A Novel Predictive Model for Alzheimer's Disease Using Multi-view Brain Networks

by

Qianhao Wei

Bachelor of Science, China University of Petroleum (Beijing), 2019

Submitted to the Graduate Faculty of the

Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Qianhao Wei

It was defended on

December 8, 2020

and approved by

Liang Zhan, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

Feng Xiong, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

Jingtong Hu, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

Thesis Advisor: Liang Zhan, Ph.D., Assistant Professor, Department of Electrical and Computer Engineering

A Novel Predictive Model for Alzheimer's Disease

Using Multi-view Brain Networks

Qianhao Wei, MS

University of Pittsburgh, 2021

Alzheimer's disease (***AD***) is an irreversible and progressive disease that begins with mild memory loss possibly leading to loss of the ability to respond to the environment. Diffusion-weighted magnetic resonance images (DW-***MRI***) derived brain network has been widely used as an important tool to study ***AD***. Many methods have been proposed to reconstruct brain structural networks in the literature. However, due to the variance of each method, the resulted brain networks are distinct. Consequently, the downstream analyses and results are varied, which significantly affect the clinical implications of the conclusions. Moreover, due to the limitation of the available imaging data, it may also be a challenge to effectively capture highly discriminative features related to ***AD***, which can further affect the predictive model's performance. In this thesis, we address these challenges by proposing a new ***CNN*** (Convolution Neural Network) and ***LSTM*** (Long Short-Term Memory) mixed model for ***AD*** classification. Here we treat different networks from the same subject as multi-view data and apply our new model on two independent publicly available AD cohorts (202 subjects from the 2nd stage of Alzheimer's disease neuroimaging initiative or ADNI2 and 445 subjects from National Alzheimer's Coordinating Center or NACC). ADNI2 has four group data (normal control or NC, early mild cognitive impairment or EMCI, late mild cognitive impairment or LMCI and AD) while NACC has only three group data (NC, MCI, and AD). Our experimental results show that our new model can achieve around 96% accuracy for multi-group

classifications for both cohorts, significantly outperformed those baseline methods whose accuracy is around 75%.

Table of Contents

1.0 Introduction **1**

2.0 A Novel Network Based on Combined CNN-LSTM **4**

2.1 2D-CNN for Feature Extraction **4**

2.2 Long Short-Term Memory for Classification **6**

2.3 A Novel Network Based on Combined CNN and LSTM **8**

3.0 Experiment **9**

3.1 Data Description and Augmentation **9**

3.2 Experiment Setup and Parameter Settings **10**

3.3 Summary of Results **11**

4.0 Conclusion **17**

Bibliography **18**

List of Tables

Table 1 The number of patients in different stages in the ANDI2 dataset 10

Table 2 The number of patients in different stages in the NACC dataset..... 10

List of Figures

Figure 1 The main structure and process of the Novel Network based on CNN and LSTM 3

Figure 2 The Basic Architecture of Convolution Neural Networks 5

Figure 3 The basic structure of a single LSTM memory cell..... 7

Figure 4 The main workflow of the proposed method 8

Figure 5 The result of Accuracy test on NACC dataset based on combined CNN-LSTM .. 12

Figure 6 The result of Loss test on NACC dataset based on combined CNN-LSTM..... 12

Figure 7 The result of Accuracy test on ADNI dataset based on combined CNN-LSTM ... 12

Figure 8 The result of the Loss test on the ADNI dataset based on combined CNN-LSTM 13

Figure 9 Test Accuracy in ADNI dataset with the change of 113 ROIs..... 14

Figure 10 Test Accuracy in NACC dataset with the change of 113 ROIs 14

Figure 11 Test Accuracy in NACC dataset with the change of 86 ROIs 15

Figure 12 The structural connectivity networks of Top 10 related ROIs in the NACC dataset 16

Figure 13 The structural connectivity networks of Top 10 related ROIs in the ADNI dataset 16

1.0 Introduction

Alzheimer's Disease (AD) is an irreversible and progressive disease that begins with mild memory loss, which can result in severe cognitive dissonance and behavior disorder. Recent research suggests that various degrees of the progression of degenerative changes in brain networks result in different symptom[1]. Standard MRI techniques have been used to capture changes of brain integrity and connectivity of the white matter to distinguish patients in several stages (NC, MCI, and AD). Recent studies based on MRI can be roughly divided into two groups i) ROI-based feature analysis ii) structural and functional connectivity analysis[9]. Diffusion-weighted magnetic resonance images (DW-MRI) methods, including diffusion tensor magnetic resonance images (DT-MRI), also allows exploring and tracking water diffusion among the connecting cortical and subcortical regions of the brain, characterizing the progressive effects of AD on microstructure[2][3]. Moreover, whole-brain tractography is an application of DW-MRI that reconstruct white matter tracts, and it is sensitive to the progressive deterioration of the white matter. Hence these tractography methods have been widely used as tools to solve AD-related structural and molecular problems in the brain network, tracking disease progression, and observing how AD affects the brain[4][5][6].

