferior temporal gyrus, middle frontal gyrus and precentral gyrus (all left-only). For the negative regions in blue, we see the clusters overlap with the cerebellar crura, the superior temporal gyrus, calcarine fissure, cuneus, middle and superior occipital gyrus, dorsolateral superior frontal gyrus, thalamus and paracentral lobule (all bilateral), and the lingual gyrus, fusiform gyrus, lobule IX of the cerebellum (all right-only) and the inferior frontal gyrus (left-only). Of note is the putamen, which shows a very clear anterior-posterior division of positive and negative changes in voxel intensity affecting classification – the only brain structure to do so in our visualizations. See Table 3 for a summary of these results.

Effect on score for stable classification	Region	Notes
Positive (red)	Posterior cingulate gyrus	Bilateral
	Posterior orbital gyrus	Bilateral
	Precuneus	Bilateral
	Superior parietal gyrus	Bilateral
	Inferior temporal gyrus	Left-only
	Middle frontal gyrus	Left-only
	Precentral gyrus	Left-only
Negative (blue)	Calcarine fissure	Bilateral
	Cerebellar crura	Bilateral
	Cuneus	Bilateral
	Dorsolateral superior frontal gyrus	Bilateral
	Middle occipital gyrus	Bilateral
	Paracentral lobule	Bilateral
	Superior occipital gyrus	Bilateral
	Superior temporal gyrus	Bilateral
	Thalamus	Bilateral
	Inferior frontal gyrus	Left-only
	Fusiform gyrus	Right-only
	Lingual gyrus	Right-only
	Lobule IX of the cerebellum	Right-only
Special	Putamen	Bilateral. Anterior has positive contribution, posterior has negative contribution.

Table 3. Important structures for predicting cognitive stability. GBP identified several regions which contribute to the probability of predicting an individual sample as cognitively stable. Red/positive regions are those where increasing voxel values increase the probability of a stable class label while blue/negative regions are those where decreasing voxel values increase the probability of stable classification.

Discussion

The results of the present work propose that FDG PET studies assisted by machine learning can forecast cognition within 3 years of baseline in patients with prodromal AD. 3 years was chosen as a clinically meaningful cut-off period for which predicting cognitive decline would be practical and relevant to treatment with anti-amyloid pharmaceuticals. The sensitivity and specificity were 73% and 100% respectively for differentiating $SMCI_{A+T+}$ and $PMCI_{A+T+}$ within the test set, with AUC of 88.6%. This improves upon previous studies (AUC varied from 80.0% to 86.8%; Table 2). Because of the difficulty for physicians in predicting cognition from the MCI stage, and the upcoming demand on specialists in nuclear medicine or cognitive neurology, our model may significantly reduce overall physician workload and societal burden if implemented as a screening tool for identifying candidates for anti-AD treatment, of which there are approximately 6 million in the USA alone[26–28].

Many studies have investigated the difficult problem of predicting cognition in patients with MCI using various combinations of medical data and different approaches in machine/deep learning[32–37]. Previous work in our lab created the Machine learning-based Alzheimer's disease Detection (MAD) algorithm, which the present work extends considerably[33]. We will therefore refer to M-2 as MAD2. We chose M-2 over M-1 for its significantly increased specificity and exceptional PPV. In the context of selecting appropriate candidates for anti-amyloid therapies, specificity has been previously demonstrated to vastly increase cost-efficiency at the societal scale[26]. The full details of our analysis may be found in the supplementary materials, however, based on models from previous research, we can demonstrate that if MAD2 were used as a screening tool for anti-amyloid therapy in the USA, compared to the status quo, it would cost an additional 40 billion USD to screen every amyloid-positive patient with MCI and treat those with identified incipient dementia over a 3 year period. Compared to prophylactic prescription, it would save roughly 363 billion USD, driven primarily by the specificity of the model and the consequent ability to reasonably forgo treatment in patients labeled as cognitively stable for the interim. There is likely additional cost effectiveness since our cost simulation does not account for quality-adjusted life-years, nor for the increased economic productivity of patients or their caregivers, nor the cost of delaying onset of severe dementia (which requires constant care), nor the cost of treatment for therapeutic side-effects (e.g. cerebral microbleeding, AKA amyloid-related imaging abnormality). The PPV of MAD2 is also of obvious benefit to clinicians and patients who may have confidence when positive predictions are made.

Since only a minority of patients with MCI progress to dementia over a 3 year period, it is paramount to ensure treatment with anti-amyloid pharmaceuticals is motivated by reasonable evidence of incipient dementia, rather than rely solely on amyloid or tau biomarkers due to the exorbitant cost and even potentially undermining the medical principle of nonmaleficence with a possibly injurious treatment of uncertain benefit[25,26]. The results of our study show that amyloid and tau biomarkers are necessary but not sufficient to forecast cognitive decline and affirms that FDG PET is a strong predictor of cognition[38,39].

The current work is an extension to the literature on the topic of cognitive prognostication in several important ways. First, we sought to emphasize clinical applicability through exclusively using the final three frames for validation and testing sets (FDG PET images were provided in 6 × 5-minute frames, most clinical settings use a single static 15-minute acquisition). We were also able to develop a new form of data augmentation using "front" and "whole" frame images that supplemented the "back" frame summed images, further augmented by normalization with two separate PET templates. These differed primarily by their registration along the cortical surface, reflecting the populations from which they were created. For every FDG PET study for a given subject, we were able to generate six studies for training and testing our models. We were able to generate 4,092 training cases from 682 FDG PET studies in 311 unique patients through our novel data augmentation methods.

On the topic of our network architecture, it should be stated that DenseNet-BC was chosen primarily for its ability to aggregate and use features at all levels of the network, no matter where they were generated. While our network outperforms those amongst comparable projects, we do not specifically attribute this to the choice of DenseNet-BC, but rather to our novel data stratification procedure and data augmentation practices.

