Repurposing Artificial Intelligence Tools for Disease Modeling: Case Study of Face Recognition Deficits in Neurodegenerative Diseases

Gargi Singh[1] and Murali Ramanathan[1,*]

Face recognition deficits occur in diseases such as prosopagnosia, autism, Alzheimer's disease, and dementias. The objective of this study was to evaluate whether degrading the architecture of artificial intelligence (AI) face recognition algorithms can model deficits in diseases. Two established face recognition models, convolutional-classification neural network (C-CNN) and Siamese network (SN), were trained on the FEI faces data set (~14 images/person for 200 persons). The trained networks were perturbed by reducing weights (weakening) and node count (lesioning) to emulate brain tissue dysfunction and lesions, respectively. Accuracy assessments were used as surrogates for face recognition deficits. The findings were compared with clinical outcomes from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data set. Face recognition accuracy decreased gradually for weakening factors less than 0.55 for C-CNN, and 0.85 for SN. Rapid accuracy loss occurred at higher values. C-CNN accuracy was similarly affected by weakening any convolutional layer whereas SN accuracy was more sensitive to weakening of the first convolutional layer. SN accuracy declined gradually with a rapid drop when nearly all nodes were lesioned. C-CNN accuracy declined rapidly when as few as 10% of nodes were lesioned. CNN and SN were more sensitive to lesioning of the first convolutional layer. Overall, SN was more robust than C-CNN, and the findings from SN experiments were concordant with ADNI results. As predicted from modeling, brain network failure quotient was related to key clinical outcome measures for cognition and functioning. Perturbation of AI networks is a promising method for modeling disease progression effects on complex cognitive outcomes.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Artificial intelligence (AI) algorithms have become increasingly proficient at tasks, such as face recognition, interpreting handwriting, and conversation, which are important features of human cognition and functioning. People with neurodegenerative diseases, such as Alzheimer's disease (AD), often experience cognitive deficits in these domains.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ To evaluate whether degrading the architecture of AI methods effective at face recognition can model face recognition deficits in disease states. The larger goal was to evaluate whether this novel AI-based strategy could be used to build models for complex cognitive outcome measures in AD.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ The results indicate that perturbing the architecture of AI methods can emulate the emergence of face recognition deficits. Neuroimaging-derived measures of network failure are associated with neuropsychological outcome measures of cognition and function that are commonly used in AD clinical trials.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ The results show that the failure of neural signaling networks due to neurodegenerative processes can account for cognitive deficits in activities, such as face recognition, that are important for patient functioning in AD. The findings represent an innovative strategy for building models for cognitive outcomes in neurodegenerative diseases.

[1]Department of Pharmaceutical Sciences, University at Buffalo, The State University of New York, Buffalo, New York, USA. *Correspondence: Murali Ramanathan (murali@buffalo.edu)

Received March 8, 2023; accepted June 20, 2023. doi:10.1002/cpt.2987

Diseases that cause neurodegeneration in brain tissue adversely affect neurocognitive processing and result in cognitive decline. There are no treatments for reversing neurodegeneration that is the pathophysiological hallmark of many diseases including Alzheimer's disease (AD), Parkinson's disease, multiple sclerosis, and dementias.

Face recognition is a cognitive function that is critical for everyday social and interpersonal interactions. Visual cues from faces are also used to infer emotion. Face recognition deficits are the defining clinical feature in developmental and acquired forms of prosopagnosia. People with prosopagnosia can adequately recognize other everyday objects, and rely on non-facial cues (e.g., voice, gait, etc.) to recognize familiar individuals.[1,2] Many other neurodevelopmental and neurodegenerative diseases ranging from autism spectrum disorder to AD can affect the ability to identify familiar faces, and this can severely impact the quality of the patient's social interactions and cause emotional pain to family members and caregivers.

The potential utility of artificial intelligence (AI) models, particularly neural networks[3] for delineating the hidden computational frameworks used by the brain to perform cognitive tasks has long been recognized but were constrained by hardware limitations, lack of training data, and computational complexity.[4] As early as 1980, Fukushima proposed the "neurocognitron," a three-layer convolutional neural network architecture inspired by a representation of visual signal transduction in the eyes.[5] Physiological activity evoked in different brain regions by cognitive tasks can now be imaged with electroencephalograms (EEGs), functional magnetic resonance imaging (fMRI), and positron-emission tomography (PET). Brain activity data are typically analyzed using correlation matrices, network analyses, and Bayesian models to identify the inter-dependencies between different regions when performing cognitive tasks.[6,7] In clinical practice, EEG and fMRI mapping are used to localize the brain regions involved with epileptogenic activity and speech to guide resection surgery for epilepsy.[8] Computational and cognitive neuroscience take bottom-up and top-down approaches[4] that are complementary with each other, and with emerging deep learning-driven pattern recognition and generative AI methods.

AI methods have become proficient at face recognition and are being deployed in real-world applications, such as photographic image curation, crime fighting, and national security. Convolutional neural networks are particularly effective for face recognition (and other object recognition) tasks from images because they can reduce data complexity while retaining key aspects of the local correlation structure; they are also capable of learning, and robust to translation, scaling, and distortions of the pattern. However, there are knowledge gaps that need to be bridged so that AI can be leveraged in clinical pharmacology for building useful models for face recognition and other cognitive deficits in neurodegenerative diseases.

The specific goal was to assess whether AI approaches could be utilized to model face recognition deficits in neurodegenerative diseases. We obtained performance accuracy for two AI face recognition models, convolutional-classification neural network (C-CNN) and Siamese network (SN), that implement convolutional neural networks with different architectures. The architecture was perturbed, and the ensuing loss of accuracy was used to emulate the effects of neurodegeneration in causing face recognition deficits. The parallels between performance loss in AI models and the severity of deficits in neurological diseases were assessed.

METHODS

Face recognition algorithms

We assessed C-CNN and SN developed by others for face recognition. The structures of C-CNN and SN are summarized in **Table 1**.

The C-CNN had three convolutional layers for feature extraction followed by two dense layers. Each convolutional layer was followed by a pooling layer to reduce the size of feature maps. The number of neurons in the final layer was set to the number of subjects to be identified – this was 200 in this case. The output of the final layer was softmax transformed. The peak value from the softmax probability mass vector was used to identify the subject of the test image. The C-CNN code was sourced from ref. 9.

Table 1 Details of the architecture of the neural networks

Layer	Structure	Parameters
Convolutional-classification neural network structure		
Layers 1–3	Convolution	Filters: 32; filter size: 3 ×3 stride: (1, 1); activation: linear
	Pooling	Type: max; size: 2 ×2; stride: (1, 1)
Layer 4	Dense layer	Neurons: 512
Dropout	Dropout	Used only during training, to avoid overfitting
Layer 5	Dense layer	Neurons: 200 (equivalent to number of classes)
Siamese network structure		
Layer 1	Convolution	Number of filters: 96; filter size: 11 ×11 stride: (4, 4); activation: ReLU
	Batch normalization	Momentum: 0.1
	Pooling	type: max; size: 3 ×3; stride: (2, 2)
Layer 2	Convolution	Number of filters: 256; filter size: 5 ×5 stride: (1, 1); activation: ReLU
	Batch normalization	Momentum: 0.1
	Pooling	Type: max; size: 3 ×3; stride: (2, 2)
Layer 3	Convolution	Number of filters: 384; filter size: 3 ×3 stride: (1, 1); activation: ReLU
	Batch normalization	Momentum: 0.1
Layer 4	Dense layer	Neurons: 1024
Layer 5	Dense layer	Neurons: 256
Layer 6	Dense layer	Neurons: 16

Batch Normalization: Applied before passing the inputs to the activation function. Activation: Applied after convolution.
ReLU, rectified linear units.