The challenges for detecting AD based on artificial intelligence mainly have three main reasons i) low-related ROIs in different stages of AD ii) The complexity of medical images iii) the limited amount of data. With the rapid development of machine learning and clinical data, deep learning methods have been widely applied in research on brain disease, not just Alzheimer's Disease, performing well in the field of extracting features of medical images for disease detection. Several machine learning algorithms based on different types of classifiers, multi-modal models,

or feature selection methods have been proposed to classify AD and MCI from NC (Normal Control)[6][7][8]. Stacked auto-encoder (SAE) can combine the latent high-level feature information with the original feature representations and use multi-kernel SVM to improve the classification performance[10]. Multimodal stacked deep polynomial networks (MM-SDPN) have been proposed to effectively fuse and learn features from original ROI features for better performance on multimodal neuroimaging and other medical data[11]. And many studies use the zero-masking strategy to extract features from different modalities, putting them into deep SAE networks to combine multimodal features. After obtaining high-level features, using SoftMax regression to evaluate classification performance[12]. Moreover, many studies based on convolution neural network (CNN), using a stack of convolutional and pooling layers to find the compositional hierarchy of objects and learn spatial features of MRI for achieving better results in AD classification. The combination of Sparse autoencoder and 3D-CNN have been presented to capture variational anatomical shapes for multimodality learning[13]. Also, the LSTM (Long short-term memory) method has been used to connect the previous and present information for predicting the development of AD progression, instead of focusing on detecting the current stage[17].

Although these mentioned methods have shown their advantages in AD classification, there are still some factors that limit their performance. Some spatial information extracted from features of ROI will be lost when applying in the flattening layer. Moreover, more convolution layer needed to be stacked in the feature fusion stage for better classification results, which will be time-consuming due to huge computational cost.

Inspired by recent studies, a novel network consisting of CNN and LSTM as shown in figure 1 has been implemented for multi-class diagnosis of Alzheimer's disease. First, we need to normalize and transform the data which are based on nine different tractography methods. Then, the deep 2D-CNN is used to capture the highly discriminative features while LSTM is used to extract and remember the long dependence information which is related to AD progression[20]. Here we apply our new model on two independent publicly available AD cohorts (202 subjects from the 2nd stage of Alzheimer's disease neuroimaging initiative or ADNI2 and 445 subjects from the National Alzheimer's Coordinating Center or NACC). Our network represents feasibility and effectiveness in training and testing when using multi-view brain networks in different experiments.

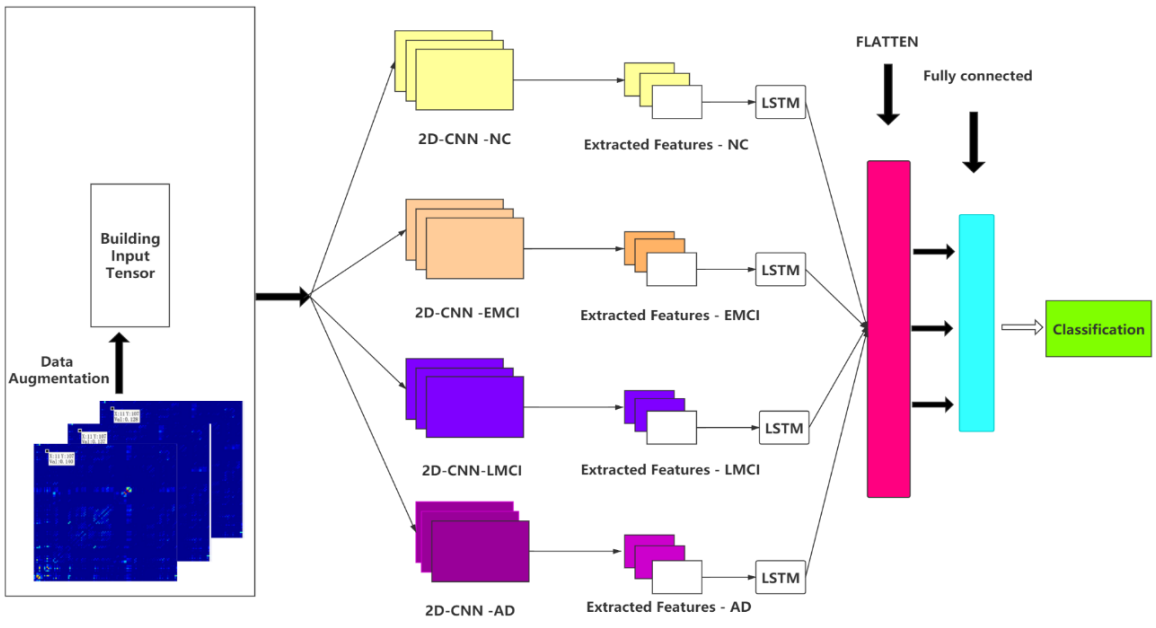

Figure 1 The main structure and process of the Novel Network based on CNN and LSTM

2.0 A Novel Network Based on Combined CNN-LSTM

The mixed structure involves the use of the convolutional neural network layer for extracting both low-level and high-level features, and the combination of CNN and LSTM to support sequence prediction. We propose one AD classification method based on multi-factor by integrating $\boldsymbol{CNN}$ and $\boldsymbol{LSTM}$ after data preprocessing and augmentation. Specifically, we first explain the details of the $\boldsymbol{CNN}$ model in section 2.1. Then, in sections 2.2 and 2.3, we introduce the LSTM method and combine it with 2D-CNN for AD classification.