We also visualized the activations of MAD2 using GBP for post-hoc investigation of the regions of interest. MAD2 emphasizes several regions of the brain well-known within the pathological progression of AD but also highlights regions not commonly associated with AD-related cognitive decline. We hypothesize this is related to the heterogeneity of presentation and progression in patients with MCI. Those (red) clusters where increasing voxels values (increasing metabolism) improve the probability of SMCI classification are easily interpretable as regions traditionally associated with hypometabolism in AD. The precuneus, cingulate gyrus and parietotemporal regions found therein are expected and well-known with the progression of AD dementia as prominent regions of hypometabolism[31,34,40,41]. The inclusion of frontal regions is also congruent with the observation that frontal hypometabolism is associated with age-related hypometabolism in general and with memory deficits in people with AD[40,42]. It is reasonable that preserved metabolism within these brain regions is associated with forecasting cognitive stability at the MCI stage.

Conversely, there is significant and surprising heterogeneity in the blue regions. These indicate that decreasing metabolism (decreasing voxel values) is positively correlated with SMCI classification, and so are hypermetabolically associated with cognitive decline. These regions include the cerebellum and various clusters within the frontal and occipital lobes, the thalamus, and the superior temporal gyri. While the hypermetabolic nature of the cerebellum in AD dementia has been well-confirmed, the presence of extracerebellar hypermetabolism in AD receives much less attention than the well-known hypometabolism associated with the disease[33,43]. Regional hypermetabolism has been shown within the literature in MCI and AD subject groups, but these regions are usually reported in isolation with one another[41,44–48]. The reported regions span the thalamus, frontal and the parietal lobes, and the superior temporal gyri – regions observed to be hypermetabolic in our visualizations. The hypermetabolic occipital lobe is particularly interesting, however, two studies found only a relative increase within their MCI cohorts[47,48]. A possible explanation is the association of regional deposition of amyloid and glycolytic hypermetabolism during the early stages of accumulation, however, the mechanisms for this effect remain unknown[45–47].

In brief, we interpret these results (Table 3) as a set of brain regions whose metabolic gradient are predictive of cognitive trajectory in patients with biologically confirmed prodromal AD over the medium term (less than 3 years). It should be noted that these regions are not necessarily prescriptive, especially since there is substantial variability in presentation and progression at the MCI stage, however, the regions identified do align well with previous research.

We do note there are some asymmetries in the laterality of the visualizations computed through GBP. The red clusters (preserved metabolism associated with cognitive stability) show some significant leanings toward the left-side of the brain. This observed asymmetry is consistent with previously reported findings on selective vulnerability of the left side of the brain to AD pathological progression, specifically with respect to reduced glucose metabolism caused by accelerated and asymmetric cortical thinning beyond that of normal aging[49,50].

An interesting result is the dual metabolic pattern found within the putamen where the anterior and posterior portions are red/hypometabolic and blue/hypermetabolic, respectively. The anterior putamen has dense connectivity with the frontal and prefrontal areas, especially the motor regions, while the posterior portions have connectivity with sensory regions[51]. The GBP maps do corroborate an inverse metabolic relationship between motor/precentral regions and the anterior putamen, and sensory/postcentral regions and the posterior putamen. The existence of a posterior-anterior gradient of putaminal dopaminergic neurodegeneration has been well-documented in Parkinson's disease yet the relevance of putaminal degeneration in AD dementia progression is unclear, though the putamen itself is known to be involved in certain cognitive roles[51]. While it is possible that this effect is caused by disturbances in anatomical standardization to a template space caused by enlarged ventricles, the use of a dementia-specific normalization template will likely account for AD-specific variances in anatomy.

Finally, the metabolism of the hippocampal and entorhinal regions may have been hypothesized from the beginning as being of great importance in forecast cognition, however, these regions are decidedly absent from the GBP maps. Previous research by Yee et al. which used GBP to realize the decision boundary of their neural network show that the hippocampus, angular gyrus, fusiform and pons are important regions in stratifying SHC and AD groups[34]. Their visualizations did agree on many other regions, such as the thalamus, putamen, lingual gyrus and cerebellum (all hypermetabolic) and cingulate gyrus (hypometabolic). This work did not stratify subjects by AD-related biomarkers and was specifically focused on differentiating demented and healthy patients; therefore, their results may not be sensitive to the transition from MCI to dementia among patients with biologically confirmed AD.

Since our GBP maps were produced exclusively from subjects with an A + T + biomarker profile and considering our emphasis on predicting cognitive stability (those more likely to be earlier in disease progression),

it is possible that we have visualized the compensatory mechanisms of brain metabolism in response to early AD, or else the modulatory effects of the same. The strict sparsity of the corrected contrast map (Fig. 1) does stand in contrast to the GBP heatmaps, however, from the uncorrected image it would appear the difference between these may be related to significance thresholding in SPM12. We must, however, urge some caution, since it is known that GBP may be influenced by discontinuities in network gradient flow, however, our procedure of averaging the maps over the entire ensemble would help alleviate this issue by balancing contributions from each subnetwork. Furthermore, the model's architecture contains many design elements to prevent gradient discontinuities.

We were able to discern several distinct areas that require attention or improvement within our study. First and foremost is the very modest sensitivity of 73.7%, which introduces concerns for false negatives if MAD2 were used as a screening tool. While this is suboptimal, this sensitivity is within the range commonly seen in literature (see Table 2 and Perron et al.)[24]. Follow-up monitoring with a physician is both appropriate and vital for this use case and should take clinical priority over the forecasts of the model if a patient labeled as stable experiences further cognitive decline, especially if this decline were sudden. We stress that the value of MAD2 as a screening tool derives from its exceptionally high PPV. The low probability of false positives occurring protects patients from undergoing unnecessary, expensive and potentially injurious treatment. This also frees healthcare systems from the significant burden of treatment for these patients[26]. Furthermore, it would also be possible to significantly enrich pharmaceutical trials by having a reliable means of forecasting cognition over a clinical meaningful timespan.

