Predicting Alzheimer’s disease in prodromal-stage by multimodal convolutional recurrent neural network

Author list

Affiliation

ABSTRACT

Alzheimer's disease (AD) is a chronic neurodegenerative disease that impairs patient's memory and cognition capability. Positron-emission tomography (PET) scan enables tracking temporoparietal hypometabolism and collecting morphological feature ***in vivo***. We proposed an interpretative multi-modalities model for classifying CN (cognitive normal), MCI (mild cognitive impairment), AD subjects. This model is constituted by a convolutional recurrent neural network that resolves time-sequential PET images for extracting spatiotemporal correlation, and a multilayer perceptron disposes of other correlation information of subjects such as the apolipoprotein E allele. Results indicate that our model can classify AD patients, MCI subjects with CN group for over 86% accuracy on 1671 subjects with 7280 test images.

Introduction

Alzheimer's disease (AD) is a progressive and irreversible brain degeneration, impairs memory and cognition capability. In contrast to other diseases, current medicines and treatments merely alleviate symptoms or retard their progression and cannot supply effective therapy. The number of AD patients was estimated to be approximately 30 million in 2015 (Vos et al., 2016), which has placed a substantial socioeconomic burden on those taking care of AD patients.

Pre-diagnosis of AD in the prodromal stage is essential to intervene in disease progression. Mild Cognitive Impairment (MCI) as the transition stage of CN and AD is segmented into two sub-classes: progressive MCI (pMCI) and stable MCI(sMCI). Abnormal accumulation of beta-amyloid (Aβ) peptide is widely believed to be the underlying mechanism of pathological and clinical changes seen in AD (Cummings, 2004). Initially, the function and structure of the transentorhinal, entorhinal cortex, and hippocampus will be changed (Delacourte et al., 1999) in the prodromal stage of AD, when the earliest impairments manifested with episodic memory deficits. Within this stage, some patients show cognitive complaints with objective cognitive impairment sporadically, which is classified as MCI in clinical view (Petersen et al., 2001). In disease progression, the structure and function of the temporal cortex and later parietal and frontal association cortices will first deteriorate. Finally, primary motor and sensory cortices and the neocortex will also degrade (Delacourte et al., 1999). Symptomatically, patients progress from episodic memory loss to additional, progressive deficits in visuospatial and semantic abilities, mood and behavioral disorders.

For tracing those latent evolutions of brain atlas and hypermetabolism within disease progression, Positron-emission tomography (PET) is a powerful tool by using ^{18}F-fluorodeoxyglucose (FDG) and other markers of amyloid pathology to investigate temporoparietal hypometabolism, which has been suggested as a core biomarker for AD. Abnormal accumulation of Aβ peptide is concerned as the underlying mechanism of pathological and clinical changes seen in AD. In these studies, *in vivo* amyloid, pathology is assessed using cerebrospinal fluid (CSF) or through specific ligands by PET imaging.

Alzheimer's Disease Neuroimaging Initiative (ADNI) database attributes to many co-investigators from a broad range of academic institutions and private corporations. It provides large samples of subjects with clinical test and assessment results, such as MR imaging, PET imaging, relevant biomarkers, and clinical or neuropsychological assessment. With the ADNI database consummating, deep learning methods such as convolutional neural networks have been employed to extract the features of medical images for classification. A convolutional neural network as a representative architecture has been applied to learn latent features and discrimination for image analysis and feature retrieval.

In AD etiology, disease progression has been associated with structural change in the hippocampal formation, entorhinal cortex, subcortical structures, and parahippocampal white matter. To investigate those regions of interests (ROIs) atrophy by features transformation and metabolism variation along with AD progression, we need a high-performance model with sufficient depth for distilling high-level features from images where those latent and complex features can be extracted by hierarchical learning. Typically, convolutional neural networks are stacked by convolutional and pooling layers followed by fully connected and softmax layers. The Stochastic depth improves deep residual network training by randomly dropping layers during training, facilitating residual neural networks (ResNets) to obtain superior information and gradient flow. However, in those legacies, feedforward architectures gradient will vanish during training over time. It means features of each ROI may vanish along with feedforward propagation. Notwithstanding that 3D PET brain image provides affluent information (256 × 256 × 256 voxels) for learning, its onerous load in computation in image preprocessing, training, and inference tasks restricts technique development. Therefore, we decide to employ a specific slice of PET images for obtaining the correlation information for learning, which the group-wise patterns of amyloid deposition and morphological changes among different regions are used to assist the model in performing signal decoding in the neural network. This reflects underlying short/long-range metabolic connectivity and possible pathological progression pathways of the hippocampus, ventricle, entorhinal cortex, and temporal lobe.