Each arm of the SN was comprised of three convolutional and two pooling layers followed by three dense layers. The distance between the final output vector from each arm was computed as a measure of dissimilarity between the two faces. Binary classification based on this distance was used to assess whether the two faces were from the same subject or not.

The SN code was sourced from ref. 10. This code was modified by adding a batch normalization layer before each activation layer and by increasing the size of the last layer from 2 neurons to 16 neurons. Batch normalization improves the learning rate by standardizing input values and addressing internal covariate shift between layers. We reasoned that a higher-dimensional output would be better at encoding nuanced differences between faces. Experiments (data not shown) confirmed that these changes improved the performance of SN.

The minimum of the distances of the test image from the set of three reference images (frontal, left profile, and right profile) for each subject was computed. The subject with the smallest distance from the test image was identified as the subject of the test image.

For C-CNN and SN, face recognition accuracy was defined as the fraction of correctly identified test images.

Face data set

The FEI face database (https://fei.edu.br/~cet/facedatabase.html), which contains color photograph images of 200 subjects (100 male and 100 female subjects) against a white background, was used.[11,12] The data set has ~14 images per subject that differ in face angles, facial expressions, and lighting.

Image data pre-processing. Images were cropped to contain faces using the Viola-Jones Haar cascade face detector.[13] Images were then resized to 100 × 100 pixels, converted to grayscale, and pixel intensity was normalized using the median intensity.

Network training

C-CNN training. The test data set consisted of three images for each subject. The training set was comprised of the remaining images.

Data augmentation was used for the C-CNN training set because the C-CNN, which relies on a classification strategy, is susceptible to over-fitting when trained on inadequate data. Augmentation increases training sample size and variety in the training input and can improve robustness of the C-CNN. Built-in transforms in the Pytorch Dataloader were used to produce 10 additional images for each image in the training set.

The model was trained on the augmented training set for 50 epochs with a batch size of 32, and saved whenever there was an increase in validation accuracy.

SN training. The SN training set was comprised of images for 197 subjects; 3 subjects were withheld from the training process. SN was trained in batches of size 64. The contrastive loss-based training process reduces the distance for image pairs from the same subject and increases the distance for image pairs from differing subjects. The SN was trained for 50 epochs. After each batch of training, a test batch of size 8 was loaded by sourcing images from the test subjects that were withheld from training. The model was saved whenever the contrastive loss decreased.

Effect of perturbing trained C-CNN and SN

We compared the effects of systematically perturbing the network architecture of the trained C-CNN and SN on the accuracy of face recognition to the corresponding unperturbed trained network. The overall goal was to assess the robustness to perturbations and to assess whether the repurposing of AI face recognition algorithms could plausibly emulate clinical features of face recognition deficits in neurological diseases.

Accuracy experiments were conducted on a common subset of 20 individuals (10 male and 10 female individuals) for the C-CNN and SN to manage time needed for the computations.

In "weakening" experiments, all node weights of the trained C-CNN or SN were multiplied by a weakening factor, with values between 0 and 1 in increments of 0.05. A weakening factor value of 1 denotes complete loss of the weights and a weakening factor value of zero denotes no change to the trained network weights.

In "lesioning" experiments, weights for a fraction of nodes selected at random from the trained C-CNN or SN were set to zero. The lesioning factor is the fraction of nodes selected: a value of 1 denotes complete loss of nodes and a value of zero denotes no change to the nodes of the trained network. Lesioning factor values between 0 and 1 in increments of 0.05 were evaluated; the accuracy results were based on averaging 5 lesioning experiments.

Contribution of individual facial landmarks

The contribution of facial landmarks was assessed for the SN by masking different combinations of the eyes and mouth. The DLib facial landmark identification algorithm was used to identify rectangular regions corresponding to the eyes and mouth.[14] The rectangular regions were black filled to create masks.

These experiments were conducted on the faces for 20 individuals (10 male and 10 female individuals). Seven copies of each face were created with the following masks: right eye, left eye, mouth, both eyes, left eye and mouth, right eye and mouth, and both eyes and mouth. The recognition accuracy across the different sets of masked faces was evaluated for SN.

Model evaluation with real-world data

Publicly available data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a multicenter study with clinical, imaging, and cerebrospinal fluid (CSF) protein biomarker measurements, were downloaded with the ADNIMERGE R package.[15] Network Failure Quotient (NFQ) data were downloaded separately. NFQ and key prognostic AD biomarkers at baseline (e.g., CSF tau and amyloid β42 (Aβ42)), MRI-derived entorhinal cortex, hippocampus, and fusiform gyrus volumes, were categorized into quintiles for graphical analyses.

RESULTS

Baseline performance of the algorithms

C-CNN and SN, which are two distinctive AI architectures (**Figure 1**) proven effective for face recognition tasks, were evaluated.

C-CNN directly predict the subject of the test image with a classification strategy. C-CNN computes high probability values when the index image (**Figure 2a**, column 2) is matched to the same individual. In comparison, C-CNN probabilities are lower for other mismatched individuals; representative results for two female (**Figure 2a**, Columns 3–4) and two male individuals (**Figure 2a**, columns 5–6) are shown.

SN architecture consists of two identical networks that take two images as input and outputs a distance measure for their dissimilarity. We created reference image sets of the front, left, and right profiles for each subject and used the minimum distance value to the test image. The SN distance for a representative female index image from the reference image for the same subject was 0.43 (**Figure 2b**, column 2). As expected, the distance of the index image from itself is zero (data not shown). The distances of

