

RESEARCH ARTICLE

Neuroimaging modality fusion in Alzheimer's classification using convolutional neural networks

Arjun Punjabi^{1*}, Adam Martersteck^{2,3}, Yanran Wang¹, Todd B. Parrish², Aggelos K. Katsaggelos¹, and the Alzheimer's Disease Neuroimaging Initiative[¶]

1 Department of Electrical Engineering and Computer Science/McCormick School of Engineering, Northwestern University, Evanston, Illinois, United States of America, **2** Department of Radiology/Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States of America, **3** Mesulam Cognitive Neurology and Alzheimer's Disease Center, Northwestern University, Chicago, Illinois, United States of America

[¶] A link to the complete listing of ADNI investigators can be found in the Acknowledgments section.

* arjunpunjabi2015@u.northwestern.edu



OPEN ACCESS

Citation: Punjabi A, Martersteck A, Wang Y, Parrish TB, Katsaggelos AK, and the Alzheimer's Disease Neuroimaging Initiative (2019) Neuroimaging modality fusion in Alzheimer's classification using convolutional neural networks. *PLoS ONE* 14(12): e0225759. <https://doi.org/10.1371/journal.pone.0225759>

Editor: Stephen D Ginsberg, Nathan S Kline Institute, UNITED STATES

Received: February 19, 2019

Accepted: November 12, 2019

Published: December 5, 2019

Copyright: © 2019 Punjabi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data is public and may be found at the Alzheimer's Neuroimaging Initiative (ADNI) website: <http://adni.loni.usc.edu/>.

Funding: While the authors received no specific funding for this work, some of the authors were supported by the Integrated Data Driven Discovery in Earth and Astrophysical Sciences (IDEAS) program at Northwestern University (NSF Research Traineeship Grant 1450006), the Biomedical Data Driven Discovery (BD3) training

Abstract

Automated methods for Alzheimer's disease (AD) classification have the potential for great clinical benefits and may provide insight for combating the disease. Machine learning, and more specifically deep neural networks, have been shown to have great efficacy in this domain. These algorithms often use neurological imaging data such as MRI and FDG PET, but a comprehensive and balanced comparison of the MRI and amyloid PET modalities has not been performed. In order to accurately determine the relative strength of each imaging variant, this work performs a comparison study in the context of Alzheimer's dementia classification using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset with identical neural network architectures. Furthermore, this work analyzes the benefits of using both modalities in a fusion setting and discusses how these data types may be leveraged in future AD studies using deep learning.

Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by cognitive decline and dementia. The number of individuals living with AD in the United States is expected to reach 10 million by the year 2025 [1]. As a result, automated methods for computer aided diagnosis could greatly improve the ability to screen at-risk individuals.

Such methods typically take as input patient data including demographics, medical history, genetic sequencing, and neurological images among others. The resulting output is health status indicated by a diagnosis label, which may also include a probabilistic uncertainty on the prediction. This particular investigation will focus on two different neuroimaging modalities: structural T1-weighted MRI and AV-45 amyloid PET. The primary goal of the investigation is to compare the efficacy of each of these modalities in isolation as well as when both are used as simultaneous input to a fusion system. While other studies make use of T1-weighted MRI and

program at Northwestern University (NIH Grant 5T32LM012203-02), and the National Institute on Aging (T32AG020506). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

Competing interests: The authors have declared that no competing interests exist.

FDG PET, we believe that, to the best of our knowledge, this is the first comparison and fusion deep learning study using AV-45 amyloid PET. Because FDG and amyloid PET have different biological sources, their ability to aid in Alzheimer's diagnosis may greatly differ.

The algorithmic design of these methods can vary, but recent successes in machine learning have opened the floodgates for a plethora of deep neural networks trained for computer aided diagnosis. Given the visual nature of the input data, this work opted to apply a model well suited for computer vision tasks: the convolutional neural network (CNN). The following sections will focus on related approaches to the AD classification problem, the methodology of both the network and data pre-processing pipeline, and a discussion of the classification results.

Related work

Computer aided diagnosis methods in this domain have spanned the gamut of algorithmic design. Earlier methods often applied linear classifiers like support vector machines (SVM) to hand-crafted biological features [2]. These features can be defined at the individual voxel level, as in the case for tissue probability maps, or at the regional level, including cortical thickness and hippocampal shape or volume. The 2011 comparison performed in [2] found that whole brain methods generally achieved higher classification accuracy than their region-based counterparts. Additionally, there was evidence to suggest that certain data pre-processing methods, namely the DARTEL registration package [3], can substantially impact classification results. These two findings informed the decision to use whole brain volumes in this work and design a robust registration pipeline before the classification algorithm.

Similar linear classifier or SVM-based methods exist that align with these ideas. In [4], gray matter tissue maps were classified with an SVM. A more complex scheme exists in [5], where template selection was performed on gray matter density maps and these features were clustered in preparation for SVM classification. As previously discussed, regional features can also be used as input to an SVM, such as spherical harmonic coefficients calculated from the hippocampus [6]. In [7], the analysis is extended to other linear classifiers, primarily comparing the performance between SVMs and variations of random forest classifiers on a large conglomerate of Alzheimer's datasets. These models can also extend to multiple data modalities as in [8], where features from MRI and FDG PET data were extracted and combined with a kernel-based approach. In [9], the procedure was modified with a custom loss function in order to perform both diagnosis classification and cognitive score regression simultaneously using a modified support vector-based model trained with MRI, PET, and cerebrospinal fluid (CSF) images.