Different modalities of biomarkers reflect the AD-related pathological changes in different aspects. Thus there may be complementary information among several modalities. Combining multimodal biomarkers would provide more information and improve the accuracy of AD prediction. A simple way to fuse different modalities is to concatenate multimodal features and feed them into a classifier directly (Kohannim et al., 2010; Walhovd et al., 2010; Westman et al., 2012). However, it can lead to bias of the modality with a more significant number of features. Mapping these multimodal features into a kernel space before concatenation (Hinrichs et al., 2011; Zhang et al., 2011; Young et al., 2013) performs better but sensitive to the weight assigned to each modality. In recent years, deep learning architecture has been employed to extract multimodal feature representations. Liu et al. (2015) used stacked auto-encoders and a zero-mask strategy to fuse MRI and PET data. Suk et al. (2014) proposed a joint feature representation of MRI and PET with a multimodal deep Boltzmann machine. Liu et al. (2018) constructed multiple deep three-dimensional (3D) convolutional neural networks to transform MRI and PET images into compact high-level features. These deep learning-based methods achieved promising results in the classification of AD/NC, but the accuracy of classifying pMCI/sMCI was just 74.58% (Suk et al., 2014). To exploit the complementarity across multimodal data, Tong et al. (2017b) employed a non-linear graph fusion that achieved better results in the diagnosis of AD and a three-way classification of AD/MCI/NC than the approaches based on a linear combination, but the classification of pMCI from sMCI was not validated. Although all of these multimodal data-based methods achieved promising results in the diagnosis of AD, the performance of AD prediction needs to be further improved for clinical use with the help of an efficient fusion of multimodal biomarkers.

This paper proposed a multi-modalities machine learning model that combines a hybrid convolutional recurrent neural network to extract latent features and high-level information to find a longitudinal correlation and a multilayer perception for resolving its corresponding biomarkers and identification information. This multiple-input single-output (MISO) significantly reduces the computational burden in image processing and image training without influencing inference results. Slice of PET image with amyloid load, predominant structure, and latent is used. We fed different stage PET images of identity region for a longitudinal study. It leads the model to learn various structures, formations, and amyloid loads of target regions in time-sequence. A multilayer perceptron (MLP) is a classical feedforward artificial neural network consisting of an input layer, a hidden layer, and an output layer. Demographic information, clinical exam, ROIs analysis results of MRI and PET are fed into MLP leading towards co-learning with the Convolutional Long-short term memory (ConvLSTM) network for processing related information from multiple modalities.

Materials and Methods

Data for this work is from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu). Data from ADNI 1, GO (Grand Opportunities), 2, 3 cohorts are used in this paper, and all subjects with preprocessed PET imaging were extracted that were cognitively normal or diagnosed with MCI or AD. In total, there are 1671 unique subjects (540 CN, 876 MCI, and 255 AD) with longitudinal follow-up resulting in 7280 imaging volumes and 3 × 7280 two-dimensional axial scans. The subjects are from 55-96 years of age and from 58 centers worldwide.

We propose a three-slice ROI hierarchy learning matrix to improve the classification performance of volumetric measurements. This enables us to release information that corresponds to brain atlas, such as the entorhinal cortex, temporal lobe, and parahippocampal white matter for machine learning. Florbetapir (AV-45) is considered an effective measure to assess amyloid load in patients with Alzheimer's disease (AD) at dementia stages and is capable of binding amyloid in the early stages of AD.