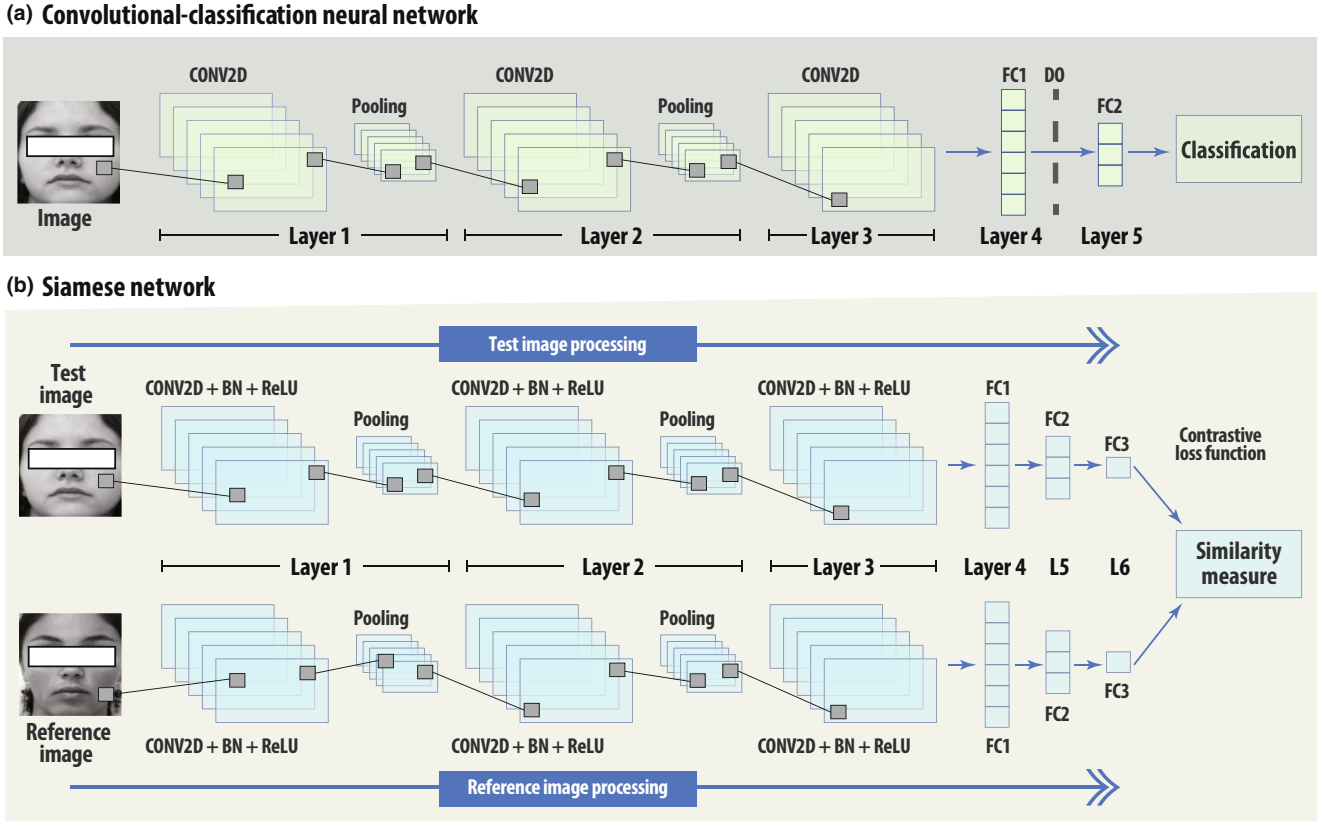


Figure 1 (**a, b**) are schematics of the architectures of the convolutional-classification neural network and the Siamese network (SN), respectively. Both the networks use three convolutional layers (denoted by CONV2D) with pooling. The convolutional neural network has a pooling layer after every convolutional layer whereas the SN directly flattens the input of the last convolutional layer before feeding it to the next dense layer. In (**b**), the reference images are selected sets containing left, right, and frontal view of each person. BN, batch normalization; FC, fully connected or dense layer; DO, dropout layer (used only during training); ReLU, rectified linear units. The subjects' eyes are masked in the figure to protect privacy in the publication. [Color figure can be viewed at wileyonlinelibrary.com]

the female index image from images of two other female subjects were 5.08 and 5.48 (**Figure 2b**, columns 3–4), and the distances from images of two male subjects were 5.97 and 8.95 (**Figure 2b**, columns 5–6). In **Figure 2c**, the histogram of within-individual distances (lighter gray bars) is closer to zero and narrow, whereas the histograms of the distances of the index image from images of other female subjects (white bars) and male subjects (darker gray bars) are centered at greater distance values.

Repurposing face recognition algorithms to emulate neurodegenerative diseases

Effects of weakening. Weakening reduces node weights in the neural network and can be viewed as a computational analog for loss of neurological function in diseases. **Figure 3a** summarizes the effects of global weakening on the face recognition accuracy of C-CNN and SN. The accuracy of both networks decreased only gradually over the broad range of weakening factors between 0 and 0.55 for C-CNN and 0 and 0.85 for SN. However, at higher values of weakening factors, greater rate of accuracy loss was observed. The onset of rapid loss occurred at ~0.55 for C-CNN and ~0.85 for SN suggesting that the accuracy of SN was more robust to large weakening factors than C-CNN.

Figure 3b shows the effects of weakening that was restricted to all 3 convolutional layers; the dense layers were not weakened. The trends in **Figure 3b** were concordant with **Figure 3a**. This indicates that the convolutional layers, which are involved in feature extraction, are important determinants of robustness of face recognition accuracy when weakening occurs. **Figure 3c** shows that the accuracy of the C-CNN and SN was relatively unaffected by weakening of all dense layers.

In **Figure 3d**, we assessed the effects of weakening of either the first, second, or third convolutional layers of the C-CNN and SN. The face recognition accuracy of the C-CNN was similarly affected by weakening of any one of the three convolutional layers. The accuracy of the SN was more sensitive to weakening of the first convolutional layer vis-a-vis weakening of the second and third convolutional layers. The accuracy of the C-CNN and SN methods was relatively stable to weakening of the dense layers (**Figure 3e**).

Effects of lesioning. Node lesioning can be viewed as a computational analog for loss of neuronal cells in diseases. Weights for a fraction of nodes selected at random were set to zero causing deletion of these nodes and their incoming and outgoing edges.

Figure 4a shows the effects of increasing lesioning across all layers of the C-CNN and SN networks on face recognition accuracy.

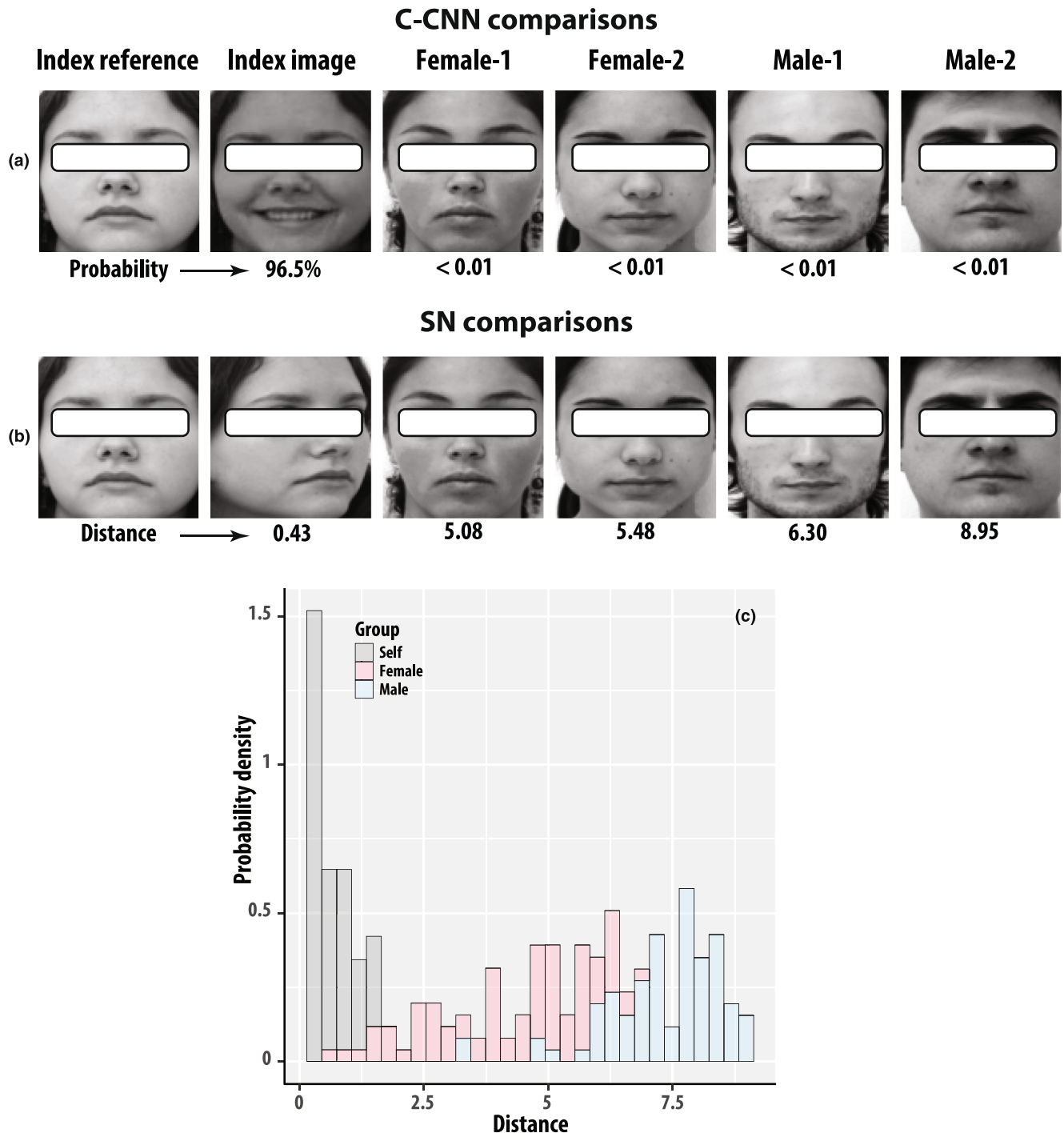


Figure 2 (**a**) Is graphical montage summarizing representative probabilities obtained from the convolutional-classification neural network (C-CNN). The first column contains a reference image for the subject of the index image in column 2. Columns 3 and 4 contain images of 2 other female subjects and columns 5 and 6 contain images of male subjects. The numbers below each image are probabilities obtained from the C-CNN. (**b**) Is a graphical montage summarizing representative distance (dissimilarity measure) results from the Siamese network. As in (**a**), the first column contains a reference image for the subject of the index image in column 2; columns 3 and 4 contain images of 2 other female subjects and columns 5 and 6 contain images of male subjects. The gray bars in (**c**) show the probability density histograms for the distance distributions of the index image (column 1 in **b**) from other images of the same individual (within-individual or intra-class distances). The pink and blue gray bars are the probability density histograms for the distance distributions of the index image from the reference images of other female and male subjects, respectively. The subjects' eyes are masked in the figure to protect privacy in the publication. [Color figure can be viewed at wileyonlinelibrary.com]