Despite the initial popularity of SVMs and linear classifiers, there has been a transition in the last several years toward more non-linear approaches. Namely, the introduction of artificial neural networks has transformed the landscape of automated Alzheimer's dementia diagnosis. However, even these methods have varied in construction. The works in [10, 11] used a deep Boltzmann machine (DBM) to extract features from MRI and FDG PET data which are then classified using an SVM. Similarly, a DBM was also used in [12] to extract features from MRI and FDG PET, but additionally included CSF and cognitive test scores. The features are still classified with an SVM. A more standard fully-connected neural network was trained on MRI images in [13], but performance was improved by adding spatial neighborhood regularization similar to the receptive field of convolutional kernels.

This leads to the current preferred machine learning model, the CNN. These models are well suited to tasks with 2D or 3D data due to the shared filter weights within each convolutional layer. A CNN was proposed in [14] that takes fMRI slices as input to a modified LeNet-5

CNN architecture [15]. The DeepAD paper [16] further developed this notion by utilizing the more complex GoogleNet CNN [17]. In [18], MRI and FDG PET data were used to train a multimodal CNN for classification, but it also allowed for missing modalities and modality completion. Some methods opted to use autoencoders [19] which can employ convolutional filters, but structurally differ from CNNs. While CNNs are trained to map input images to some given representation, autoencoders are trained to perform dimensionality reduction and reconstruct the input image. In this manner, the features learned in the middle layer of an autoencoder can be extracted and classified with either linear or non-linear methods. In [20], features from MRI and FDG PET images were extracted using a stacked autoencoder which were then classified with softmax regression. On the other hand, the work in [21] used an autoencoder on 2D MRI slices to learn basis features that are then used as CNN filters. A similar procedure was performed in [22] that compared the performance between both 2D and 3D systems. An autoencoder was used in [23] on full 3D MRI images to pre-train the layers of a CNN model, and this was expanded in [24] to include the FDG PET modality. The authors in [25] use a scheme of stacked polynomial networks on MRI and FDG PET data, and use similar cascaded network approaches in [26] and [27] when tackling Parkinson's diagnosis. Some of these results are shown in Table 1.

Fundamentally, while methods exist that take advantage of multiple data types and apply state-of-the-art neural network architectures, comparison studies between modalities have been haphazard in their use of datasets and lacking in explanations of model efficacy. In some instances, subsets of larger databases were used without explanations of why certain images were included or excluded. The deep learning comparisons that have been performed examine MRI and FDG PET scans, whereas none have addressed fusion of MRI with AV-45 PET scans. Because FDG PET measures metabolism whereas AV-45 PET measures beta amyloid (the buildup of which is a precursor to Alzheimer's disease), the modalities are drastically different in their information content [28]. Consequently, the added benefit to classification performance when combined with MRI data may differ as well. Additionally, pre-processing pipelines differ between these various studies. These factors contribute to incongruous modality comparison results between papers. Furthermore, the biological explanations for such discrepancies are often lacking or non-existent. This work is novel in these respects. First, the pre-processing used in this work is clearly explained and the rationale for each step is provided. Also, the modality comparison results are discussed within a biological context that more effectively describes the relative performance of each data type.

Methodology

As previously alluded to in the discussion of related work, pre-processing operations can have a major impact on final classification performance. As a result, a pipeline was developed to correct several of the biases inherent in the imaging data. While the components of the pipeline employ existing algorithms, the overall structure differs from previous work and allows for a more fair comparison between the T1-weighted MRI and AV-45 PET modalities.

Table 1. MRI and FDG PET fusion classification accuracies (%).

	MRI	FDG PET	Fusion
[11] DBM	92.38	92.20	95.35
[24] CNN	80.62	81.93	84.72
[24] SAE	85.24	85.53	91.14
[25] SDPN	95.44	95.11	97.13

<https://doi.org/10.1371/journal.pone.0225759.t001>

This section also discusses the neural network architecture. The design of the network is similar to the CNN-based approaches discussed previously. Again, because the primary goal of the investigation is a comparison of data modalities rather than network styles, the CNN was designed to be representative of comparable methods comprised of standard network layers.

Pre-processing

The pre-processing pipeline aimed to correct several biases that can exist in raw MRI and PET data. This also removes the additional burden of the network learning methods to correct or overlook these biases. Instead, the network has the isolated task of finding patterns between healthy and Alzheimer's patients. The vast majority of related work also employs similar pre-processing techniques in order to combat standard problems; namely, most methods perform some kind of MRI bias field correction, volumetric skull stripping, and affine registration. This approach is nascent in its registration scheme in order to prepare data for longitudinal studies in addition to traditional single time instance analyses. This manifests itself in two ways. First, our current investigation that treats each of these scanning instances as distinct samples in the dataset is less biased by differences in pre-processing for each modality. Second, when the scanning instances are viewed jointly as a single sample in the dataset for a longitudinal study, the images are normalized both within the subject and among all subjects in the set. Future longitudinal studies that take advantage of this processed data will be discussed at the end of the paper. The building blocks of the pipeline are as follows:

MRI bias field correction. MRI images can have a low frequency bias component as a result of transmit/receive inhomogeneities of the scanner [29]. This spatial non-uniformity, while not always visually apparent, can cause problems for image processing pipelines. As a result, many MRI processing schemes begin by applying a bias field correction algorithm. Non-parametric non-uniform intensity normalization (N3) [30] is a robust and well-established approach for removing this bias field. It optimizes for the slowly varying multiplicative field that, when removed, restores the high frequency components of the true signal. This work opted to employ a more recent update to this method known as N4 [31], which makes use of B-spline fitting for improved corrections. This step is performed on the raw MRI images and is unnecessary for the PET images.