The ADNI dataset is widely considered to be unrepresentative of overall North American demographics. A majority of ADNI subjects are non-Hispanic Whites that have significant post-secondary education. This oversampling issue is expected to be addressed with ADNI4 specifically recruiting participants from populations underrepresented in dementia research[52,53]. Before this data becomes available, a possible solution would be the targeted usage of additional data augmentation strategies to oversample these underrepresented groups. To elucidate possible race bias in our work, a simple post-hoc inspection of classification accuracy by race was performed. All subjects who passed all inclusion and exclusion criteria were either White ($n = 537$) or Black ($n = 57$). Of the White subjects, only 7 identified as Hispanic/Latino. Overall, White (all categories) and Black subjects were classified with similar accuracies of 86.41% (464/537) and 85.96% (49/57), respectively.

We were forced to severely restrict the number of baseline healthy control subjects who later progressed to either MCI or dementia in our study due to a lack of FDG PET studies meeting inclusion criteria. The large number of subjects excluded due to a lack of available biomarker measures is another issue, however, this is currently under investigation for a new project on the most current ADNI data available at the time of writing. Some of the sample sizes available were quite uneven and this limits our ability to provide generalizable conclusions about these groups, especially regarding biomarker profiles.

One positive note is the lack of distinction between A + T- and A + T + profiles, specifically within PMCI and DMNT cohorts, is in line with guidelines recently published by Jack et al.[4]. It is possible that greater performance would have been achieved in a cognition-only stratification, however, only those who have biological AD or are at least amyloid-positive are relevant to our goal of determining candidacy for anti-amyloid therapy.

Covariate data were also not used in training. These could be incorporated into future versions of our model with careful consideration of the universality of such additions. A balance between the cost and availability of diagnostic information is necessary and so it was decided to eschew all data aside from the primary FDG PET neuroimaging studies because (i) FDG PET is the primary modality used for differential diagnosis of dementia, and (ii) it is the modality most effectively supplemented by covariates in a prognostic model[29,31,54].

Finally, though we emphasize using multiple independent test sets, we acknowledge the necessity of validation on a fully external dataset outside of ADNI and efforts in finding a suitable database are currently underway for a dedicated project to validate MAD2.

Materials and methods
Study approval
For this study we downloaded images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (ClinicalTrials.gov Identifier: NCT00106899; adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

Ethics approval for this study was granted by the Health Research Ethics Committee at the University of Manitoba. All research was performed in accordance with approved procedures and with principles of the Declaration of Helsinki. All human subjects provided informed consent for usage of their data and samples.

Data stratification
In our retrospective cross-sectional study, a total of 594 subjects (data accessed December 2020) who underwent FDG PET studies for whom biomarker data were available and had a minimum of 3 years of clinical follow-up were selected for use in this study. See Table 4 for baseline demographics. Clinical baseline diagnosis was used as a starting point and longitudinal progression used to determine cognitive cohort. Subjects were classified as described in Popuri et al. with some modifications[32]. First, subjects were classified by cognitive status with clinical dementia rating (CDR) into one of the following groups: stable healthy controls (SHC), unstable healthy controls (UHC), progressive healthy controls (PHC), stable mild cognitive impairment (SMCI), progressive mild cognitive impairment (PMCI) and dementia (DMNT). Healthy controls all enrolled with baseline CDR of 0 but differed by longitudinal trajectory. SHCs had a longitudinal CDR of 0, UHC was defined as those with

Cognitive cohort	Number	Sex	Age	MMSE	Education	Race	Biomarker profiles
SHC	129	M = 61￼F = 68	73.24 ± 5.83	29.03 ± 0.82	16.49 ± 2.71	W = 122￼NW = 7	68 (A-T-)￼30 (A + T-)￼31 (A + T+)
UHC	29	M = 18￼F = 11	71.64 ± 8.49	28.64 ± 0.96	16.28 ± 2.69	W = 25￼NW = 4	18 (A-T-)￼2 (A + T-)￼9 (A + T+)
SMCI	156	M = 103￼F = 63	72.42 ± 7.04	27.73 ± 1.59	16.31￼± 2.83	W = 147￼NW = 9	49 (A-T-)￼30 (A + T-)￼77 (A + T+)
PMCI	114	M = 69￼F = 45	72.67 ± 6.96	24.20 ± 2.74	16.13 ± 2.57	W = 113￼NW = 1	4 (A-T-)￼14 (A + T-)￼96 (A + T+)
DMNT	166	M = 103￼F = 73	74.50 ± 8.32	21.91 ± 2.96	15.53 ± 2.76	W = 16￼NW = 10	6 (A-T-)￼18 (A + T-)￼142 (A + T+)

Table 4. Demographics and sample sizes. Demographic values reported are means with standard deviation. M = male, F = female, W = white (including Hispanic/Latino), NW = non-white. A = amyloid and T = tau.

longitudinal CDR = 0.5 and PHC was defined as longitudinal CDR > = 1. SMCI and PMCI had CDR = 0.5 at baseline with PMCI eventually reaching CDR > = 1. DMNT enrolled with baseline CDR > = 1 and remained there longitudinally.

Only those PMCI subjects who progressed within 3 years were included for this study, and only SMCI subjects who were cognitively stable for a minimum of 3 years were included. UHC subjects who remained stable in MCI for at least 3 years were reclassified as SMCI subjects if FDG imaging was available within 3 months of MCI diagnosis.