2.1 2D-CNN for Feature Extraction

In recent years, many deep learning methods have been utilized for image processing and feature extraction. In this thesis, CNN is composed of several convolutional layers and pooling layers. The convolutional layer uses multiple convolution kernels to extract image features and generate feature map groups. Pooling layer to carry out the down-sampling, merging features to reduce dimensions; $\boldsymbol{LRU}$ layer is used to enhance the ability to generate which sections have a relatively high correlation. Especially, in the convolutional layer's connection, outputs of each layer could be shared and all the operation results would be decided by the $\boldsymbol{RELU}$ activation function. Then the pooling layer can reduce and regular the dimensionality of the sample data. The general formula for the convolutional layer can be written as:

$$\begin{cases} \text{Height} = \dfrac{\text{Matrix length} - \text{kernel height} + 2(\text{padding})}{\text{strides}} + 1, \\ \text{Width} = \dfrac{\text{Matrix length} - \text{kernel width} + 2(\text{padding})}{\text{strides}} + 1, \end{cases} \qquad (2-1)$$

Training the CNN model is difficult and complicated because we should observe the change of accuracy and loss curve based on the performance through training experiments. We choose the binary cross-entropy function to get the loss which can be used to update the parameters in each training loop. And due to the characteristic of the ***RELU*** activation function, the loss will increase rapidly, we should change the learning rate frequently and use the batch normalization can produce a higher learning rate on the model and reduce the parameter initialization. After through the SoftMax layer, it will export the probability of AD which would be used to update parameters for minimizing the loss. Moreover, in order to improve the robustness of the model when meeting abnormal data, Gaussian noise is used and it can produce effective results during the training step. Figure.3 shows the basic structure of CNN. In conclusion, this model can reduce the computational cost and improve accuracy.

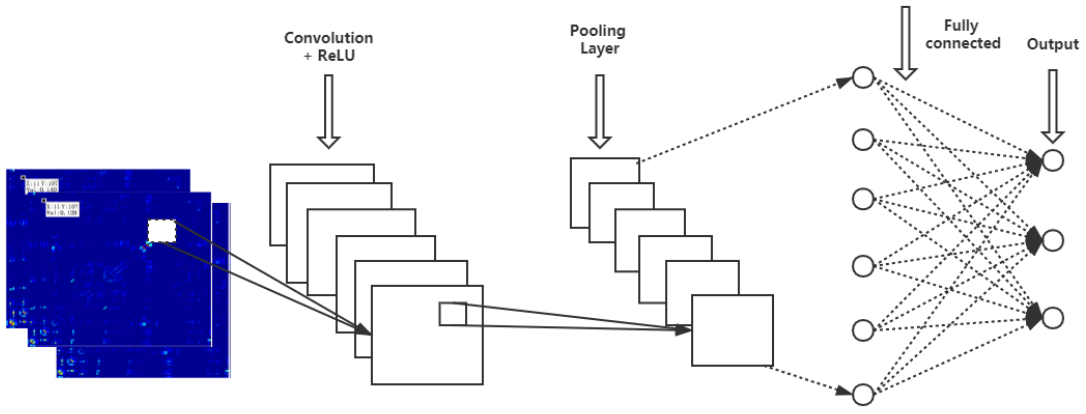

Figure 2 The Basic Architecture of Convolution Neural Networks

2.2 Long Short-Term Memory for Classification

Recurrent neural network (RNN) which contains four layers has powerful performance in sequence analysis. And LSTM provides the memory blocks instead of conventional RNN units, which can be used in solving the exploding gradient problem. By adding one cell state, the LSTM network can save long-term states, remembering and connecting the previous and current information extracted from output layers of CNN[21]. The LSTM network includes an input layer, a hidden layer that contains the memory cell, and an output layer. In each memory cell, a forget unit, an input unit, and an output unit have been comprised to effectively capture valuable and spatial information in a sequence, deciding which information would be removed from or added to the cell states. After the previous process, the output layer determines which information would be selected as new information.