Affine registration. Both image modalities are registered using a linear affine transformation. Registration aims to remove any spatial discrepancies between individuals in the scanner, namely minor translations and rotations from a standard orientation. Typically, scans are registered to a brain atlas template, such as MNI152 [32]. While this procedure is perfectly acceptable for traditional single time point analyses, this pipeline was designed to accommodate longitudinal studies as well. In such a setting, a patient in the dataset will have multiple scanning sessions at different times, but these images are aggregated and treated as a more complex representation of a single data point. As a result, it is beneficial to have congruence between the temporal scans in addition to registration to the standard template. Consequently, MRI and PET scans in the pipeline are registered first to an average template created from all MRI scans from a single patient, and then once more to the standard MNI152 space. The average template is created by registering all scans from one patient to a single scanning instance and then taking the mean of these images. Therefore, each subject will have unique average templates. Every MRI and PET scan is registered to the respective average template before the traditional registration with the MNI152 template. This ensures that all of the scans are registered both temporally within each patient's history and generally across the entire dataset. FSL FLIRT was used to perform the registrations [33].

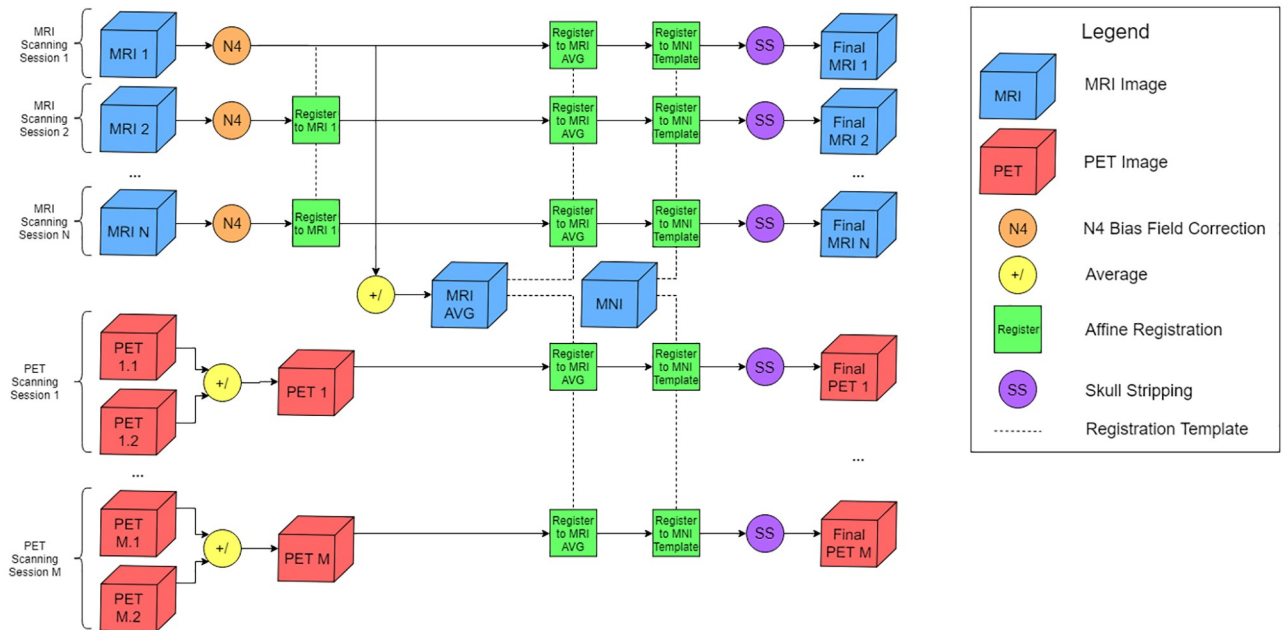


Fig 1. Pre-processing pipeline for a single subject. A subject has N MRI scanning sessions and M PET scanning sessions; therefore, the pipeline yields N MRI images and M PET images. The pipeline is repeated for each subject in the dataset.

<https://doi.org/10.1371/journal.pone.0225759.g001>

Skull stripping. Skull stripping is used to remove non-brain tissue voxels from the images. This is generally framed as a segmentation problem wherein clustering can be used to separate the voxels accordingly, as in FSL's brain extraction tool (BET) [34]. However, given that the scans were already registered to a standard space, skull stripping was a straightforward task. A brain mask in MNI152 space was used to zero out any non-brain voxels in both the MRI and PET images.

Fig 1 shows the pipeline in its entirety. The process is performed for all MRI and PET images for a single patient in the dataset before proceeding to the next. N4 correction is applied to all of the MRI scans before any registration steps. All MRI scans are registered to the first scanning time point, and the resulting images are averaged to create the average template. The N4 corrected scans are registered to this space before being registered with the MNI152 template. The resulting images are then skull stripped using a binary mask.

Amyloid AV-45 PET scans were collected over 20 minutes in dynamic list-mode 50 minutes post-injection of 370 MBq 18F-florbetapir. PET scans were attenuation corrected using a computed tomography scan. The first 10 minutes of PET acquisition was reconstruction into two 5 minute frames. Frames were motion corrected together and referenced (normalized) by the whole cerebellum. Each PET scan was registered to the individual's average T1 template with a 6 DOF registration and then the pre-computed 12 DOF registration from average T1 to MNI152 was concatenated and applied to the PET images to move them from native PET to MNI152 space. Finally, the PET images were skull stripped as above.