Subjects were then stratified within their cognitive cohort by biomarker status according to the AT(N) classification system with CSF assays[55]. For amyloid-beta (Aβ), a cut-off value of < 980 pg ml^{-1} was decided upon, along with a cut-off value of > 21.6 pg ml^{-1} for phosphorylated-tau (p-tau)[56,57]. Any subjects for whom biomarker data were not available were not included in the present study. Labels were then combined, such as PMCI$_{A+T+}$ or SHC$_{A-T-}$ or DMNT$_{A+T+}$. For an overview of the stratification procedure see Fig. 3.

Exclusion criteria

Subjects who underwent reversion were included in the dataset so long as their status eventually returned to baseline status (e.g. MCI to SHC to MCI) to account for the variability of clinical presentation. Subjects with A-T + pathology were excluded from consideration since these are commonly recognized as non-Alzheimer pathologies. All PHC subjects were excluded owing to their sparsity of usable data fulfilling inclusion criteria.

Data processing and training strategy

All FDG PET images underwent preprocessing operations in MATLAB 2022b (MathWorks Inc, Sherborn, MA, USA) using SPM12 (Wellcome Centre for Human Neuroimaging, London, UK). All dynamics study frames were aligned to their average before summation. The final three frames of each dynamic study were point-wise summed to create static studies mimicking clinical norm of a single 15-minute imaging session ("back"). Additional static images were created from the first three frames ("front") and all six frames ("whole") for the purpose of increasing training sample size. The summed images were normalized to a dementia specific FDG PET template to ensure spatial congruence[58]. Images were also normalized to the canonical ^{15}O perfusion PET template in SPM12 to further increase training sample size. Each FDG PET had 6 unique data representations (^{15}O: front, back, whole; ^{18}F: front, back, whole). These templates were binarized and used to remove any voxels outside the brain. All images were smoothed with scanner-specific kernels and intensity scaled by global cerebral mean[59]. Images were visually inspected at each step for quality control and processing errors corrected.

For both M-1 and M-2, the baseline imaging data of all PMCI$_{A+T+}$ and SMCI$_{A+T+}$ subjects were divided into testing and training sets at the level of the subject to prevent data leakage. All subjects were randomly assigned a train/test split index (80% training, 20% testing) and checked for imbalances (age, sex, race/ethnicity, years of education, APOE genetics and mini-mental status exam score). The training set was split into cross-validation folds and checked for imbalances as described above. Longitudinal imaging sessions were included in the training splits but not validation or testing splits. FDG PET data from SHC$_{A+T+}$ and DMNT$_{A+T+}$ subjects was also used to supplement the training data as a further form of data augmentation for M-2 but not for M-1. Validation folds exclusively contained baseline images of MCI patients normalized to the ^{18}F dementia-specific template and summed from the final 3 frames of the study to replicate clinical study duration. Subjects with healthy and AD continuum biomarker profiles (e.g. SMCI$_{A+T-}$, UHC$_{A-T-}$, SHC$_{A-T-}$) were used to create independent test sets consisting solely of baseline images without data augmentation. All datasets (training, validation and testing) were checked for leakage at the level of the subject.

Every UHC$_{A+T+}$ subject meeting criteria for SMCI was considered as clinically equivalent and was used to supplement the prodromal test set if FDG imaging was available within 3 months of MCI baseline. The independent test set therefore included 19 PMCI$_{A+T+}$ subjects and 25 (16 + 9) SMCI$_{A+T+}$ subjects. The FDG PET studies post dementia diagnosis from PMCI$_{A+T+}$ subjects were also used to supplement the augmented dataset to
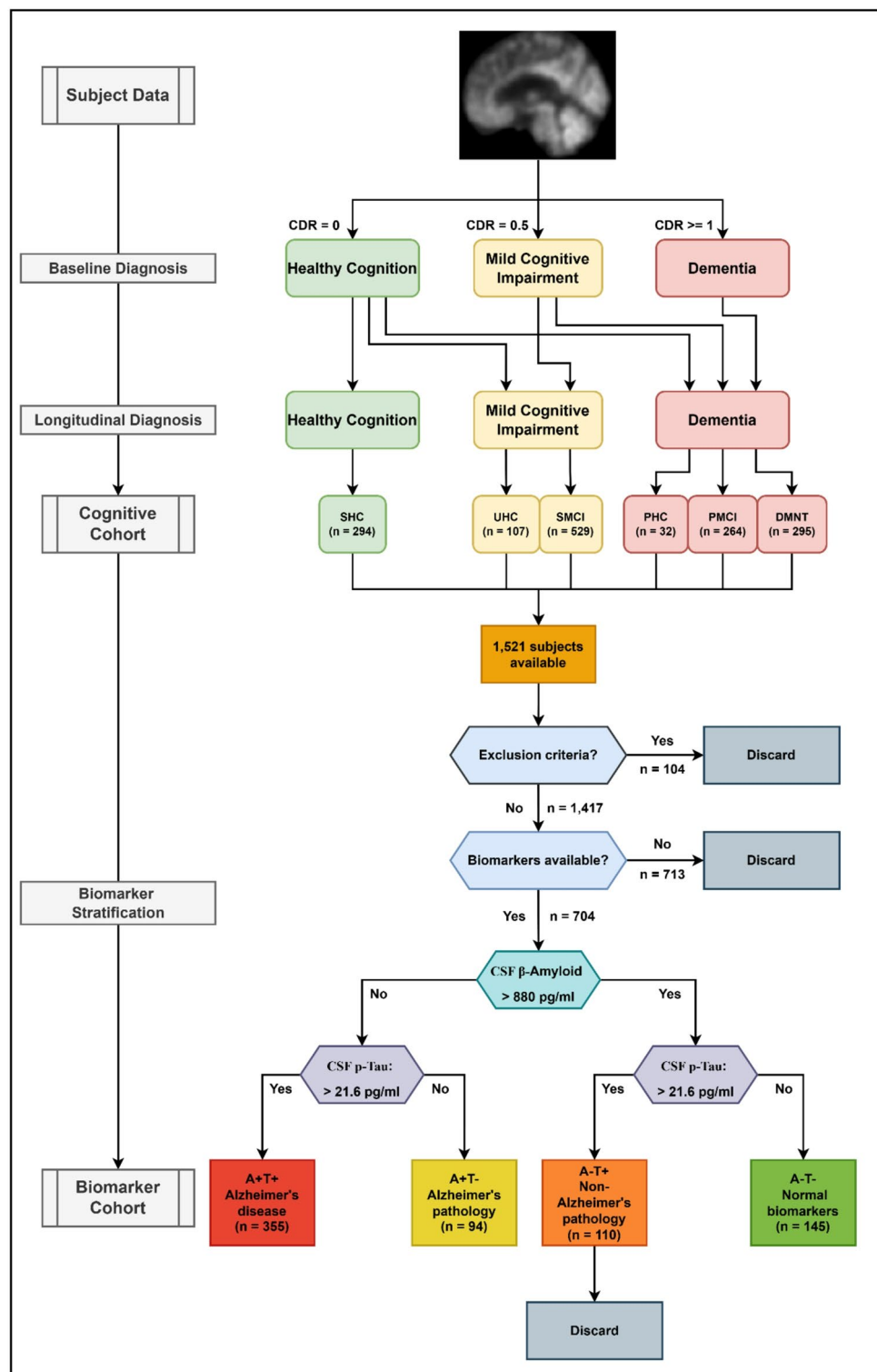
Fig. 3. Flowchart of cognitive-biomarker stratification for ADNI cohort. Cognitive stratification followed by biomarker groupings allowed for the population of cognitive-biomarker cohorts that were used as a primary variable-of-interest within the present work. All subjects with A-T+biomarker profiles were excluded from our study since these likely represent non-Alzheimer's pathology. A total of 927 subjects were excluded from any further analysis in our selection procedure, leaving 594 subjects suitable for our purposes.