The principle of LSTM that contains five important parts is shown in the following six equations from (2) to (7). Where *w* means weight matrices and x_t represents the input at the time step t. i_t f_t means the forget unit which allows loading abstract information into a long-term memory unit at the time t. By extracting abstract information from the output of neurons in the previous time series and putting it into neurons with inputs, neurons can be given memory which affects the following outputs. If the input door keeps open, a large amount of information pours into the memory, causing the value of the memory C to become very large. And when the input value is large, the gradient basically disappears. We should add a mechanism to get rid of the information and selecting the information using (3), in which θ is the sigmoid function. Therefore, combining the short-term and long-term dependencies, the new information can be stored into next time state using (5) and (6).

$$i_t = \theta(w_i h_{t-1} + w_i x_t + b_i) \tag{2-2}$$

$$f_t = \theta\left(w_f h_{t-1} + w_f x_t + b_f\right) \tag{2-3}$$

$$C = tanh(\, w_c h_{t-1} + w_c x_t + b_c) \tag{2-4}$$

$$o_t = \theta(w_o h_{t-1} + x_t + b_o) \tag{2-5}$$

$$c_t = f * c_{t-1} + i * C \tag{2-6}$$

$$h_t = o_t * tanh(\, c_t) \tag{2-7}$$

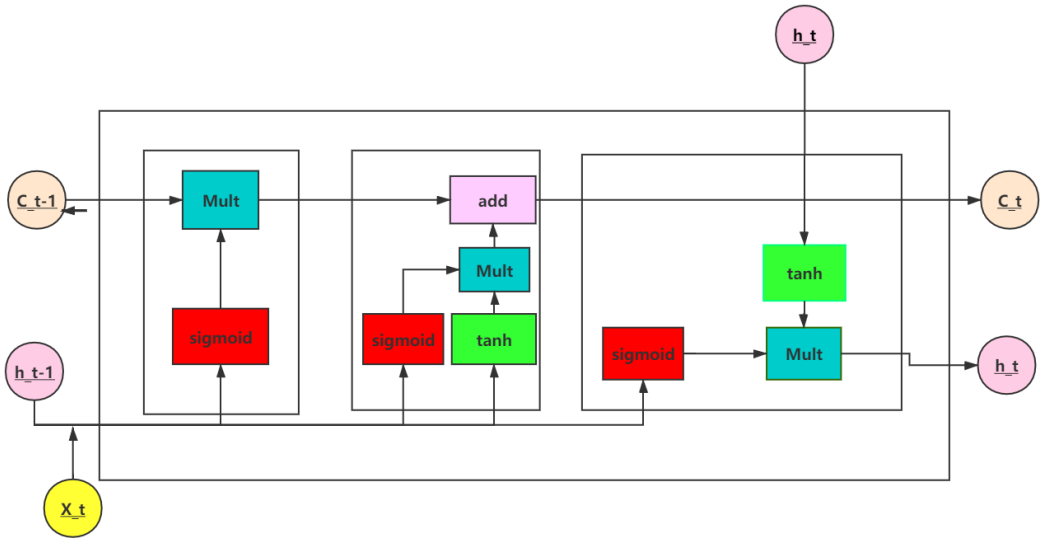

Figure 3 The basic structure of a single LSTM memory cell

Figure 3 shows the internal structure of the LSTM memory cell which could be formulated to process time series of AD progression. The output unit can determine the states that are required for continuation by h_{t-1} and x_t. Finally, h_t can represent that the output will be multiplied by the vector c_t and $tanh$ layer using (7).

2.3 A Novel Network Based on Combined CNN and LSTM

In this thesis, a novel hybrid network was developed by combining CNN and LSTM for AD classification, and the whole algorithm can be divided into two stages i) feature extraction stage based on CNN ii) feature fusion stage using LSTM as shown in Figure 4. The combined novel network can extract features using the relationships between ROIs and different stages of Alzheimer's disease for deeper learning. Most researchers used PCA (Principal Component Analysis) technique to extract main features from the original images, however, the main feature will be abstracted as the relationships between different ROIs.

Algorithm: A novel network based on multi-level CNN-LSTM
Required data: ADNI (86*86*9*445), NACC (113*113*9*202) 1. Building the training dataset using data normalization and transformation. 2. Construct CNN and LSTM hybrid neural network architecture. 3. Extract the feature of the training dataset through the convolution layer of CNN and perform the max-pooling operation. 4. Input the extracted features from CNN into LSTM. 5. Using the trained parameters and test data to identify the unknown brain condition.

Figure 4 The main workflow of the proposed method

3.0 Experiment

3.1 Data Description and Augmentation

In recent studies, Structural Magnetic Resonance Imaging (MRI) has been mostly utilized as a research tool to extract and analyze effective information for AD classification[22]. It captures the basic progressive deterioration in the brain, detecting the degenerative changes caused due to AD progression. In recent studies, many researchers apply whole-brain tractography algorithms on detecting large-scale relationships based on voxel in the brain network. And tractography methods mainly consist of two steps i) fitting the diffusion model ii) fiber tracking[2]. Figure 2 shows the typical example of a connectome brain network based on nine different tractography methods. In this thesis, to diagnose the brain network changes in different stages of AD, we assessed the importance of using original image features to find the connectivity of different ROIs in order to address the negative impact of numerous irrelevance parts of raw MRI data. Then the imbalanced data sample distribution would be overcome by using batch normalization and data augmentation techniques in different conditions with different model parameters[28].