Network

The CNN architecture is fairly traditional in its construction and is most similar to that in [23]. Because the goal of this investigation is modality comparison, a representative CNN architecture was used rather than one with very specific modifications aimed at maximizing classification scores. In this manner, the modality comparison would not be obfuscated by

the nuances of the network. The network takes as input a full 3D MRI or PET image and outputs a diagnosis label. While several processing layers exist in the network, there are only three different varieties: convolutional layers, max pooling layers, and fully connected layers. Convolutional layers constitute the backbone of the CNN. As the name suggests, 3D filters are convolved with the input to the layer. Each kernel is made of learned weights that are shared across the whole input image; yet, each processing layer can have multiple trainable kernels. This allows kernel specialization while still affording the ability to capture variations at each layer. Following convolutional layers, it is common to have max pooling layers. These layers downsample an input image by outputting the maximum response in a given region. For example, a max pooling layer with a kernel size of 2x2x2 will result in an output image that is half the input size in each dimension. Each voxel in the output will correspond to the maximum value of the input image in the associated 2x2x2 window. Fully connected layers are often placed at the end of a CNN. These layers take the region specific convolutional features learned earlier in the network and allow connections between every feature. The weights in these layers are also trainable; therefore, these layers aggregate the region features and learn global connections between them. As a result, the output of the final fully connected layer in the CNN is the final diagnosis label.

Fig 2 is a diagram of the final CNN architecture for a single modality. In this instance, the network accepts MRI or PET images of size 182x218x182 (due to the MNI template size), but in principle a CNN can accept an image of any size. The image is then processed by three pairs of alternating convolutional (20 kernels of size 5x5x5) and max pooling layers (kernel size 2x2x2). The convolutional layers use the ReLU [35] activation function. Following these layers, the feature vector is flattened before being passed as input to a fully connected layer with 1024 nodes, a second fully connected layer of 128 nodes, and finally a fully connected layer with the number of diagnosis categories. In this case, there are 2 diagnosis categories corresponding to individuals with AD and healthy controls. The two fully connected layers also use the ReLU activation function, but the final classification is done with the softmax function.

Fig 3 shows the extension of the network for the fusion case. In this setting, the network takes both an MRI and PET image of size 182x218x182 as input into parallel branches. These branches are structured in the same manner as in the former case, but an additional fully connected layer of 128 nodes is added at the end in order to fuse the information from both modalities before the final classification is made. Additionally, the number of kernels in each

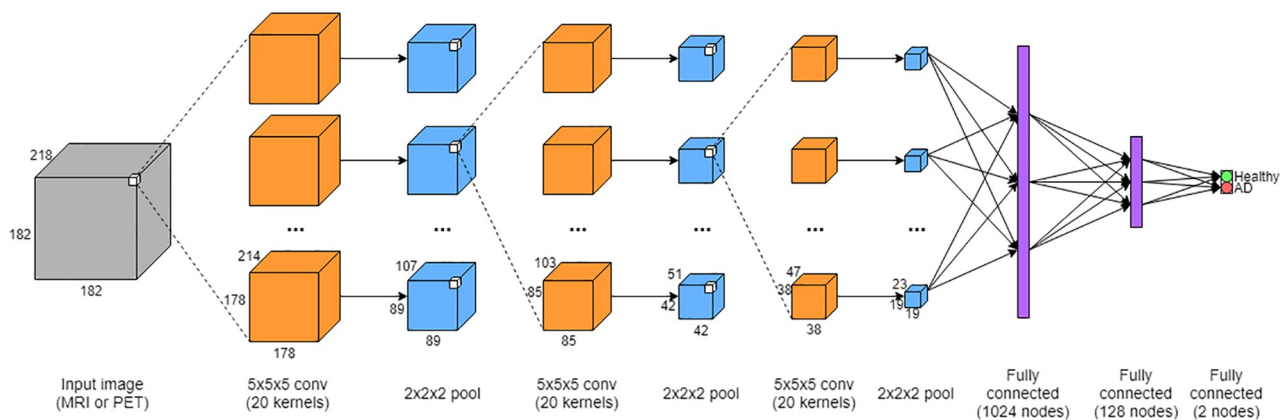


Fig 2. Convolutional neural network for one modality. A single MRI or PET volume is taken as input, and the output is a binary diagnosis label of either “Healthy” or “AD”.

<https://doi.org/10.1371/journal.pone.0225759.g002>

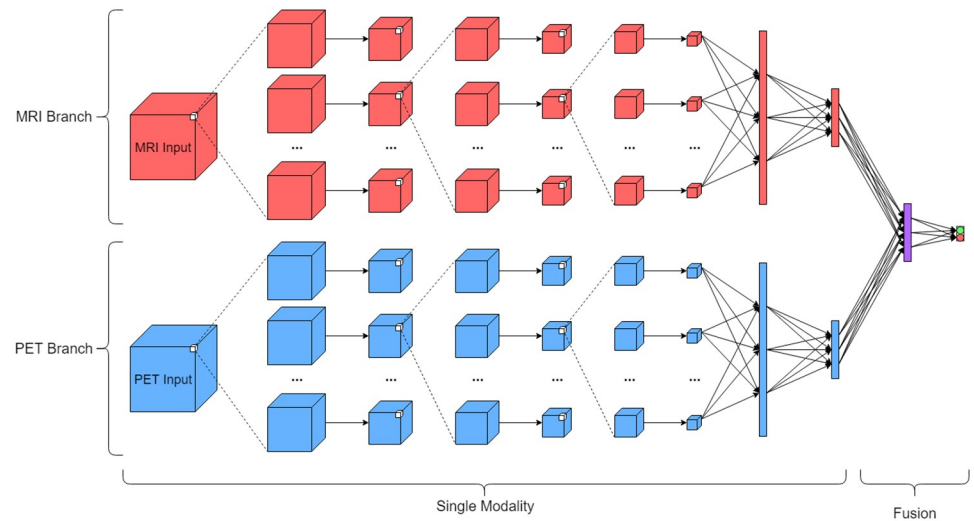


Fig 3. Convolutional neural network for fusing MRI and PET modalities. An MRI and PET scan from a single patient is taken as input, and the output is again a binary diagnosis label.