increase the number of AD subjects where available. Post dementia diagnosis studies from $PMCI_{A+T+}$ pre-MCI UHC_{A+T+} test set subjects were discarded.

Statistical parametric mapping

A group-wise two-sample t-test was performed prior to machine learning-based investigation. All A + T + subjects at the MCI stage were included. The control group was composed of SMCI and UHC subjects at MCI stage while the active group was composed exclusively of PMCI subjects. Only baseline images were used. Corrected t-maps were FWE adjusted for $p < 0.05$ and an explicit grey matter mask was used. Uncorrected results ($p < 0.001$) were also visualized for reference.

Machine learning architecture

The machine learning model used for the present study was a deep learning architecture based on DenseNet-BC[60]. The most notable property of the DenseNet architecture is the connectivity between every layer within a convolutional block. The network functions as a feature aggregator where each convolutional layer creates features which are depth-wise concatenated along a vector of feature maps from the previous layer within a dense block. Each convolutional layer may therefore access the features output by each previous convolutional layer within a block. This encourages feature reuse, strengthens feature propagation between layers and helps alleviate vanishing gradients through identity mappings[60,61]. Due to this inherent design of feature aggregation over a deep network, we estimated that a densely connected architecture could aid in finding useful features for predicting cognition in patients with prodromal AD. Comparisons of deep learning architectures for neuroimaging applications have been performed, and we recommend Bloch et al.[62].

We implemented a growth rate parameter which increases exponentially between each dense block to encourage use of high-level features. Every subsequent dense block accumulates twice the number of features between convolutional layers as the block before it. Bottlenecking layers and compression layers were implemented in the transition blocks. Each convolutional layer within a dense block was preactivated with LeakyReLU and instance normalization to ensure convolutional layers always operate on appropriate distributed data[61,63]. The final dense block outputs a large series of feature maps into a classification block for global average pooling and convolution into a softmax classifier, eschewing fully connected layers to reduce the number of parameters within the network[64]. Dropout layers were used after transition blocks and immediately after global average pooling in the classification block[65]. The primary advantage in using dropout layers is that they disallow coadaptation of features and prevent overfitting at the cost of some increased training time. See Fig. 4 for a graphical representation of the network architecture.
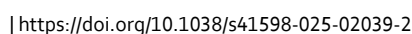
Network loss was calculated by class-weighted cross-entropy with L_2 penalty to encourage smaller network weights and reduce overfitting[66]. Each cross-validated model was trained by backpropagation with Adaptive Moment Estimation (ADAM) using an initial learning rate of 10^{-3}. The learning rate decreased by a multiplicative factor of 0.9 every epoch. L_2 constant was set to 10^{-2} and epsilon was set to 10^{-8}. Gradient decay and squared gradient decay were 0.9 and 0.999 respectively while the gradient threshold was set to the L_2 norm. The network was validated twice per epoch for a maximum of 50 epochs, and training was stopped if validation loss failed to decrease over 6 epochs. The network iteration with the lowest validation loss was taken as the output of a training cycle. All output network loss curves were visually inspected to ensure good training. Default softmax probability threshold of 0.5 was used to determine class label. Threshold sensitivity analysis was performed.

Statistical analysis

The group classification by machine learning models were compared by area-under-curve (AUC), accuracy (ACC), sensitivity (SEN) and specificity (SPE). Both positive and negative predictive values (PPV/NPV) are also discussed for prodromal subjects. The outputs (class probability scores) of the best performing model were tested by analysis of variance (ANOVA) with Bonferroni post-hoc tests.

Post-hoc visualization of machine class activations

Visualization of neural network classification was performed with guided backpropagation (GBP)[67]. GBP is a modification of the standard backpropagation algorithm such that after performing a forward pass through the network for a prediction, any negative gradients are set to zero, ensuring that only positive gradients (and therefore positive contributions towards a class) are allowed to backpropagate. The result is a visualization of those input features (voxels) that affect a neural network's output, offering insight into the features learned by the model. Each prodromal MCI subjects was used to compute an average map of the activations for the SMCI class. The Automated Anatomical Labeling atlas (AAL3) found within CAT12 (http://www.neuro.uni-jena.de/cat/) was used to determine regional anatomy within the clusters identified by GBP[68].