The data used in this thesis composed of nine different folders, which come from nine different tractography algorithms based on two independent AD cohorts (202 subjects from the 2nd stage of Alzheimer's disease neuroimaging initiative or ADNI2 and 445 subjects from National Alzheimer's Coordinating Center or NACC)[23][24]. Each matrix's size based on ADNI2 in each folder is 86*86, and the matrix's size based on NACC in each folder is 113*113. Therefore, the dimension size of the final data is 86*86*9*445 and 113*113*9*202. Table 1 shows a summary of the subject's information from two independent AD cohorts.

Before entering the data into the network, we should transform all of the matrix and labels to the tensor, which acts as the input. All the data will be divided into three parts, including training, validation, and testing. In order to evaluate the stability and adaptability of the proposed method, we divided the whole dataset in each epoch during the training process and tested when iterations equal to 500.

Table 1 The number of patients in different stages in the ANDI2 dataset

Diagnosis	Number
AD	38
EMCI	73
LMCI	38
NC	53

Table 2 The number of patients in different stages in the NACC dataset

Diagnosis	Number
AD	54
MCI	57
NC	328

3.2 Experiment Setup and Parameter Settings

In this experiment, the dataset was divided into three parts, the training set, validation set, and test set accounted for 65%, 20%, and 15%, respectively. The proposed method based on CNN

and LSTM was implemented using the Pytorch framework, and the tensor calculation was executed using GPU NVIDIA GTX 1660Ti.

During the process of training, batch normalization and transformation were implemented before putting the data into convolution layers [25], and dropout would be used to avoid overfitting, which is a critical challenge for effectively capturing highly discriminative features related to AD. Then, we choose many different batch sizes such as 8,16,32 to accelerate the training process while improving accuracy. When solving Alzheimer's disease classification problems, the most common way is to set n output nodes in the last layer. For each batch, the model can get an n*1 array which has the same size as the number of stages in Alzheimer's disease progression, and we should choose the one with the highest value as the predicted class. And we should use cross-entropy to determine how close the actual output is to the expected output. Moreover, the learning rate should be considered carefully, because this hyperparameter can determine the speed of effective training and avoid overfitting during the training process. Finally, all of the experiments were trained in nearly 120 epochs while the curve of accuracy and loss would tend to be stable in 50 epochs.

3.3 Summary of Results

For comparing the proposed method with other methods, we trained four different models on two AD datasets (ADNI2 and NACC). The results are shown in Table 2. Figures 5 and 6 show the accuracy and loss result test on NACC and ADNI2 datasets in the test process. Our new model can achieve around 96% accuracy for multi-group classifications, and the smooth curve represents

stability. Besides, the accuracy and training loss can converge in nearly 50 epochs. In comparison to other methods, it can achieve better performance in terms of accuracy and stability.