<https://doi.org/10.1371/journal.pone.0225759.g003>

convolutional layer was changed from 20 to 10 in order to keep the number of weights in the fusion network approximately the same as in the single modality network.

Experimental design

Classification experiments were performed on the Alzheimer's Disease Neuroimaging Initiative (ADNI) [36] database. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease. The set has clinical data from hundreds of study participants including neuroimaging modalities, demographics, medical history, and genetic sequencing. This work analyzed T1-weighted MRI and amyloid PET images in addition to the diagnosis labels given to patients at each study visit. Neurological test scores were examined in order to validate these labels, but were not used during network training. Data was only used from participants who had at least one scanning session for both MRI and PET. Additionally, scanning sessions were not considered if neurological testing was not performed within 2 months of the scanning session. This was to ensure that the diagnosis label provided during the scanning sessions had clinical justification. As a result, a subset of 723 ADNI patients were used. As in [16], individual scanning sessions from the same patient were considered separately in this work. This resulted in 1299 MRI scans, each falling into either the healthy or AD category. Patients underwent less PET scanning, with a total of 585 scans. Classification experiments were initially performed using only one modality, either MRI or PET, and using the appropriate data subset. Due to the fact that more MRI data exists than PET data, two different MRI classification experiments were performed. In one case, all of the available MRI data was used. In the other, the MRI data was limited to only use the same number of scans as the PET dataset.

These sets were further split into training and testing components in order to ascertain the generalizability of the algorithm. When splitting the data into training and testing subsets, scanning sessions from a single patient were not used in both the testing and training subsets. In other words, all of a single patient's scans were used in one of the two subsets. This was done to ensure that the algorithm would not overfit to the patient's identity rather than

learning the disease pattern. In some previous works, it is unclear whether this procedure was done. As a result, classification results in some previous work may have been inflated by models that overfit on certain individuals in the dataset.

Following this, fusion experiments were performed, where an MRI and PET scan from the same individual at a given time were used. Each scan was sent through parallel CNN branches. At the final fully connected layer of each branch, the features were merged into another fully connected layer that was used to produce the classification result. These experiments used the same number of data points as the PET experiments, albeit with each data point having an associated MRI and PET scan. Again, the testing and training subsets were made such that no patient's data was used in both subsets.

The neural network was constructed in Python using Keras [37] as a front-end and TensorFlow [38] as the back-end deep learning framework. The optimization procedure used stochastic gradient descent with a learning rate of 0.0001 and a momentum of 0.9. Categorical cross-entropy was used to classify the results of the CNN into the diagnosis labels. Training was done on an Nvidia Titan Z GPU and took approximately 20 epochs to complete each experiment. Depending on the dataset size, epoch training times ranged between approximately 45 minutes and 1.5 hours.

Results and discussion

Table 2 details the results of the classification experiments. The experiments were each performed 5 times holding out a different random subset of the data for validation. The mean age and gender splits for each validation subset are shown in Table 3. One can see that each validation subset is not biased by patient age or gender. The networks in each experiment were trained independently and from scratch using different random weight initializations. The mean validation accuracies in percentages are reported along with the corresponding standard deviations. To reiterate, the structure of the MRI and PET networks are identical, as they both take in a single volume and have the same number of trainable weights. The fusion network takes in two volumes, one from each modality, into parallel branches that each have half the number of weights as a single MRI or PET network. Aside from a few extra weights at the end of the fusion network, the total number of weights in all three networks is roughly the same. Additionally, the fusion network used the same number of data points as the PET network, but each included two volumes instead of one. The MRI network was able to use more data points due to the larger number of MRI scanning sessions. Consequently, two MRI experiments were run: one using all available MRI data and one with a limited dataset of the same size as the PET dataset.

To begin, the full data MRI network is able to classify with 87% accuracy. While this number is respectable, the performance could improve beyond 95% by employing techniques such as those described in [21–23]. However, we once again underscore that the goal is to compare

Table 2. MRI and amyloid PET fusion classification accuracies (%).

Trial	MRI (Full)	MRI (Limited)	AV-45 PET	Fusion
1	91.59	78.38	84.77	94.45
2	90.46	80.43	90.65	93.47
3	84.45	66.67	80.98	90.30
4	84.64	69.97	84.92	93.41
5	86.29	73.33	84.45	90.15
Total	87.49 ± 3.33	73.76 ± 5.72	85.15 ± 3.48	92.34 ± 1.95

<https://doi.org/10.1371/journal.pone.0225759.t002>

Table 3. Classification subject age and gender breakdown.

Trial	Age	Gender Split % (M/F)
1	74.98 ± 7.30	46.15/53.85
2	74.60 ± 6.72	44.57/55.43
3	74.29 ± 6.91	51.65/48.35
4	75.83 ± 7.72	57.87/42.13
5	73.79 ± 7.64	54.35/45.65
Total	74.70 ± 7.30	51.53/48.47

<https://doi.org/10.1371/journal.pone.0225759.t003>

the performance of the data modalities in the most balanced way possible. The inclusion of some of the more specific techniques in [21–23], such as pre-training the CNN filters with an autoencoder, does not enhance the modality comparison. Rather, the added complexity may obfuscate the findings if the pre-training effectiveness differed. That said, the full data MRI results do not tell the full story in the context of modality fusion. Because the MRI dataset is much larger than that of the PET, the potential for the network to learn is greatly increased. Thus, a direct comparison between the full data MRI network and the PET network could be misleading, as the MRI results may be inflated. Thus, one must look at the limited data MRI classification results when comparing the modalities and fusion head to head. In this case, because the dataset was limited to less than half of the available scans, the network was only able to achieve an accuracy of 74%. This discrepancy is somewhat expected, but moreover it highlights a large point about the availability of training data. Given this accuracy differential for the MRI data, one can imagine the potential benefits to the PET and fusion results as the number of available amyloid PET scans increases. On that note, it can be seen that the PET network performs much better than the MRI network trained with the equivalent data size. The accuracy of 85% is even comparable to the full data MRI network, despite being trained with far fewer examples.