Fig. 4. Network architecture and ensemble construction. Data were split 80 – 20 into training and testing sets at the level of the subjects, balanced for multiple demographic and biological characteristics. The training data underwent 5-fold cross-validation (n_{train} = 4,092). The resulting models were ensembled and then applied to the independent test data (n_{test} = 283).

Data availability

Original code and prepared data are available upon reasonable request to the corresponding author, Dr Ji Hyun Ko.

References

1. Cui, Y., Yang, W., Shuai, J., Ma, Y. & Yan, Y. Lifestyle and socioeconomic transition and health consequences of Alzheimer's disease and other dementias in global, from 1990 to 2019. *J. Prev. Alzheimers Dis.* **11**, 88–96. https://doi.org/10.14283/jpad.2023.63 (2024).
2. Nandi, A. et al. Cost of care for Alzheimer's disease and related dementias in the united States: 2016 to 2060. *Npj Aging.* **10** https://doi.org/10.1038/s41514-024-00136-6 (2024).
3. Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dement.* **3**, 186–191. https://doi.org/10.1016/j.jalz.2007.04.381 (2007).
4. Jack, C. R. et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's association workgroup. *Alzheimer's Dement.* https://doi.org/10.1002/alz.13859 (2024).
5. Therriault, J. et al. Biomarker-based staging of alzheimer disease: rationale and clinical applications. *Nat. Reviews Neurol.* **20**, 232–244. https://doi.org/10.1038/s41582-024-00942-2 (2024).
6. Arvanitakis, Z., Shah, R. C. & Bennett, D. A. Diagnosis and management of dementia: review. *JAMA: J. Am. Med. Association.* **322**, 1589–1599. https://doi.org/10.1001/jama.2019.4782 (2019).
7. Mahase, E. Alzheimer's disease: Lecanemab gets full FDA approval and black box safety warning. *BMJ (Online).* **382**, 1580–1580. https://doi.org/10.1136/bmj.p1580 (2023).
8. Manly, J. J. & Deters, K. D. Donanemab for alzheimer Disease—Who benefits and who is harmed?? *JAMA* **330**, 510. https://doi.org/10.1001/jama.2023.11704 (2023).
9. van Dyck, C. H. et al. Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* **388**, 9–21. https://doi.org/10.1056/NEJMoa2212948 (2023).
10. Knopman, D. S. & Hershey, L. Implications of the approval of Lecanemab for alzheimer disease patient care: incremental step or paradigm shift?? *Neurology* **101**, 610–620. https://doi.org/10.1212/WNL.0000000000207438 (2023).
11. Sims, J. R. et al. Donanemab in early symptomatic alzheimer disease. *JAMA* **330**, 512. https://doi.org/10.1001/jama.2023.13239 (2023).
12. Cummings, J. et al. Aducanumab: appropriate use recommendations. *J. Prev. Alzheimer's Disease.* **8**, 398–410. https://doi.org/10.14283/jpad.2021.41 (2021).