Image selection and preprocessing

The PET-image set is obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) across ADNI 1, ADNI GO(Grand Opportunities), ADNI 2, and ADNI 3 databases (dynamic 3D scan of six 5-min frames 30-60 min post-injection) with dates ranging from May 2005 to January 2017 collecting 1671 patients' clinical images for retrospective analysis.

The FDG-PET images (via averaging counts of angular, temporal, and posterior cingulate regions) in this study were preprocessed using a series of steps to mitigate inter-scanner variability and obtain FDG-PET data with a uniform spatial resolution and intensity range for further analysis. Preprocessing steps includes dynamic co-registration of images acquired in consecutive time frames, averaging, reorientation along the anterior-posterior commissure, and filtering with a scanner specific filter function to produce images of a uniform isotropic resolution of 8 mm full width at half maximum Gaussian kernel. The approximate resolution of the lowest resolution scanners used in ADNI is adopted for smoothing image with the identical norm. Of the preprocessed image set, we select three slices of the axial image of the brain from 3D image (96×160×160) for subsequent learning. All images are processed by Smooth, Calculate Coregistration, Apply Coregistration, Average Frames, AC-PC Orient Baseline, Standardize to Baseline, Average Frames, Intensity Normalization, and Variable Smooth sequentially.

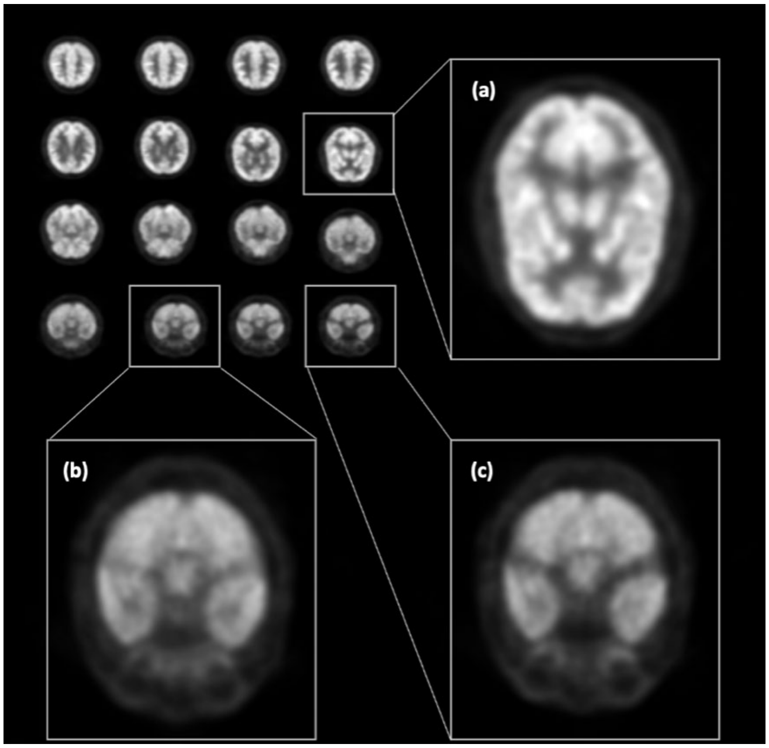


Figure 1. Example of fluorine ^{18}fluorodeoxyglucose PET images from Alzheimer's Disease Neuroimaging Initiative set preprocessed with the grid method for patients with Alzheimer disease (AD). One representative zoomed-in section was provided for each of three example regions of an individual patient.

The image selection hopes to primarily release information of relevant regions which are conductive to clinical trial. Notably, in disease progression, the structure and function temporal cortex and later parietal and frontal association cortices will deteriorate (Delacourte et al., 1999) and cause hippocampus and ventricle changed in structure and amyloid deposition.

Clinical Interpretation of Multi-Modality Model

Model interpretative are highly demanded in the clinical trial. Our scheme adopts a multi-modality neural network to trace amyloid hypometabolism and structure change by a convolutional recurrent neural network and extracts its corresponding biomarkers and clinical data by a multilayer perception from the ADNI database. ADNIMERGE is a clinical and biomarkers data package of the ADNI database that includes all subjects' characteristics, bio-information, image analysis results, and genetic data.