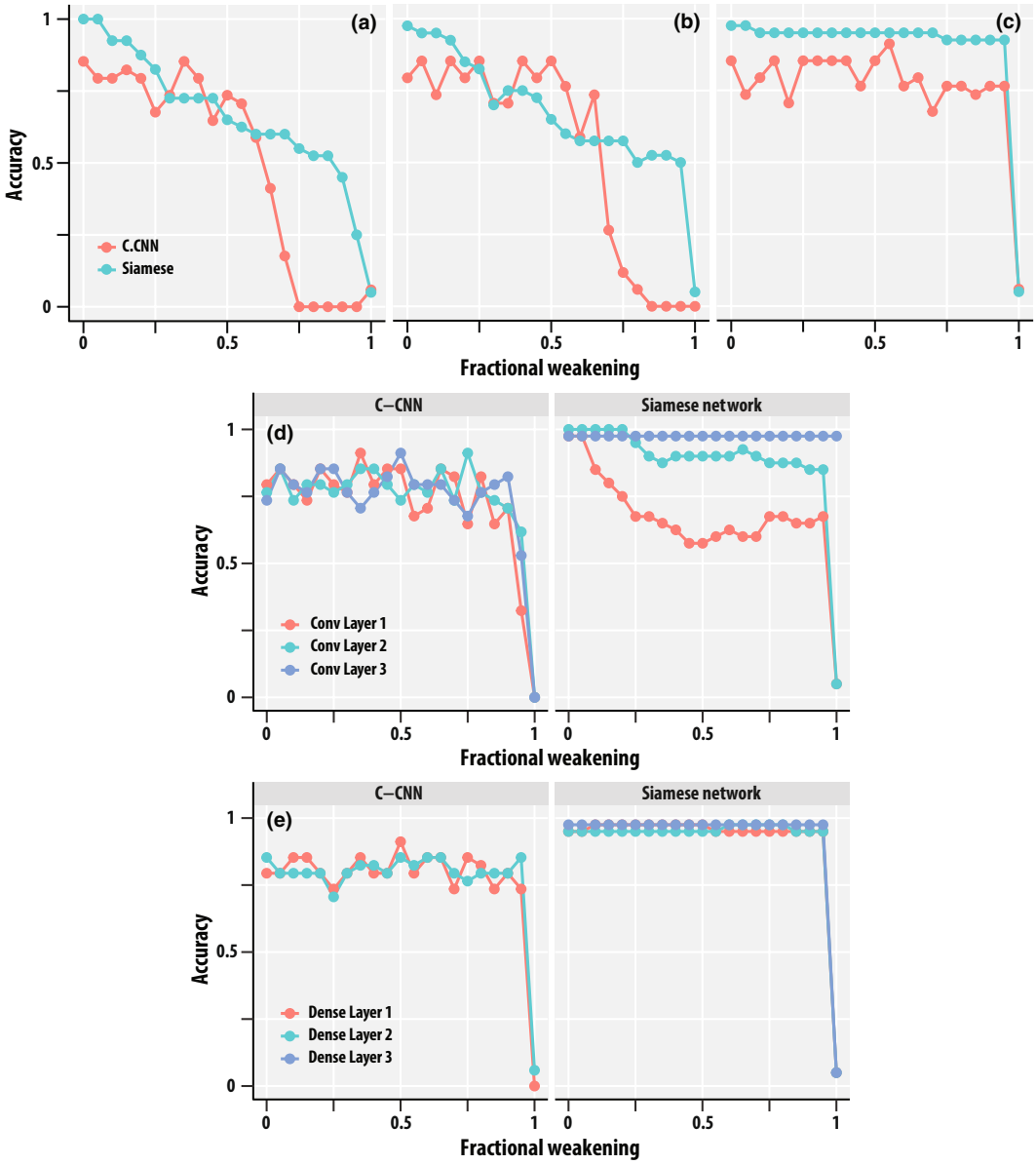


Figure 3 Effects of perturbing network weights on facial recognition accuracy. (**a**) Shows the effects of weakening the weights in every layer of the C-CNN (salmon lines and circle) and SNs (teal lines and circles). (**b**) shows the effects of weakening the weights of all the edges in all the convolutional layers in the C-CNN (salmon lines and circle) and Siamese networks (teal lines and circles). (**c**) Shows the effects of weakening the weights of all the edges in all the dense layers in the C-CNN (salmon lines and circle) and SNs (teal lines and circles). The facet plots in (**d**) shows the effects of weakening the weight of the edges in either convolutional layer 1 (salmon lines and circles), convolutional layer 2 (teal lines and circles), or convolutional layer 3 (blue lines and circle) for the C-CNN (left graph) and SNs (right graph). The facet plots in (**e**) shows the effects of weakening the weight of the edges in either dense layer 1 (salmon lines and circles), dense layer 2 (teal lines and circles), or dense layer 3 (blue lines and circles) for the C-CNN (left graph) and SNs (right graph). A fractional weakening value of zero corresponds to no changes to the trained network weights and 1 corresponds to complete elimination of the weights. C-CNN, convolutional-classification neural network; SN, Siamese network. [Color figure can be viewed at wileyonlinelibrary.com]

The accuracy of the SN declined gradually across the entire range, but the accuracy dropped rapidly when nearly all the nodes were deleted. C-CNN accuracy declined rapidly when as few as 10% of the nodes were lesioned. **Figures 4b** and **c** show the effects of random lesioning that were restricted to the convolutional layers and dense layers, respectively. In **Figure 4b**, increasing loss of nodes in the convolutional layers caused declines in accuracy that were generally like those in **Figure 4a**, except that the declines were right shifted. The face recognition accuracy of the SN network was more robust to effects of lesioning on the dense layers (**Figure 4c**) than the C-CNN whose accuracy declined gradually. **Figures 4d** shows the effects of lesioning that were limited to first, second, and third convolutional layers of C-CNN and SN, respectively. Both C-CNN and SN were more sensitive to random lesioning of the first convolutional layer, which is proximal to the input. Interestingly, the C-CNN showed markedly increased noise in accuracy with increased node deletion that was not observed for SN. **Figure 4e** shows that C-CNN accuracy was more sensitive to lesioning in