To properly discern the distinction between the MRI and PET performance, one must examine the biological facets of the modality. Amyloid accumulation has been hypothesized to begin more than two decades before symptoms occur [28]. In a longitudinal study of dominantly inherited Alzheimer's disease [39], elevated amyloid PET signals were found 22 years before expected onset of symptoms.

Separate from the CNN pipeline, a standard method, previously described [40], was used to calculate the total amyloid burden. Briefly, FreeSurfer [41, 42] was used to parcellate the T1-weighted MRI scan taken closest to the amyloid PET visit. Whole cerebellar referenced cortical regions normalized by volume were used to calculate a single weighted standard uptake value ratio (SUVR). The previously defined cutoff of ≥ 1.11 was used to define amyloid positivity [40].

In the first set of classification experiments, out of the 11 amyloid PET scans that were incorrectly classified, 7 were controls and 4 were Alzheimer's dementia cases. All 7 control cases had elevated amyloid SUVR ≥ 1.11 (average SUVR 1.42 ± 0.12). Two Alzheimer's dementia cases were amyloid positive (i.e., true misclassification) and two Alzheimer's cases were amyloid negative (average SUVR 0.95 ± 0.03) and therefore are unlikely to have underlying Alzheimer's disease neuropathology. If the 7 elevated amyloid controls and 2 amyloid negative AD cases are removed, then the effective PET classification accuracy rises from 85% to 97%.

The newly proposed NIA-AA research criteria for Alzheimer's disease [43] points out that amnesic dementia diagnoses are not sensitive or specific for AD neuropathologic change.

From 10 to 30% of individuals classified as AD dementia do not display AD neuropathology at autopsy [44] and 30 to 40% of individuals classified as unimpaired healthy have AD neuropathologic change at post-mortem examination [45, 46]. The proposed CNN here is capturing this mismatch between biomarker and diagnosis. The CNN labels healthy individuals with high amyloid PET as AD and those with Alzheimer's dementia and low amyloid PET as non-AD. Thus, while the phenomenon negatively impacts performance in this context, amyloid PET scans may be adept in a longitudinal study because elevated amyloid precedes symptom onset.

With this in mind, a few points regarding the comparison between MRI and amyloid PET can be stated. First, it is clear that the network benefited from the use of the full training set. Therefore, one can expect the PET performance to increase as well once amyloid scans become more readily available. This potential improvement may not be on the same scale, given that the PET performance is already higher than the MRI performance using the same training set size. This PET performance is likely due to the fact that amyloid accumulation may occur far ahead of symptom onset, which in turn may occur in advance of structural changes that would be detectable with an MRI. Moreover, the false positive cases of the PET network all had elevated amyloid levels. This indicates that the network is effective at deducing elevation of amyloid levels from the PET scan and converting this information into a disease status determination. Furthermore, in these false positive cases, it is quite possible that these patients develop Alzheimer's neuropathology at a later time. This in turn would support the justification for using amyloid PET in a longitudinal prediction case rather than structural MRI data alone.

The final noteworthy result of the investigation is that the fusion network outperformed both the individual MRI and PET networks. Additionally, the fusion network outperformed the full data MRI network despite the fact that less data points were used. Again, having more PET scans available in the fusion case may further improve the accuracy. The fusion performance is consistent with the other results [11, 24, 25], despite the fact that these investigations use FDG PET rather than amyloid. One can see back in Table 1 that the MRI and FDG PET classification accuracies are rather comparable in all cases, while the fusion results are greater than either individual modality. In our case, the amyloid PET results are much better than the MRI results when using the same amount of training data, and the fusion provides a similar benefit to accuracy. That said, one cannot make a direct head to head comparison between amyloid PET and FDG PET from this investigation alone due to the fact that different biological markers, data subsets, pre-processing methods, and classification algorithms were used. A further investigation that holds these factors constant would be required. Nonetheless, this investigation still clearly demonstrates the discriminative power of the amyloid PET modality and the potential for even further gains when fused with MRI.

Conclusion and future work

This work compared the effectiveness of the T1-weighted MRI and AV-45 amyloid PET modalities in the context of computer aided diagnosis using deep neural networks. Specifically, two identically structured CNNs were designed and trained on MRI and amyloid PET data that were pre-processed to be as fairly compared as possible. The classification results indicate that MRI data is less conducive to neural network training than amyloid PET data to predict clinical diagnosis. However, a network that uses both modalities, even with the same number of trainable weights, will achieve higher accuracy. This indicates that the two data types have complementary information that can be leveraged in these kinds of tasks. This phenomenon was also placed into the biological context of amyloid vs. MRI.