Apolipoprotein E Data Preprocessing

The apolipoprotein E (APOE) ε4 genotype is a decisive risk factor for late-onset Alzheimer's disease (AD). ADNI cohort previously reported significant baseline structural differences in APOE ε4 carriers as compared to non-carriers. There are three types of the APOE gene, called alleles: APOE ε2, ε3, and ε4. Everyone has two copies of the gene, and the combination determines APOE "genotype"— ε2/ε2, ε2/ε3, ε2/ ε4, ε3/ ε3, ε3/ ε4, or ε4/ ε4. The ε2 allele is the rarest form of APOE, and carrying even just one copy appears to reduce the risk of developing Alzheimer's by up to 40%. APOE ε3 is the most common allele and does not seem to influence risk. The APOE ε4 allele, present in approximately 10-15% of people, increases the risk for Alzheimer's and lowers the age of onset. Having one copy of ε4 (ε3/ ε4) can increase one's risk by 2 to 3 times while two copies (ε4/ ε4) can increase the risk by 12 times. In our case, we classify subjects as non-carrier, a single copy of ε4 allele (heterozygotes) carriers or muted copy of ε4 carriers (homozygotes) then fed it into the multilayer perception to examine the longitudinal effects of APOE genotype on brain morphology and amyloid metabolism by co-learning with ConvLSTM model.

Demographic data

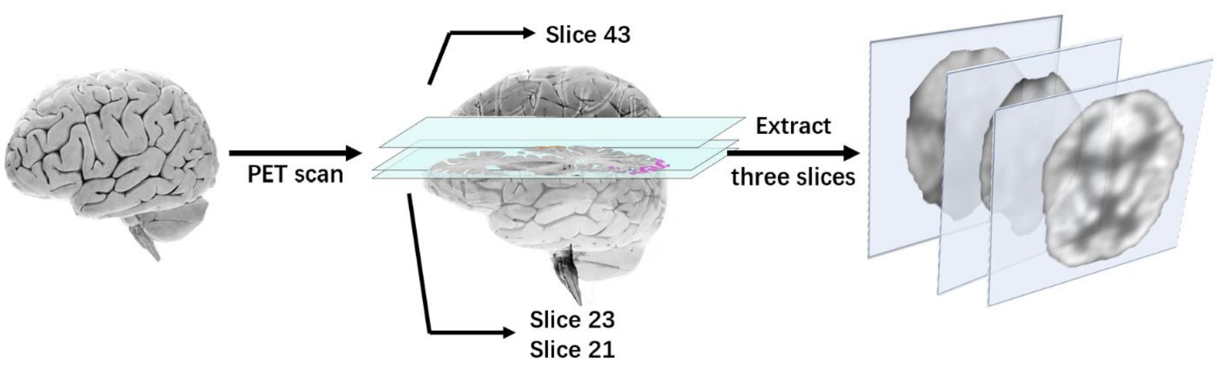
The image-analysis data was acquired from MR imaging analysis results which source from data dictionaries and method paper. It supplies reference volumes of each ROI to comparison learning with the image set in our model.

Table 1. The demographic information of subjects

Mean±SD	AD	CN	MCI
Count(F/M)	254(105/149)	403(209/194)	990(446/544)
AGE	74.9±8.0	72.4±7.2	73.7±6.0
EDUCATION	15.3±2.9	16.2±2.7	16.4±2.7
MMSE	19.9±5.2	26.6±4.5	29.1±1.1

In cells of the second row, the first number is the total number with numbers of females (F) and males (M) in brackets. AD, Alzheimer's disease; MMSE, Mini-Mental State Examination; CN, cognitive normal; SD, standard deviation;

Image Preprocessing



Model training

Model of Convolutional Long-Short term memory

Convolutional neural networks (CNNs) with several layers of convolution, pooling, and non-linear units have shown considerable success in computer vision tasks. Unlike most fully connected neural networks, weight sharing in CNN results in vastly different gradients in different layers. The advantage of RNN is to characterize the relationship between the current output of a sequence and the previous one for modeling the sequential progression. The differences and correlation of images for an individual are similar to the sequential information and useful for AD classification. Motivated by the success of RNN, this work proposes to build an RNN network to characterize the high-level correlation for disease classification.