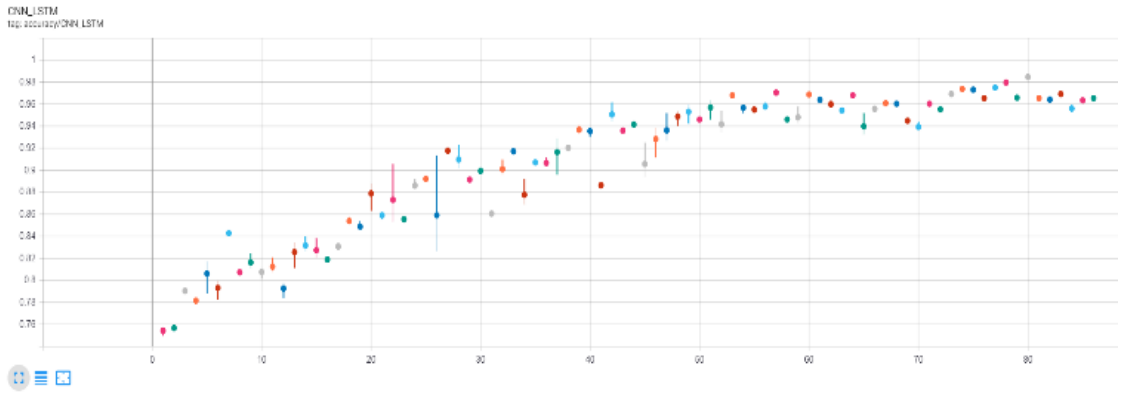

Figure 5 The result of Accuracy test on NACC dataset based on combined CNN-LSTM

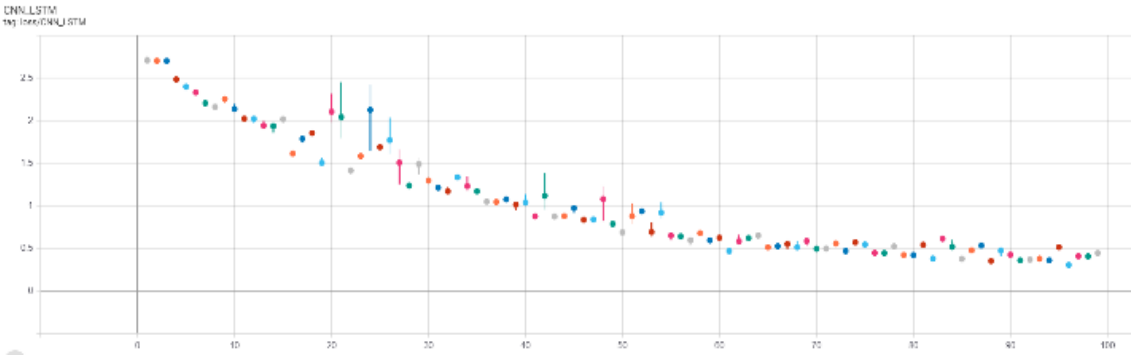

Figure 6 The result of Loss test on NACC dataset based on combined CNN-LSTM

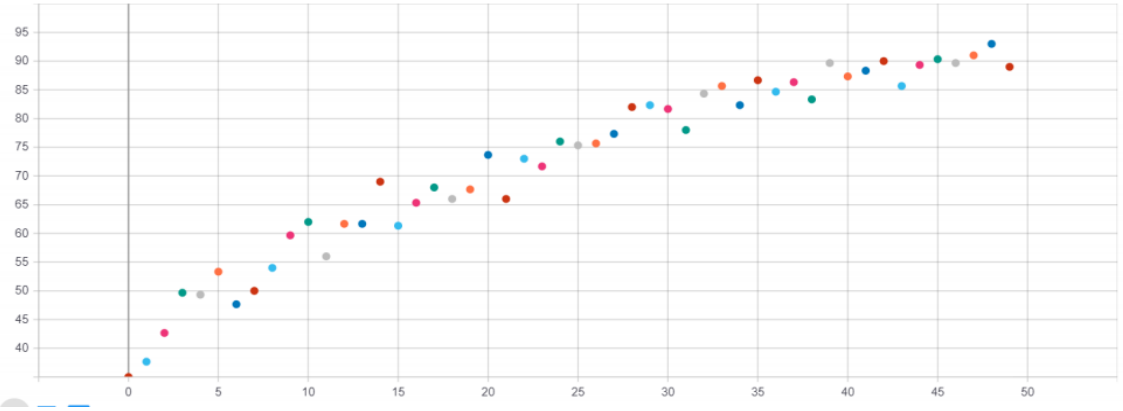

Figure 7 The result of Accuracy test on ADNI dataset based on combined CNN-LSTM

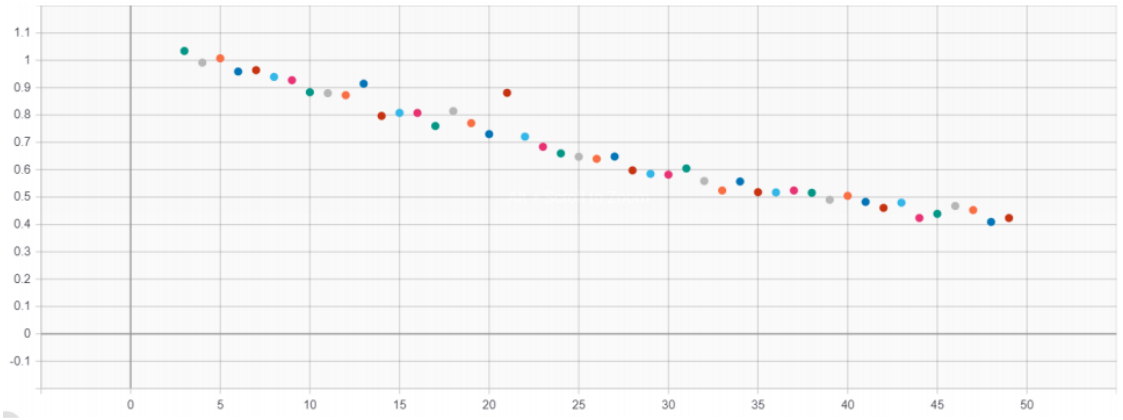

Figure 8 The result of the Loss test on the ADNI dataset based on combined CNN-LSTM

Moreover, we save all model parameters as a file after training for more than 50 epochs. In order to observe which ROIs in brain networks have the most effect on the progression of Alzheimer's disease, we have two further experiments 1)occlusion sensitivity analysis between different ROIs 2) the ROIs structure connectivity analysis. First, we should set the rows and columns of the matrix to zero in sequence, using the saved model parameters to test the accuracy of the classification. And we can rank the accuracy from low to high, to some extent, lower accuracy represents higher importance. For instance, when testing the ADNI2 dataset, some of the vertical values are very different from the others, which means that this region in the brain network can generate more effect on the progression of Alzheimer's disease.