While these results are a step forward in the optimization of computer aided diagnosis tools for AD, the value from this investigation must be utilized in further applications. To begin, the efficacy of these algorithms could be examined when the MCI state is included in classification or applying the current investigation to FDG PET data for comparison. Following this, a natural extension can be made to looking at AD patients on a functional spectrum rather than distinct diagnosis categories. Additionally, as previously alluded to, longitudinal studies that use several scanning sessions of multiple modalities may not only improve classification performance, but also allow the ability to perform more complex tasks such as predicting future cognitive decline irrespective of clinical phenotype. These results would be invaluable to clinicians, as they can directly inform decisions regarding preemptive or preventative care.

Acknowledgments

The authors would like to thank the Integrated Data Driven Discovery in Earth and Astrophysical Sciences (IDEAS) program at Northwestern University, the Biomedical Data Driven Discovery (BD3) training program at Northwestern University, and the National Institute on Aging. The authors would also like to thank the QUEST High Performance Computing Cluster at Northwestern University for computational resources.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI). This investigation was led by Michael W. Weiner (Michael.Weiner@ucsf.edu) and the complete list of collaborators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author Contributions

Conceptualization: Arjun Punjabi, Adam Martersteck, Yanran Wang, Todd B. Parrish, Aggelos K. Katsaggelos.

Data curation: Arjun Punjabi, Adam Martersteck.

Formal analysis: Arjun Punjabi, Adam Martersteck, Aggelos K. Katsaggelos.

Funding acquisition: Aggelos K. Katsaggelos.

Investigation: Arjun Punjabi, Adam Martersteck, Todd B. Parrish, Aggelos K. Katsaggelos.

Methodology: Arjun Punjabi, Adam Martersteck, Todd B. Parrish, Aggelos K. Katsaggelos.

Project administration: Todd B. Parrish, Aggelos K. Katsaggelos.

Resources: Arjun Punjabi, Todd B. Parrish, Aggelos K. Katsaggelos.

Software: Arjun Punjabi, Adam Martersteck, Yanran Wang.

Supervision: Todd B. Parrish, Aggelos K. Katsaggelos.

Validation: Arjun Punjabi.

Visualization: Arjun Punjabi.

Writing – original draft: Arjun Punjabi.

Writing – review & editing: Arjun Punjabi, Adam Martersteck, Yanran Wang, Todd B. Parrish, Aggelos K. Katsaggelos.

References

1. Hebert L, Scherr P, Bienias J, Bennett D, Evans D. State-specific projections through 2025 of Alzheimer disease prevalence. *Neurology*. 2004; 62(9):1645–1645. <https://doi.org/10.1212/01.wnl.0000123018.01306.10> PMID: 15136705
2. Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehericy S, Habert MO, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *NeuroImage*. 2011; 56(2):766–781. <https://doi.org/10.1016/j.neuroimage.2010.06.013> PMID: 20542124
3. Ashburner J. A fast diffeomorphic image registration algorithm. *NeuroImage*. 2007; 38(1):95–113. <https://doi.org/10.1016/j.neuroimage.2007.07.007> PMID: 17761438
4. Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain*. 2008; 131(3):681–689. <https://doi.org/10.1093/brain/awn319> PMID: 18202106
5. Liu M, Zhang D, Adeli E, Shen D. Inherent Structure-Based Multiview Learning With Multitemplate Feature Representation for Alzheimer's Disease Diagnosis. *IEEE Transactions on Biomedical Engineering*. 2016; 63(7):1473–1482. <https://doi.org/10.1109/TBME.2015.2496233> PMID: 26540666
6. Gerardin E, Chételat G, Chupin M, Cuingnet R, Desgranges B, Kim HS, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*. 2009; 47(4):1476–1486. <https://doi.org/10.1016/j.neuroimage.2009.05.036> PMID: 19463957
7. Sabuncu MR, Konukoglu E, Initiative ADN, et al. Clinical prediction from structural brain MRI scans: a large-scale empirical study. *Neuroinformatics*. 2015; 13(1):31–46. <https://doi.org/10.1007/s12021-014-9238-1> PMID: 25048627
8. Zu C, Jie B, Liu M, Chen S, Shen D, Zhang D, et al. Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment. *Brain imaging and behavior*. 2016; 10(4):1148–1159. <https://doi.org/10.1007/s11682-015-9480-7> PMID: 26572145
9. Zhu X, Suk HI, Shen D. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage*. 2014; 100:91–105. <https://doi.org/10.1016/j.neuroimage.2014.05.078> PMID: 24911377
10. Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2013. p. 583–590.
11. Suk HI, Lee SW, Shen D, Initiative ADN, et al. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*. 2014; 101:569–582. <https://doi.org/10.1016/j.neuroimage.2014.06.077> PMID: 25042445
12. Li F, Tran L, Thung KH, Ji S, Shen D, Li J. A robust deep model for improved classification of AD/MCI patients. *IEEE journal of biomedical and health informatics*. 2015; 19(5):1610–1616. <https://doi.org/10.1109/JBHI.2015.2429556> PMID: 25955998
13. Yang X, Wu Q, Hong D, Zou J. Spatial regularization for neural network and application in Alzheimer's disease classification. In: *Future Technologies Conference (FTC)*. IEEE; 2016. p. 831–837.
14. Sarraf S, Tofighi G. Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. *arXiv preprint arXiv:160308631*. 2016;.

15. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998; 86(11):2278–2324. <https://doi.org/10.1109/5.726791>
16. Sarraf S, Tofighi G, et al. DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI. *bioRxiv*. 2016; p. 070441.
17. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015. p. 1–9.
18. Li R, Zhang W, Suk HI, Wang L, Li J, Shen D, et al. Deep learning based imaging data completion for improved brain disease diagnosis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2014. p. 305–312.
19. Hinton GE, Zemel RS. Autoencoders, minimum description length and Helmholtz free energy. In: *Advances in neural information processing systems*; 1994. p. 3–10.
20. Liu S, Liu S, Cai W, Che H, Pujol S, Kikinis R, et al. Multimodal neuroimaging feature learning for multi-class diagnosis of Alzheimer's disease. *IEEE Transactions on Biomedical Engineering*. 2015; 62(4):1132–1140. <https://doi.org/10.1109/TBME.2014.2372011> PMID: 25423647
21. Gupta A, Ayhan M, Maida A. Natural image bases to represent neuroimaging data. In: *International Conference on Machine Learning*; 2013. p. 987–994.
22. Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. *arXiv preprint arXiv:150202506*. 2015;.
23. Hosseini-Asl E, Keynton R, El-Baz A. Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In: *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE; 2016. p. 126–130.
24. Vu TD, Yang HJ, Nguyen VQ, Oh AR, Kim MS. Multimodal learning using convolution neural network and Sparse Autoencoder. In: *Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on*. IEEE; 2017. p. 309–312.
25. Shi J, Zheng X, Li Y, Zhang Q, Ying S. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE journal of biomedical and health informatics*. 2018; 22(1):173–183. <https://doi.org/10.1109/JBHI.2017.2655720> PMID: 28113353
26. Shi J, Xue Z, Dai Y, Peng B, Dong Y, Zhang Q, et al. Cascaded Multi-Column RVFL+ Classifier for Single-Modal Neuroimaging-Based Diagnosis of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*. 2018;.
27. Gong B, Shi J, Ying S, Dai Y, Zhang Q, Dong Y, et al. Neuroimaging-based diagnosis of Parkinson's disease with deep neural mapping large margin distribution machine. *Neurocomputing*. 2018; 320:141–149. <https://doi.org/10.1016/j.neucom.2018.09.025>
28. Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*. 2013; 12(2):207–216. [https://doi.org/10.1016/S1474-4422\(12\)70291-0](https://doi.org/10.1016/S1474-4422(12)70291-0)
29. McVeigh E, Bronskill M, Henkelman R. Phase and sensitivity of receiver coils in magnetic resonance imaging. *Medical physics*. 1986; 13(6):806–814. <https://doi.org/10.1118/1.595967> PMID: 3796476
30. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*. 1998; 17(1):87–97. <https://doi.org/10.1109/42.668698> PMID: 9617910
31. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*. 2010; 29(6):1310–1320. <https://doi.org/10.1109/TMI.2010.2046908> PMID: 20378467
32. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, et al. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*. 2011; 54(1):313–327. <https://doi.org/10.1016/j.neuroimage.2010.07.033> PMID: 20656036
33. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 2002; 17(2):825–841. <https://doi.org/10.1006/nimg.2002.1132>
34. Smith SM. Fast robust automated brain extraction. *Human brain mapping*. 2002; 17(3):143–155. <https://doi.org/10.1002/hbm.10062> PMID: 12391568
35. Hahnloser RH, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*. 2000; 405(6789):947–951. <https://doi.org/10.1038/35016072> PMID: 10879535
36. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of magnetic resonance imaging*. 2008; 27(4):685–691. <https://doi.org/10.1002/jmri.21049> PMID: 18302232

37. Chollet F, et al. Keras; 2015.
38. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: a system for large-scale machine learning. In: OSDI. vol. 16; 2016. p. 265–283.
39. Gordon BA, Blazey TM, Su Y, Hari-Raj A, Dincer A, Flores S, et al. Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *The Lancet Neurology*. 2018; 17(3):241–250. [https://doi.org/10.1016/S1474-4422\(18\)30028-0](https://doi.org/10.1016/S1474-4422(18)30028-0) PMID: [29397305](https://pubmed.ncbi.nlm.nih.gov/29397305/)
40. Landau SM, Mintun MA, Joshi AD, Koeppe RA, Petersen RC, Aisen PS, et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of neurology*. 2012; 72(4):578–586. <https://doi.org/10.1002/ana.23650> PMID: [23109153](https://pubmed.ncbi.nlm.nih.gov/23109153/)
41. Fischl B. FreeSurfer. *Neuroimage*. 2012; 62(2):774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021> PMID: [22248573](https://pubmed.ncbi.nlm.nih.gov/22248573/)
42. Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE transactions on medical imaging*. 2001; 20(1):70–80. <https://doi.org/10.1109/42.906426> PMID: [11293693](https://pubmed.ncbi.nlm.nih.gov/11293693/)
43. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*. 2018; 14(4):535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>
44. Nelson PT, Head E, Schmitt FA, Davis PR, Neltner JH, Jicha GA, et al. Alzheimer's disease is not brain aging: neuropathological, genetic, and epidemiological human studies. *Acta neuropathologica*. 2011; 121(5):571–587. <https://doi.org/10.1007/s00401-011-0826-y> PMID: [21516511](https://pubmed.ncbi.nlm.nih.gov/21516511/)
45. Bennett D, Schneider J, Arvanitakis Z, Kelly J, Aggarwal N, Shah R, et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology*. 2006; 66(12):1837–1844. <https://doi.org/10.1212/01.wnl.0000219668.471116.e6> PMID: [16801647](https://pubmed.ncbi.nlm.nih.gov/16801647/)
46. Price JL, Davis P, Morris J, White D. The distribution of tangles, plaques and related immunohistochemical markers in healthy aging and Alzheimer's disease. *Neurobiology of aging*. 1991; 12(4):295–312. [https://doi.org/10.1016/0197-4580\(91\)90006-6](https://doi.org/10.1016/0197-4580(91)90006-6) PMID: [1961359](https://pubmed.ncbi.nlm.nih.gov/1961359/)