In the past decades, the RNN performance was severely restricted due to training difficulty. Gradient mass and explosion are unsolved problems until the emergence of Long Short-Term Memory (LSTM). LSTM consists of a memory cell unit and three gate units, i.e., forget, input, and output gates. By updating the state of the memory cell through three gates, LSTM can discard irrelevant information and effectively capture the valuable information in sequence.

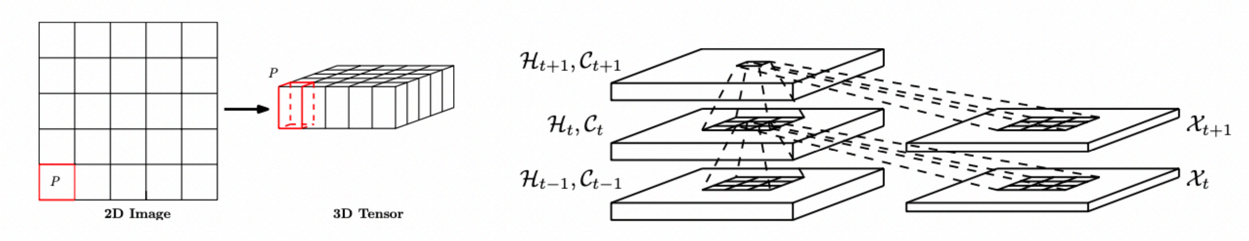

P
2D Image
P
3D Tensor
$\mathcal{H}_{t+1}, \mathcal{C}_{t+1}$
$\mathcal{H}_t, \mathcal{C}_t$
$\mathcal{H}_{t-1}, \mathcal{C}_{t-1}$
$\mathcal{X}_{t+1}$
$\mathcal{X}_t$

Figure 2. Scheme of transforming a 2D image into a 3D tensor and inner structure of ConvLSTM

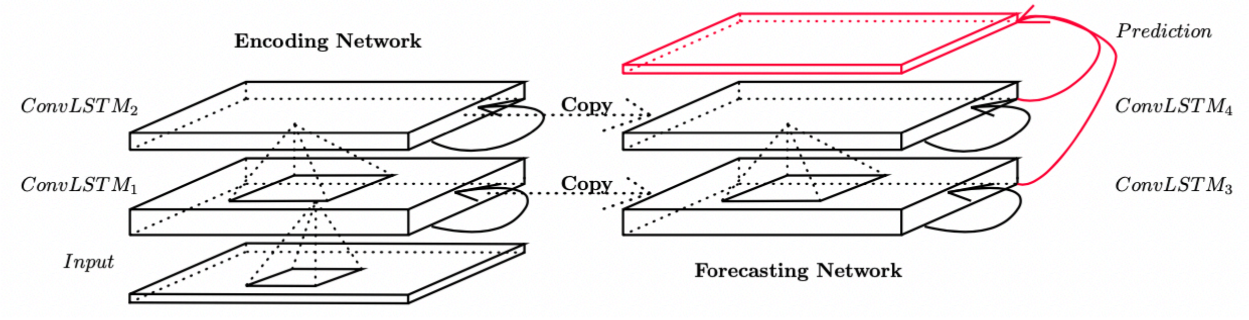


Figure 3. Decoding and forecasting network

ConvLSTM (Convolutional Long-short term memory) is a type of recurrent neural network for Spatio-temporal prediction along with convolutional structures in both the input-to-state and state-to-state transitions. The ConvLSTM determines the future state of a specific cell in the grid by the inputs and past states of its local neighbors, and possesses better spatiotemporal correlations, and consistently outperforms FC-LSTM (Fully connected-LSTM), with inputs $\mathcal{X}_1$,. . . , $\mathcal{X}_t$, cell outputs $\mathcal{C}_1$,. . . , $\mathcal{C}_t$, hidden states $\mathcal{H}_1$,. . . , $\mathcal{H}_t$,and gates i_t, f_t, o_t To alleviate gradient vanish or explosion and get a better picture of the inputs and states, the vector is supposed to be standing on a spatial grid. This spatiotemporal structure enables us to compare feature changed in each area in time-sequence. The critical equations of ConvLSTM are shown below, where '$*$'denotes the convolution operator and $' \circ '$ the Hadamard product:

$$\begin{aligned}
i_t &= \sigma(W_{xi} * \mathcal{X}_t + W_{hi}\mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i) \\
f_t &= \sigma\left(W_{xf}\mathcal{X}_t + W_{hf}\mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f\right) \\
\mathcal{C}_t &= f_t\mathcal{C}_{t-1} + i_t \circ \tanh\,(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c) \\
o_t &= \sigma(W_{xo}\mathcal{X}_t + W_{ho}\mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o) \\
\mathcal{H}_t &= o_t \circ \tanh\,(\mathcal{C}_t)
\end{aligned} \tag{1}$$

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \tanh \end{pmatrix} W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix} \tag{2}$$

$$c_t^l = f \odot c_{t-1}^l + i \odot g \tag{3}$$

$$h_t^l = o \odot \tanh\left(c_t^l\right) \tag{4}$$

The three vectors $i, f, o \in \mathrm{R}^n$ are binary gates to control each memory cell to update, reset to zero, or reveal its local state in the hidden vector. The activations of these gates are based on the sigmoid function and hence allowed to range smoothly between zero and one to keep the model differentiable. The vector $g \in \mathrm{R}^n$ ranges between -1 and 1 and is used to modify the memory contents additively. This additive interaction is a critical feature of the LSTM's design because, during backpropagation, a sum operation merely distributes gradients. This allows gradients on the memory cells to flow backward through time uninterrupted for extended time periods, or at least until the flow is disrupted with the multiplicative interaction of an active forget gate. Lastly, note that an implementation of the LSTM requires one to maintain two vectors (${h_t}^l$ and ${c_t}^l$) at every point in the network.

Therefore, a ConvLSTM can learn latent feature and correlation of each PET image and use its memory cells to remember long-range information in principle and concatenate high-dimensional correlation for tracking various attributes of the feature when it is currently processing.