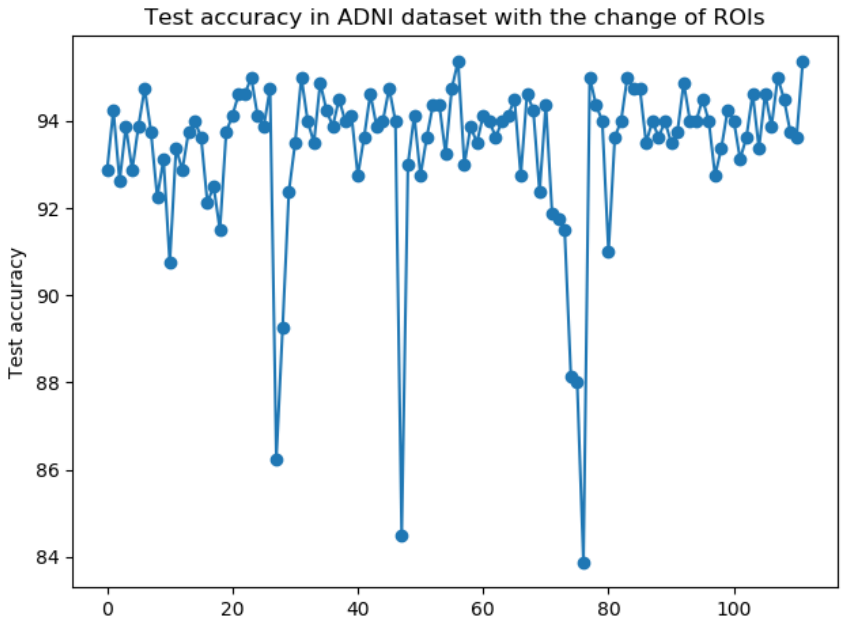

Figure 9 Test Accuracy in ADNI dataset with the change of 113 ROIs

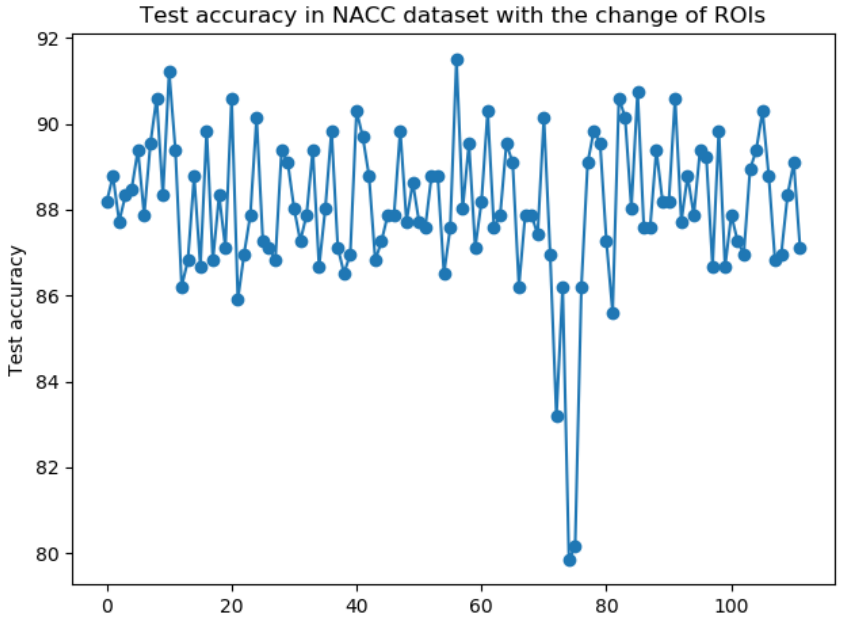

Figure 10 Test Accuracy in NACC dataset with the change of 113 ROIs

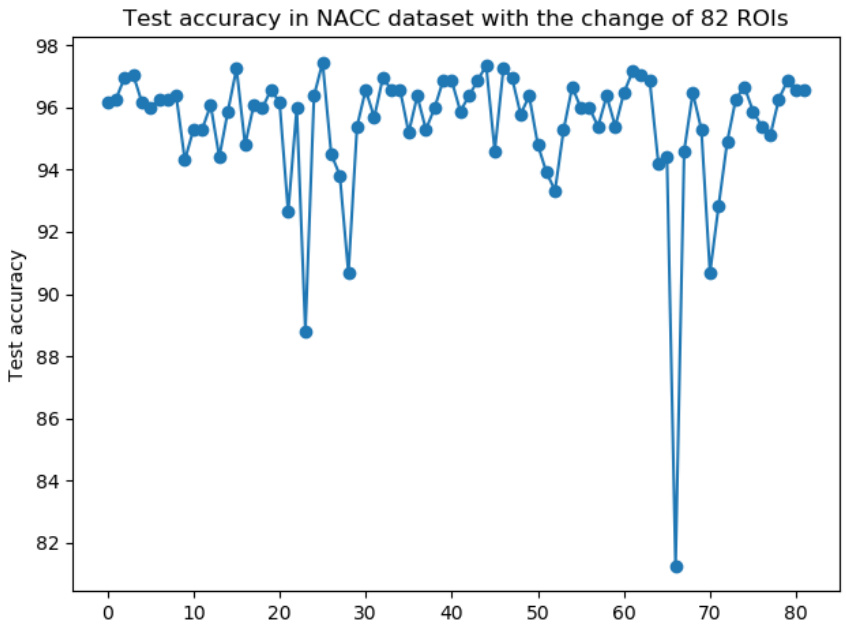

Figure 11 Test Accuracy in NACC dataset with the change of 86 ROIs

In the second experiment, we use the transformed and normalized data of different stages of patients to observe the connectivity structural networks, which can represent the performance