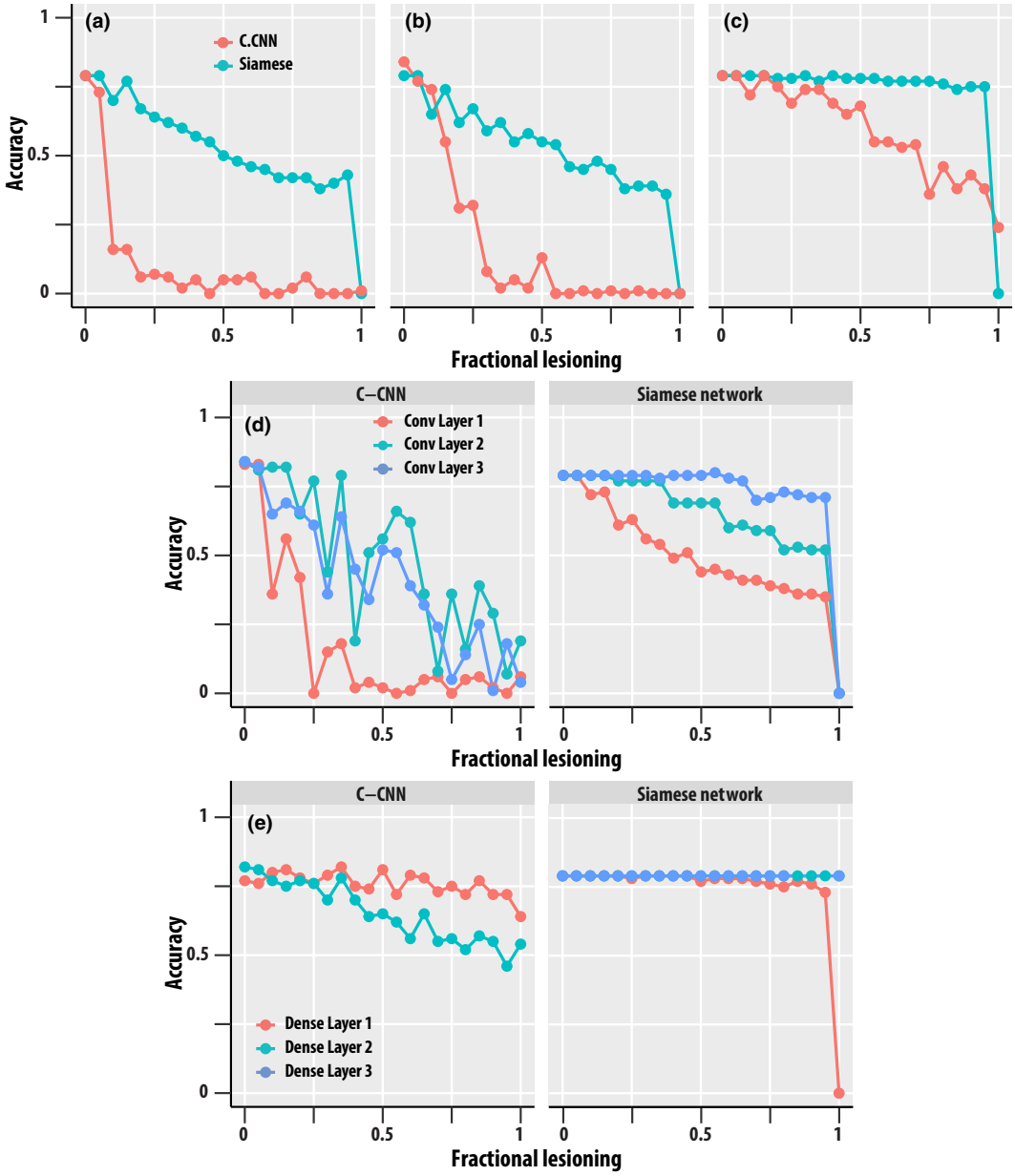


Figure 4 Effects of random lesioning of network nodes on facial recognition accuracy. (**a**) Shows the effects of lesioning every layer of the C-CNN (salmon lines and circles) and SNs (teal lines and circles). (**b**) Shows the effects of lesioning in all the convolutional layers in the C-CNN (salmon lines and circles) and SNs (teal lines and circles). (**c**) Shows the effects of lesioning in all the dense layers in the C-CNN (salmon lines and circles) and SNs (teal lines and circles). The facet plots in (**d**) shows the effects of lesioning in either convolutional layer 1 (salmon lines and circles), convolutional layer 2 (teal lines and circles), or convolutional layer 3 (blue lines and circles) for the C-CNN (left graph) and SNs (right graph). The facet plots in (**e**) shows the effects of lesioning in either dense layer 1 (salmon lines and circles), dense layer 2 (teal lines and circles), or dense layer 3 (blue lines and circles) for the C-CNN (left graph) and Siamese networks (right graph). A fractional lesioning value of zero corresponds to no lesioning in the trained network and 1 corresponds to complete lesioning. C-CNN, convolutional-classification neural network; SN, Siamese network. [Color figure can be viewed at wileyonlinelibrary.com]

dense layer 2 vs. dense layer 1; SN was relatively robust to lesioning in all dense layers.

Comparing effects of weakening vs. lesioning. For SN, the random lesioning (Figure 4) and weakening results (Figure 3) were qualitatively concordant; in both sets of experiments, the decreases in SN accuracy were gradual. C-CNN accuracy was degraded to a greater extent by random lesioning than weakening; the convolutional layers of the C-CNN were particularly affected by random lesioning. Overall, the results indicate greater robustness of the SN.

Contribution of individual facial landmarks

The contributions of individual facial landmarks were assessed for the SN (Table S1), which had greater robustness in our previous experiments. Masking an eye (46.2%–59.0 accuracy) caused a greater loss in accuracy than masking the mouth (76.9% accuracy) compared with the control (no masking, 96.3%). However,

decrease in accuracy from masking an eye and the mouth (35.9% accuracy) was comparable to masking both eyes (35.9% accuracy). The accuracy was lowest when both eyes and the mouth were masked (28.2% accuracy).

Model evaluation with real-world data

AD clinical outcomes and network failure quotient. The model simulation results were assessed using the ADNI data set for AD. The demographic characteristics of the ADNI sample are summarized in **Table S2**. The ADNI data set included neuropsychological test data from the Clinical Dementia Rating Sum of Boxes (CDRSB) scale for measuring severity of dementia, the AD Assessment Scale (ADAS) Cognitive-11 (ADAS-11), ADAS Cognitive-13 (ADAS-13), and Mini-Mental State Examination (MMSE) scales for cognitive function, and the Functional Activities Questionnaire (FAQ).

Because our simulations were focused on neural network edge weakening and deletion experiments, we selected the NFQ,[16] a robust biomarker of large-scale network failure derived from task-free fMRI. **Figure 5a** summarizes NFQ in the cognitively normal (CN), subjective memory complaints (SMCs), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and AD diagnosis groups at baseline. NFQ values were progressively worse in the EMCI, LMCI, and AD groups, which is consistent with breakdown of neural signaling pathways. NFQ also increased across increasing age tertiles.

Figure 5b–f compare the effect of network weakening as assessed by NFQ quintiles with key outcome measures used in AD clinical trials (i.e., CDRSB, ADAS-11, ADAS-13, MMSE, and FAQ). The strong associations of NFQ with AD cognitive outcome measures underscore the importance of network integrity, which is a key factor in our AI model.

The pattern of deterioration of CDRSB, ADAS-11, ADAS-13, MMSE, and FAQ scores with increased NFQ quintiles was gradual and resembled the patterns in SN simulations. We did not find evidence for the sharp deterioration in outcomes observed in C-CNN simulations.