Proposed model architecture

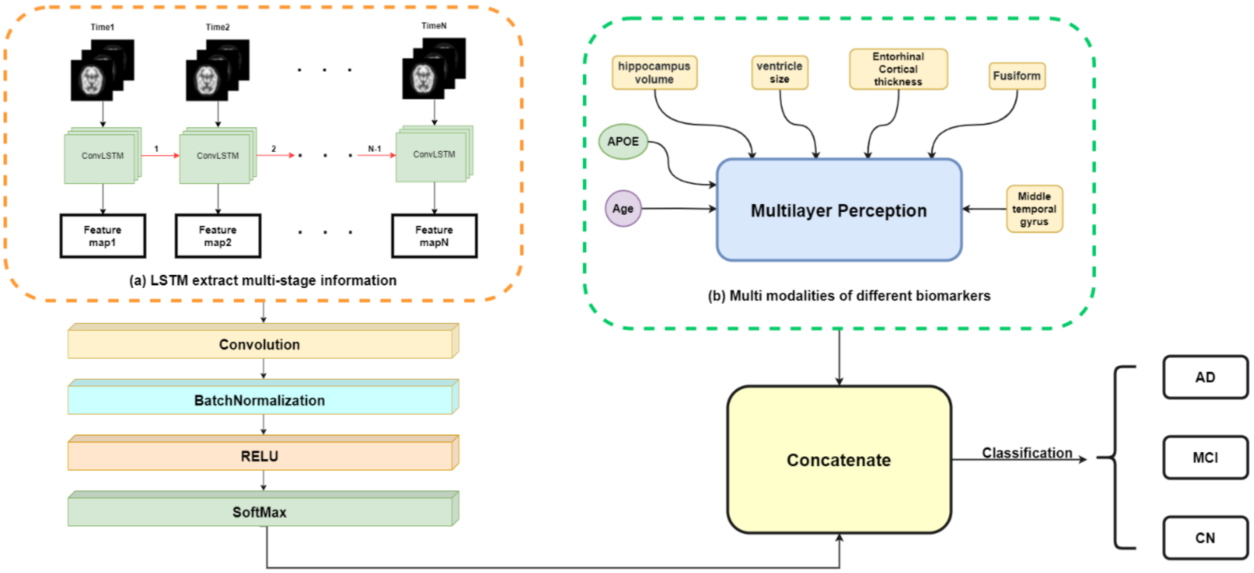


Fig 4. (a) Three Convolutional RNNs corresponding to each slice of PET scan, red array indicates that time-sequence information at different stages is transmitted by using LSTM. (b) A multilayer perception with seven factors inputs.

Figure 4. The architecture of the proposed model

The neural network consists of 4 main parts: 3 RNNs and 1 MLP. The 3 RNNs take in PET scans of different places of the brain and the MLP takes in biological information. Then there's a concatenation layer that concatenates all 4 parts together and makes the prediction. We chose 2 of the most recent PET scans for each patient and convert them into NumPy arrays. The RNN takes in the 4D array for each patient and passes it through a convolution LSTM layer which has 16 filters and a kernel size of (3,3). Then there is a batch normalization layer to prevent vanishing gradient problems. We then have maxpooling layer and dropout layer in order to prevent the overfitting problem. It is then flattened and connected with dense layers. The MLP takes in an array with 7 numbers of each patient, such as demographic information like age, genetic information like APOE, and image analysis information like volume of the temporal lobe, entorhinal cortex, ventricle, and hippocampus. It is then connected with dense layers and dropout layers in order to prevent overfitting problems, and eventually co-learning with the ConvLSTM network.

(pictures of RNN model summary and MLP model summary)

Results and Discussion

In total, 1671 subjects with 7280 images of each critical slice of PET images (160 × 160 × 3) from ADNI are used with a 70/20/10 split for training/validation/testing. Each volume is labeled with the diagnostic label from the clinical data provided with ADNI. Using training and validation data, a slice selection technique was designed to determine which slices to use for classification. The best performing slices were selected to perform the final AD classification on the hold-out test set. All models were developed on Keras-Tensorflow 2.3.0 and were loaded and trained onto a machine with Linux operating system (Ubuntu 18.04). The machine has two eight-core processors (Intel(R) Xeon(R) Gold 6126 CPU with 2.60GHz) and two NVIDIA Pascal Titan V100 graphical processing units. Convolutional neural network architecture is shown in figure 4.

Our model relies on non-parallel data in which two networks only shared categories or concepts without require to have shared instances. Non-parallel co-learning approaches can help to learn representations in object recognition with interpretability.There are 3 categories, CN, MCI, and AD. After 200 epochs of training, the overall testing accuracy for diagnosing is 86.7%. The clinical diagnosis of only AD can be as high as 90%, but our model can diagnose the three stages of the patients with an accuracy that is close enough to that.

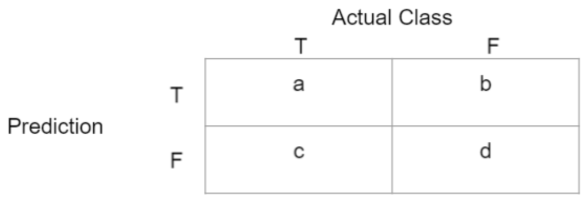
Will evaluate our classifier via additional four dimensions.

ROC = Receiver Operating Characteristic.

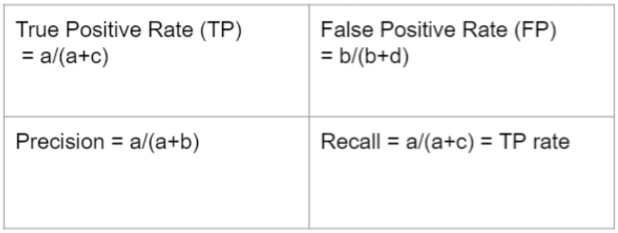
AUC = Area under the ROC Curve.