13. Liu, K. Y. et al. Key questions for the evaluation of anti-amyloid immunotherapies for Alzheimer's disease. *Brain Commun.* **5**, fcad175. https://doi.org/10.1093/braincomms/fcad175 (2023).
14. Digma, L. A., Winer, J. R. & Greicius, M. D. Substantial doubt remains about the efficacy of Anti-Amyloid antibodies. *J. Alzheimers Dis.* **97**, 567–572. https://doi.org/10.3233/JAD-231198 (2024).
15. Dantas, J. M. et al. Efficacy of anti-amyloid-ß monoclonal antibody therapy in early Alzheimer's disease: a systematic review and meta-analysis. *Neurol. Sci.* **45**, 2461–2469. https://doi.org/10.1007/s10072-023-07194-w (2024).
16. Bradshaw, A. C. & Georges, J. Anti-Amyloid therapies for alzheimer's disease: an alzheimer Europe position paper and call to action. *J. Prev. Alzheimers Dis.* **11**, 265–273. https://doi.org/10.14283/jpad.2024.37 (2024).
17. Beach, T. G., Monsell, S. E., Phillips, L. E. & Kukull, W. Accuracy of the clinical diagnosis of alzheimer disease at National Institute on aging alzheimer disease centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* **71**, 266–273. https://doi.org/10.1097/NEN.0b013e318 24b21b (2012).
18. Brunnström, H. M. D. & Englund, E. M. D. P. D. Clinicopathological concordance in dementia diagnostics. *Am. J. Geriatric Psychiatry.* **17**, 664–670. https://doi.org/10.1097/JGP.0b013e3181a6516e (2009).
19. Scheltens, P. & Rockwood, K. How golden is the gold standard of neuropathology in dementia? *Alzheimer's Dement.* **7**, 486–489. https://doi.org/10.1016/j.jalz.2011.04.011 (2011).
20. Tangalos, E. G. M. D. & Petersen, R. C. M. D. P. Mild cognitive impairment in geriatrics. *Clin. Geriatr. Med.* **34**, 563–589. https://doi.org/10.1016/j.cger.2018.06.005 (2018).
21. Thaipisuttikul, P., Jaikla, K., Satthong, S. & Wisajun, P. Rate of conversion from mild cognitive impairment to dementia in a Thai hospital-based population: A retrospective cohort. *Alzheimer's Dementia: Translational Res. Clin. Interventions.* **8**, e12272. https://doi.org/10.1002/trc2.12272 (2022).
22. Ward, A., Tardiff, S., Dye, C. & Arrighi, H. M. Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: A systematic review of the literature. *Dement. Geriatric Cogn. Disorders Extra.* **3**, 320–332. https://doi.org/10.1159/000354370 (2013).
23. Shimada, H., Doi, T., Lee, S. & Makizako, H. Reversible predictors of reversion from mild cognitive impairment to normal cognition: a 4-year longitudinal study. *Alzheimers Res. Ther.* **11**, 24–24. https://doi.org/10.1186/s13195-019-0480-5 (2019).
24. Perron, J. & H Ko, J. Review of quantitative methods for the detection of Alzheimer's disease with positron emission tomography. *Appl. Sci.* **12**, 11463. https://doi.org/10.3390/app122211463 (2022).
25. Villain, N. & Planche, V. Disentangling clinical and biological trajectories of neurodegenerative diseases. *Nat. Reviews Neurol.* https://doi.org/10.1038/s41582-024-01004-3 (2024).
26. Perron, J., Scramstad, C. & Ko, J. H. Analysis of costs for Imaging-Assisted pharmaceutical intervention in Alzheimer's disease with Lecanemab: snapshot of the first 3 years. *J. Alzheimers Dis.* **96**, 1305–1315. https://doi.org/10.3233/JAD-230633 (2023).
27. Graham, M. M. The future of nuclear medicine in the united States. *J. Nucl. Med.* **64**, 1352–1353. https://doi.org/10.2967/jnumed.122.265314 (2023).
28. Verger, A. et al. FDA approval of Lecanemab: the real start of widespread amyloid PET use? — the EANM neuroimaging committee perspective. *Eur. J. Nucl. Med. Mol. Imaging.* **50**, 1553–1555. https://doi.org/10.1007/s00259-023-06177-5 (2023).
29. Juengling, F. et al. PET imaging in dementia: Mini-Review and Canadian perspective for clinical use. *Can. J. Neurol. Sci.* 1–13. https://doi.org/10.1017/cjn.2024.31 (2024).
30. Zukotynski, K. et al. PET/CT of Dementia. *Am. J. Roentgenol.* **211**, 246–259. https://doi.org/10.2214/AJR.18.19822 (2018).
31. Minoshima, S., Mosci, K., Cross, D. & Thientunyakit, T. Brain [F-18]FDG PET for clinical dementia workup: differential diagnosis of Alzheimer's disease and other types of dementing disorders. *Semin. Nucl. Med.* **51**, 230–240. https://doi.org/10.1053/j.semnuclmed.2021.01.002 (2021).
32. Popuri, K. et al. Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. *NeuroImage Clin.* **18**, 802–813. https://doi.org/10.1016/j.nicl.2018.03.007 (2018).
33. Katako, A. et al. Machine learning identified an Alzheimer's disease-related FDG-PET pattern which is also expressed in lewy body dementia and Parkinson's disease dementia. *Sci. Rep.* **8**, 13236–13213. https://doi.org/10.1038/s41598-018-31653-6 (2018).
34. Yee, E., Popuri, K. & Beg, M. F. Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer's dementia score. *Hum. Brain. Mapp.* **41**, 5–16. https://doi.org/10.1002/hbm.24783 (2020).
35. Pan, X. et al. Multi-View separable pyramid network for AD prediction at MCI stage by 18 F-FDG brain PET imaging. *IEEE Trans. Med. Imaging.* **40**, 81–92. https://doi.org/10.1109/TMI.2020.3022591 (2021).
36. Ma, D. et al. Blinded clinical evaluation for dementia of Alzheimer's type classification using FDG-PET: A comparison between Feature-Engineered and Non-Feature-Engineered machine learning methods. *J. Alzheimers Dis.* **80**, 715–726. https://doi.org/10.3233/JAD-201591 (2021).
37. Lu, D., Popuri, K., Ding, G. W., Balachandar, R. & Beg, M. F. Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease. *Med. Image. Anal.* **46**, 26–34. https://doi.org/10.1016/j.media.2018.02.002 (2018).
38. Heyer, S. et al. F-FDG PET can effectively rule out conversion to dementia and the presence of CSF biomarker of neurodegeneration: a real-world data analysis. *Alzheimers Res. Ther.* **16**, 182. https://doi.org/10.1186/s13195-024-01535-3 (2024).
39. Ou, Y. N. et al. FDG-PET as an independent biomarker for Alzheimer's biological diagnosis: a longitudinal study. *Alzheimers Res. Ther.* **11**, 57. https://doi.org/10.1186/s13195-019-0512-1 (2019).
40. Habeck, C. et al. Relationship between baseline brain metabolism measured using [18F]FDG PET and memory and executive function in prodromal and early Alzheimer's disease. *Brain Imaging Behav.* **6**, 568–583. https://doi.org/10.1007/s11682-012-9208-x (2012).
41. Kato, T., Inui, Y., Nakamura, A. & Ito, K. Brain Fluorodeoxyglucose (FDG) PET in dementia. *Ageing Res. Rev.* **30**, 73–84. https://doi.org/10.1016/j.arr.2016.02.003 (2016).
42. Knopman, D. S. et al. 18F-fluorodeoxyglucose positron emission tomography, aging, and Apolipoprotein E genotype in cognitively normal persons. *Neurobiol. Aging.* **35**, 2096–2106. https://doi.org/10.1016/j.neurobiolaging.2014.03.006 (2014).
43. Gupta, V., Booth, S. & Ko, J. H. Hypermetabolic cerebellar connectome in Alzheimer's disease. *Brain Connect.* **13**, 356–366. https://doi.org/10.1089/brain.2020.0937 (2023).
44. Johnson, S. C. et al. Amyloid burden and neural function in people at risk for Alzheimer's disease. *Neurobiol. Aging.* **35**, 576–584. https://doi.org/10.1016/j.neurobiolaging.2013.09.028 (2014).
45. Oh, H., Habeck, C., Madison, C. & Jagust, W. Covarying alterations in Aβ deposition, glucose metabolism, and Gray matter volume in cognitively normal elderly. *Hum. Brain. Mapp.* **35**, 297–308. https://doi.org/10.1002/hbm.22173 (2014).
46. Bakhtiari, A. et al. Early cerebral amyloid-β accumulation and hypermetabolism are associated with subtle cognitive deficits before accelerated cerebral atrophy. *GeroScience* **46**, 769–782. https://doi.org/10.1007/s11357-023-01031-w (2023).
47. Kemppainen, N. et al. Long-Term interrelationship between brain metabolism and amyloid deposition in mild cognitive impairment. *J. Alzheimers Dis.* **48**, 123–133. https://doi.org/10.3233/JAD-150190 (2015).
48. Ashraf, A., Fan, Z., Brooks, D. J. & Edison, P. Cortical hypermetabolism in MCI subjects: a compensatory mechanism? *Eur. J. Nucl. Med. Mol. Imaging.* **42**, 447–458. https://doi.org/10.1007/s00259-014-2919-z (2015).
49. Roe, J. M. et al. Asymmetric thinning of the cerebral cortex across the adult lifespan is accelerated in Alzheimer's disease. *Nat. Commun.* **12**, 721. https://doi.org/10.1038/s41467-021-21057-y (2021).
50. Lubben, N., Ensink, E., Coetzee, G. A. & Labrie, V. The enigma and implications of brain hemispheric asymmetry in neurodegenerative diseases. *Brain Commun.* **3**, fcab211. https://doi.org/10.1093/braincomms/fcab211 (2021).
51. Costa, A. & Villalba, E. *Horizons in Neuroscience Research* (Nova Science Publishers, Inc, 2012).