Biomarkers and AD clinical outcomes. The progression of AD clinical symptoms is preceded by temporally ordered changes in prognostic CSF and MRI biomarkers.[17] Abnormalities in CSF

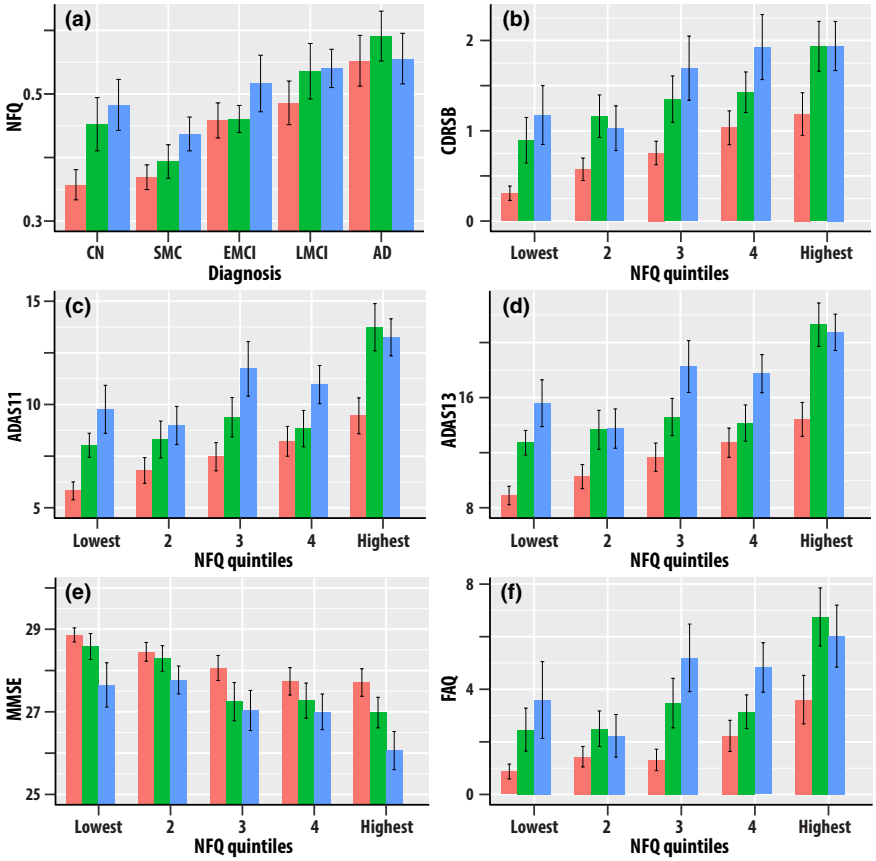


Figure 5 (**a**) Shows the mean values of the network failure quotient (NFQ) in cognitively normal (CN), subjective memory complaints (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD) groups at baseline. (**b–f**) Shows the dependence of NFQ quintiles on the key neuropsychological outcome measures: clinical dementia rating scale-sum of boxes score (CDRSB, **b**), ADAS-11 (**c**) is the 11-item cognitive subscale of the Alzheimer's disease assessment Scale (ADAS), ADAS-13 (**d**) is the 13-item cognitive subscale of ADAS, MMSE (**e**) is the mini-mental status score, and the functional activities questionnaire (FAQ) (**f**). The bars represent mean values, and the error bars are standard errors. The highest and lowest quintiles of NFQ are indicated. Separate bar colors are used for lowest (red bars, tertile 1 < 70.7 years), intermediate (green bars, 70.7 years ≤ tertile 2 < 76.5 years), and highest (blue bars, tertile 3 ≥ 76.5 years) tertiles of age at baseline. [Color figure can be viewed at wileyonlinelibrary.com]

Aβ42 occur in the prodromal phase of AD, followed by increases in CSF tau and phospho-tau-181. Atrophy resulting from the neurodegeneration triggered by pathological amyloid and tau deposition occurs early in the entorhinal cortex and spreads to the hippocampus at the mild cognitive impairment (MCI) stage. Subsequently, atrophy occurs at the fusiform gyrus, which is important for face recognition. Memory impairments precede the occurrence of dementia in AD.

Tau-amyloid β42 (tau-Aβ42) ratio was used as a CSF biomarker because tau (and phospho-tau-181, which behaves similarly) is associated with MRI outcomes and Aβ42 decreases in AD. Baseline tau-Aβ42 ratio, and MRI-derived entorhinal cortex, hippocampus, and fusiform gyrus volumes in the CN, SMC, EMCI, LMCI, and AD diagnosis groups were worse in the more severe disease groups (**Figure 6a–d**).

Memory impairment, as assessed by the Rey Auditory Verbal Learning Test percent forgetting score (RAVLT%F), was greater in the SMC, EMCI, LMCI, and AD groups (**Figure 6e**). The CDRSB score of dementia severity was greater in the LMCI and AD groups (**Figure 6f**).

RAVLT%F worsened with increasing tau-Aβ42 ratio and decreasing entorhinal cortex and hippocampus volumes quintiles (**Figure 6g–i**); RAVLT%F dependence on fusiform gyrus volume quintiles was weaker in comparison, which is attributable to the role of this brain region in face recognition not memory (**Figure 6j**). CDRSB also worsened with higher tau-Aβ42 ratio, and lower entorhinal cortex and hippocampus volumes quintiles (**Figure 6k–m**). Worsening of CDRSB occurred primarily at the lowest quintile of fusiform gyrus volume (**Figure 6n**) likely because it is downstream of other prognostic MRI biomarkers in AD pathophysiology.

DISCUSSION

In this research, we evaluated the counterintuitive and contrarian strategy of systematically degrading the performance of effective AI methods to obtain model-based insights into face recognition deficits in neurodegenerative diseases.

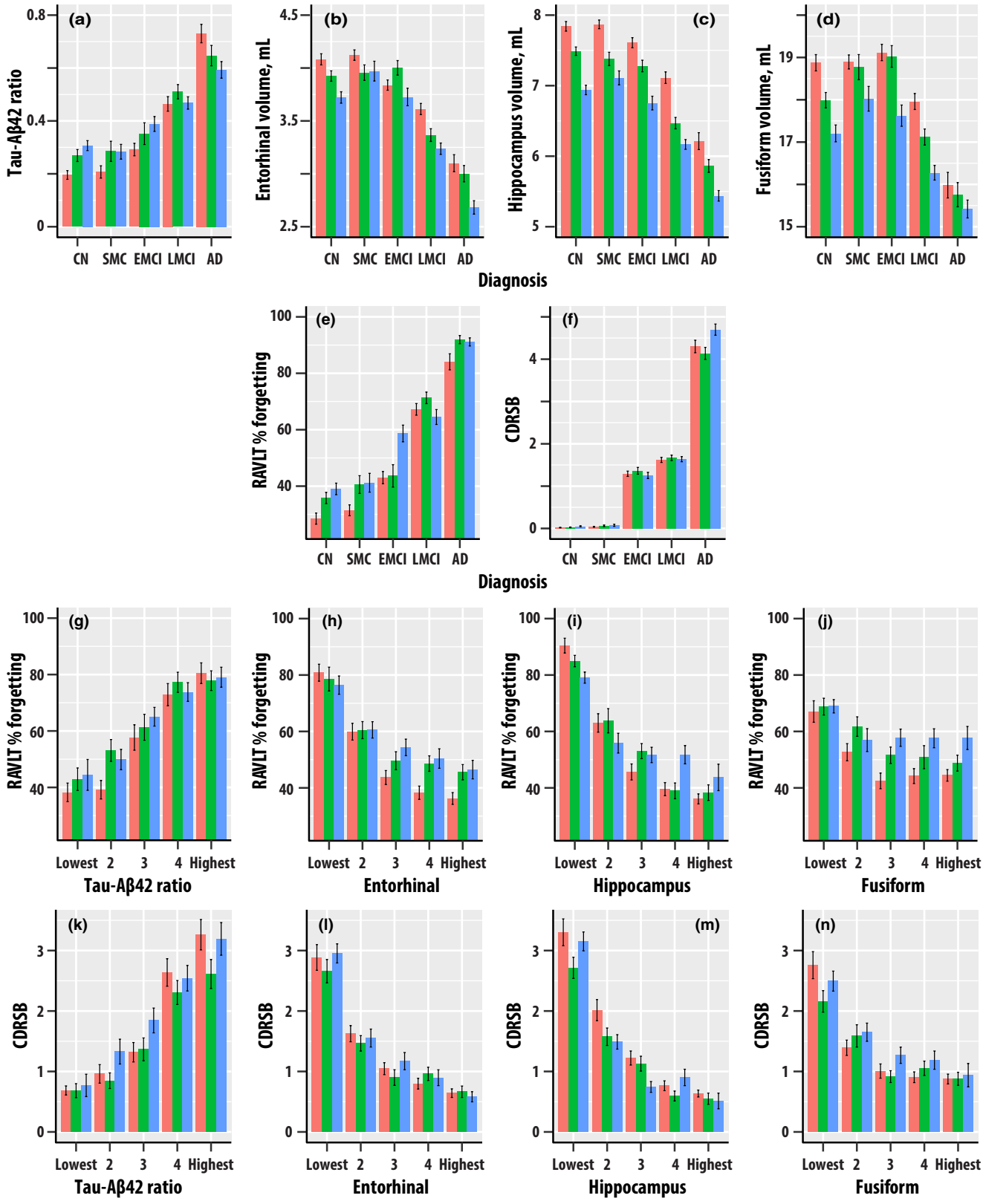
It was challenging to obtain *in vivo* measures of network failure to assess our approach. We selected the NFQ as a measure of network failure; because NFQ is derived from resting fMRI, it is relatively independent of the other AD-relevant neuropsychological measures in the ADNI data set.[15] The US Food and Drug Administration requires the primary outcome measure in AD drug trials to include cognition and functioning.[18] CDRSB, which measures both cognition and functioning, was the primary outcome in the trials of the anti-Aβ antibodies, lecanemab,[19] and aducanumab,[20] which received accelerated marketing approvals; ADAS-11 or ADAS-13, and MMSE were secondary outcome measures. We included the FAQ, as it explicitly assesses functioning. We also evaluated MRI data on the fusiform gyrus, which is important for face recognition. However, it would have been ideal to have a large data set with neuropsychological assessments of face recognition in prosopagnosia and AD.