SEN= Sensitivity

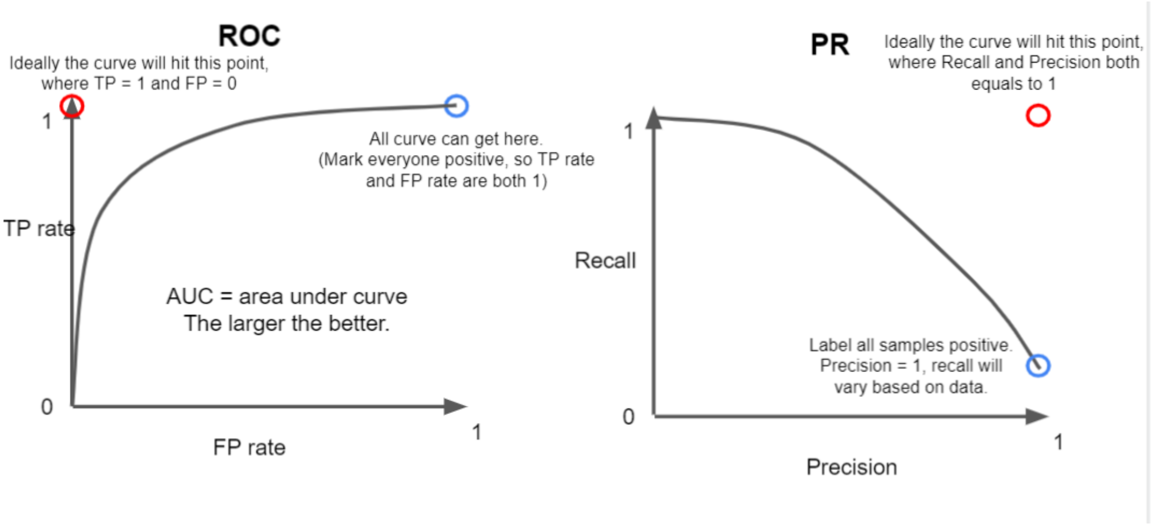
SPE= specificity



		Actual Class	
		T	F
Prediction	T	a	b
	F	c	d



True Positive Rate (TP) = a/(a+c)	False Positive Rate (FP) = b/(b+d)
Precision = a/(a+b)	Recall = a/(a+c) = TP rate


ROC
Ideally the curve will hit this point, where TP = 1 and FP = 0
1
All curve can get here. (Mark everyone positive, so TP rate and FP rate are both 1)
TP rate
AUC = area under curve
The larger the better.
0
FP rate
1
PR
Ideally the curve will hit this point, where Recall and Precision both equals to 1
1
Recall
Label all samples positive. Precision = 1, recall will vary based on data.
0
Precision
1

Conclusion

Alzheimer's Disease (AD), the most common type of dementia and one of the leading causes of death in America, impairs memory and other important cognitive functions. Through early diagnosis and clinical intervention in advance, even if the average age of Alzheimer disease onsets only delays 5 years, the cases can be reduced by up to 50%. However, it’ s always a big challenge to do the early diagnosing. The model we created and tested has shown promising diagnosing accuracy. We combined deep learning with human knowledge about diagnosing by selecting PET slices and adding biological information like APOE. Furthermore, although some other studies have achieved a higher accuracy with 3D brain scans, they require a lot more computation powers and longer training times. Our model needs less computation power because we only used slices from 3 different places for one person. Moreover, the cost of conducting such a diagnosis is also cheaper than normal since you don’ t need a radiologist to diagnose, and most health insurances should cover the cost of PET scans. With a larger dataset and more training, this model should be able to improve and maybe outperform radiologists and provide an opportunity for early intervention.

Limitations: uneven dataset. Missing data points. Real-life scenario: some biological information might miss, and the neural network might not be able to handle that.

Overall, the model shows the possibility of doing such tasks with lower computation power.

References

Disclosure

Acknowledgements