52. Weiner, M. W. et al. Increasing participant diversity in AD research: plans for digital screening, blood testing, and a community-engaged approach in the Alzheimer's disease neuroimaging initiative 4. *Alzheimer's Dement.* **19**, 307–317. https://doi.org/10.1002/alz.12797 (2023).

53. Mitchell, A. K. et al. Analysis of eligibility criteria in Alzheimer's and related dementias clinical trials. *Sci. Rep.* **14** https://doi.org/10.1038/s41598-024-65767-x (2024).

54. Shaffer, J. L. et al. Predicting cognitive decline in subjects at risk for alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology* **266**, 583–591. https://doi.org/10.1148/radiol.12120010 (2013).

55. Jack, C. R. et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* **14**, 535–562. https://doi.org/10.1016/j.jalz.2018.02.018 (2018).

56. Deatsch, A., Perovnik, M., Namías, M., Trošt, M. & Jeraj, R. Development of a deep learning network for Alzheimer's disease classification with evaluation of imaging modality and longitudinal data. *Phys. Med. Biol.* **67**, 195014. https://doi.org/10.1088/1361-6560/ac8f10 (2022).

57. Blazhenets, G. et al. Validation of the alzheimer disease dementia Conversion-Related pattern as an ATN biomarker of neurodegeneration. *Neurology* **96**, e1358–e1368. https://doi.org/10.1212/WNL.0000000000011521 (2021).

58. Della Rosa, P. A. et al. A standardized [18F]-FDG-PET template for Spatial normalization in statistical parametric mapping of dementia. *Neuroinformatics (Totowa N J).* **12**, 575–593. https://doi.org/10.1007/s12021-014-9235-4 (2014).

59. Joshi, A., Koeppe, R. A. & Fessler, J. A. Reducing between scanner differences in multi-center PET studies. *NeuroImage (Orlando Fla).* **46**, 154–159. https://doi.org/10.1016/j.neuroimage.2009.01.057 (2009).

60. Huang, G., Liu, Z., van der Laurens, M. & Weinberger, K. Q. Densely Connected Convolutional Networks. *arXiv.org*, (2018). https://doi.org/10.48550/arxiv.1608.06993

61. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. *ArXiv Org.* https://doi.org/10.48550/arxiv.1603.05027 (2016).

62. Bloch, L., Friedrich, C. M. & Initiative, A. s. D. N. Systematic comparison of 3D deep learning and classical machine learning explanations for Alzheimer's disease detection. *Comput. Biol. Med.* **170**, 108029. https://doi.org/10.1016/j.compbiomed.2024.108029 (2024).

63. Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance normalization: the missing ingredient for fast stylization. *ArXiv Org.* https://doi.org/10.48550/arxiv.1607.08022 (2017).

64. Lin, M., Chen, Q. & Yan, S. Network in network. https://doi.org/10.48550/arxiv.1312.4400 (2013).

65. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv Org.* https://doi.org/10.48550/arxiv.1207.0580 (2012).

66. Kukačka, J., Golkov, V. & Cremers, D. Regularization for deep learning: A taxonomy. *ArXiv Org.* https://doi.org/10.48550/arxiv.1710.10686 (2017).

67. Springenberg, J. T., Dosovitskiy, A., Brox, T. & Riedmiller, M. Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806 (2014).

68. Rolls, E. T., Huang, C. C., Lin, C. P., Feng, J. & Joliot, M. Automated anatomical labelling atlas 3. *Neuroimage* **206**, 116189 (2020).

Acknowledgements

Author contributions

Conceptualization, J.P. and J.H.K.; methodology, J.P.; investigation, J.P., C.S.; resources, J.P. and J.H.K, writing-original draft preparation, J.P.; writing-review and editing, C.S., J.P. and J.H.K.; visualization, J.P.; supervision, C.S. and J.H.K.; funding acquisition, J.P. and J.H.K. All authors have read and agreed to the published version of the manuscript.

Funding

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-02039-2.

Correspondence and requests for materials should be addressed to J.H.K.

Reprints and permissions information is available at www.nature.com/reprints.