Perceptual psychology experiments investigating face recognition by masking features in photographs have found that accuracy decreases substantially when eyes are masked.[21–24] Our results with masking of images are concordant with these findings. Interestingly, masking eyes has long been used to protect the identity of patients in medical photographs.

The size of our faces data set compares favorably with the estimated value of 150 people for Dunbar's number, the number of faces an average human can recognize by name.[25] Dunbar's number was obtained from evolutionary studies investigating social group size vs. neocortical volume[26] but it has empirical support from other contexts.[27] There is interindividual variability, however, and the Dunbar's number estimate has very wide confidence intervals.[25] The number of faces a person can recognize (~5,000) is much greater than Dunbar's number.[25]

Public domain data sets, such as ADNI, have been a catalyst and accelerant for AI research in AD. Hojjati *et al*. used a neural network-based regression approach to predict the associations of MRI and PET biomarkers with cognitive decline in the ADNI data set.[28] The best predictors of ADAS-13 and CDRSB scores were entorhinal cortex and hippocampus volumes and fluorodeoxyglucose-PET intensities of the angular gyrus, temporal gyrus, and posterior cingulate regions. Stricker *et al*.[29] found that mirrored cascaded architectures had superior performance characteristics vs. feed-forward neural networks in simulating word learning deficits. Grassi *et al*.[30] used a diverse range of ensemble machine learning to predict progression of patients with MCI to AD using demographic factors and neuropsychological test scores. Yang *et al*.[31] used a semi-supervised generative adversarial network to identify atrophy progression pathways in AD. Our research goals and experimental strategy differ fundamentally and distinctively from these prior AI works in AD. The underlying strategy was motivated by the clinical pharmacology work of Hwang *et al*.[32] who leveraged systematic network disruption for drug target identification. Li *et al*.[33] constructed GPT-D, a degraded version of the GPT-2 transformer-based natural language generator, and demonstrated that GPT-D was capable of

Figure 6 (**a–d**) Respectively, show the bar graphs of tau to amyloid β42 (Aβ42) tau to amyloid β42 (Aβ42) ratio, entorhinal cortex volume, hippocampus volume, fusiform gyrus volume, which are prognostic CSF and MRI biomarkers in cognitively normal (CN), subjective memory complaints (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD) groups at baseline. (**e, f**) Are the corresponding bar graphs for the Rey Auditory Verbal Learning Test (RAVLT) percent forgetting score, which measures memory, and the Clinical Dementia Rating Scale Sum-of-Boxes (CDRSB) scores, which is a dementia severity scale. The bar graphs in (**g–j**) plot the dependence of RAVLT percent forgetting score on quintiles of tau to Aβ42 ratio, entorhinal cortex volume, hippocampus volume and fusiform gyrus volumes. The bar graphs in (**k–n**) plot the dependence of CDRSB score on quintiles of tau to Aβ42 ratio, entorhinal cortex volume, hippocampus volume and fusiform gyrus volumes. The bars represent mean values, and the error bars are standard errors. The highest and lowest quintiles are indicated. Separate bar colors are shown for lowest (red bars, tertile 1 < 70.7 years), intermediate (green bars, 70.7 years ≤ tertile 2 < 76.5 years), and highest (blue bars, tertile 3 ≥ 76.5 years) tertiles of age at baseline. CSF, cerebrospinal fluid; MRI, magnetic resonance imaging. [Color figure can be viewed at wileyonlinelibrary.com]



synthesizing speech patterns similar to patients with dementia.[34] Adamczyk utilized a degradation strategy that used weight scrambling during training and altered activation functions to emulate learning disability.[35]

We acknowledge that there are vast differences between AI deep learning neural networks and the human brain. The human brain has greater structural complexity, versatility, and functional capabilities. The AI method is reductionist and likely overly

parsimonious: for example, our current AI model is trained for a single task – face recognition. However, neurological diseases can cause impairments in multiple cognitive domains and involve interactions between different brain regions. Given the complexity of the brain, AI models should be used with caution as they could be simplistic and even inappropriate.

We evaluated the C-CNN and SN architectures reasoning that different approaches for solving a given task can have convergent needs for certain shared features. Both C-CNN and SN use convolutional layers for feature extraction but use distinct strategies. The C-CNN architecture has a single neural network and uses classification, whereas the SN, which was more robust to perturbation, has twin networks and uses pairwise similarity. We attribute the performance differences to these distinctive architectures and recognition strategies. In disease modeling, functional robustness resulting from network architecture and topology could relate to the variability of age of symptom onset, and to physical resilience, the ability to maintain function during aging and disease.[36]

A potential criticism is that the weakening and lesioning experiments were conducted in a subset of 20 subjects to manage computational burden given the number of comparisons, weakening/lesioning levels, replicates, and network layers. To assess generalizability differences, we conducted the weakening (**Figure 3a**) and lesioning (**Figure 4a**) experiments for the C-CNN with all 200 subjects and found similar results (data not shown). As loss of accuracy is to be expected when trained AI models are perturbed, the findings from our weakening and lesioning experiments might be inadvertently viewed as unsurprising. However, because AI methods differ in their face recognition effectiveness, robustness, algorithmic frameworks, and architecture, the performance will deteriorate at different rates when perturbed, and certain components in the architecture may be important determinants of accuracy. For example, models that are too parsimonious may be sensitive to small perturbations whereas models with excessive complexity may not generalize. The utility of our strategy is derived from the investigating the dependence of accuracy loss on the extent of weakening and lesioning. An additional limitation is that our current approach is not a disease progression model because time is not an explicit predictor in the modeling: the CN, SMC, EMCI, LMCI, and AD groups used are disease stages with increasing severity of cognitive impairment. The time to progression between stages does not occur at the same rate and has interindividual variability. The variability is only partly explained by neurodegeneration measures from structural MRI (e.g., entorhinal cortex and hippocampal atrophy) because brain reserve differences, which are harder to measure, also contribute.[17]

Our computational experiments provide proof-of-concept for the strategy of perturbing neural networks to obtain models for neurological diseases. The AD results show that failure of neural signaling networks due to neurodegeneration can account for cognitive deficits, such as face recognition, that are important for patient functioning. This may represent an innovative strategy for building AI models for other cognitive outcome measures. Rigorous research and scientific equipoise are warranted regarding the potential of AI models for clinical applications.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr Carlos E. Thomaz, Head of the Image Processing Lab (IPL), Department of Electrical Engineering, Centro Universitario da FEI, Sao Paulo, Brazil for granting permission to utilize the images in our publication. We also thank the Alzheimer's Disease Neuroimaging Initiative (ADNI) for creating a publicly available resource that enabled this research.

FUNDING

Grant MS190096 from the Department of Defense Multiple Sclerosis Research Program for the Office of the Congressionally Directed Medical Research Programs (CDMRP) to the Ramanathan Laboratory is gratefully acknowledged.

CONFLICT OF INTEREST

G.S. has no conflicts. M.R. received research funding from the National Science Foundation, Department of Defense, and the National Institutes of Health.

AUTHOR CONTRIBUTIONS

G.S. and M.R. wrote the manuscript. M.R. designed the research. G.S. and M.R. performed the research. G.S. and M.R. analyzed the data.

PATIENT CONSENT STATEMENT

Dr Carlos E. Thomaz, Head of the Image Processing Lab (IPL), Department of Electrical Engineering, Centro Universitario da FEI, Sao Paulo, Brazil, has granted permission to utilize the images in our publication.

PERMISSION TO REPRODUCE MATERIAL FROM OTHER SOURCES

Dr Carlos E. Thomaz, Centro Universitario da FEI, Sao Paulo, Brazil has granted permission to utilize the images in our publication.

1. Barton, J.J.S., Davies-Thompson, J. & Corrow, S.L. Prosopagnosia and disorders of face processing. *Handb. Clin. Neurol.* **178**, 175–193 (2021).
2. Lahiri, D. Prosopagnosia. *Cortex* **132**, 479 (2020).
3. McClelland, J.L. & Rumelhart, D.E. *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises* (MIT Press, Cambridge, Mass, 1988).
4. Kriegeskorte, N. & Douglas, P.K. Cognitive computational neuroscience. *Nat. Neurosci.* **21**, 1148–1160 (2018).
5. Fukushima, K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**, 193–202 (1980).
6. Bullmore, E.T. & Bassett, D.S. Brain graphs: graphical models of the human brain connectome. *Annu. Rev. Clin. Psychol.* **7**, 113–140 (2011).
7. Chye, Y. *et al*. Examining the relationship between altered brain functional connectome and disinhibition across 33 impulsive and compulsive behaviours. *Br. J. Psychiatry* **220**, 1–3 (2021).
8. Dwivedi, R. *et al*. Surgery for drug-resistant epilepsy in children. *N. Engl. J. Med.* **377**, 1639–1647 (2017).
9. Singh, S. 6CC555 – Face Recognition (computer program). Kaggle.com software repository 2022 <https://www.kaggle.com/code/ssingh9908/6cc555-face-recognition>.
10. CaLeSS0 (username). Siamese networks (computer program) *Github software repository* (Siamese neworks. <https://github.com/maticvl/dataHacker/blob/master/pyTorch/014_siameseNetwork.ipynb>, 2021).
11. Thomaz, C.E. FEI Face Database. <https://fei.edu.br/~cet/facedatabase.html> (2012). Accessed July 16, 2022.

12. Thomaz, C.E. & Giraldi, G.A. A new ranking method for principal components analysis and its application to face image analysis. *Image Vis. Comput.* **28**, 902–913 (2010).
13. Viola, P. & Jones, M. Rapid object detection using a boosted cascade of simple features. *Paper presented at: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001; 8–14 Dec. 2001* (2001).
14. King, D.E. Dlib-ml: a machine learning toolkit. *J. Mach. Learn. Res.* **10**, 1755–1758 (2009).
15. *ADNIMER: Alzheimer's disease neuroimaging initiative* [computer program]. Version R package version 0.0.12023.
16. Wiepert, D.A. *et al*. A robust biomarker of large-scale network failure in Alzheimer's disease. *Alzheimers Dement (Amst).* **6**, 152–161 (2017).
17. Jack, C.R. Jr. *et al*. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12**, 207–216 (2013).
18. Food and Drug Administration Early Alzheimer's disease: developing drugs for treatment: guidelines for industry. In *Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation* (Department of Health and Human Services, Silver Spring, MD: U.S, 2018).
19. van Dyck, C.H. *et al*. Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* **388**, 9–21 (2023).
20. Budd Haeberlein, S. *et al*. Two randomized phase 3 studies of Aducanumab in early Alzheimer's disease. *J. Prev Alzheimers Dis.* **9**, 197–210 (2022).
21. Keil, M.S. "I look in your eyes, honey": internal face features induce spatial frequency preference for human face processing. *PLoS Comput. Biol.* **5**, e1000329 (2009).
22. McKelvie, S.J. The role of eyes and mouth in the memory of a face. *Am. J. Psychol.* **89**, 311–323 (1976).
23. Peterson, M.F. & Eckstein, M.P. Looking just below the eyes is optimal across face recognition tasks. *Proc. Natl. Acad. Sci. USA* **109**, E3314–E3323 (2012).
24. de Haas, B. *et al*. Perception and processing of faces in the human brain is tuned to typical feature locations. *J. Neurosci.* **36**, 9289–9302 (2016).
25. Jenkins, R., Dowsett, A.J. & Burton, A.M. How many faces do people know? *Proc. Biol. Sci.* **285**, 20181319 (2018).
26. Dunbar, R.I.M. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **22**, 469–493 (1992).
27. Goncalves, B., Perra, N. & Vespignani, A. Modeling users' activity on twitter networks: validation of Dunbar's number. *PloS One* **6**, e22656 (2011).
28. Hojjati, S.H. & Babajani-Feremi, A. Alzheimer's disease Neuroimaging I. prediction and modeling of neuropsychological scores in Alzheimer's disease using multimodal Neuroimaging data and artificial neural networks. *Front. Comput. Neurosci.* **15**, 769982 (2021).
29. Stricker, J.L., Corriveau-Lecavalier, N., Wiepert, D.A., Botha, H., Jones, D.T. & Stricker, N.H. Neural network process simulations support a distributed memory system and aid design of a novel computer adaptive digital memory test for preclinical and prodromal Alzheimer's disease. *Neuropsychology* **29**, 10.1037/neu0000847 (2022).
30. Grassi, M. *et al*. A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to Alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures. *Front. Neurol.* **10**, 756 (2019).
31. Yang, Z. *et al*. A deep learning framework identifies dimensional representations of Alzheimer's disease from brain structure. *Nat. Commun.* **12**, 7065 (2021).
32. Hwang, W.C., Zhang, A. & Ramanathan, M. Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clin. Pharmacol. Ther.* **84**, 563–572 (2008).
33. Radford, A., Jeffrey, W., Child, R., Luan, D., Amodei, D. & Ilya, S. Language models are unsupervised multitask learners. *OpenAI Blog.* **1**, 9 (2019).
34. Li, C., Knopman, D., Xu, W., Cohen, T. & Pakhomov, S. GPT-D: inducing dementia-related linguistic anomalies by deliberate degradation of artificial neural language models. *ArXiv* (2022):2203.13397.
35. Adamczyk, J. Neural network degeneration and its relationship to the brain. *ArXiv* (2020):2008.000053.
36. Merchant, R.A., Aprahamian, I., Woo, J., Vellas, B. & Morley, J.E. Editorial: Resilience and successful aging. *J. Nutr. Health Aging* **26**, 652–656 (2022).