evaluation of the proposed method which is based on the 2D-CNN and LSTM algorithms. For instance, Fig 13 shows that the connectivity between the left cingulate gyrus and the left precuneus can make more impress on the symptoms of the patient.

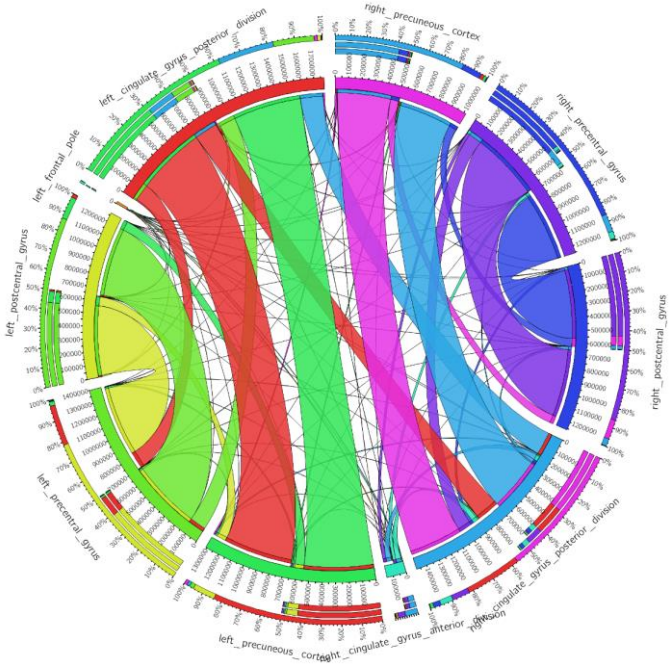

Figure 12 The structural connectivity networks of Top 10 related ROIs in the NACC dataset

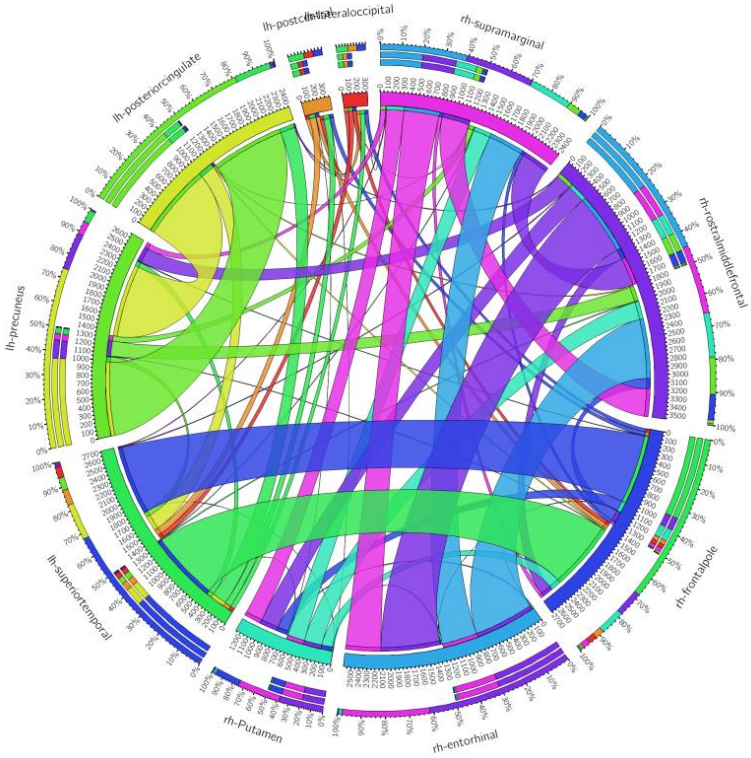

Figure 13 The structural connectivity networks of Top 10 related ROIs in the ADNI dataset

4.0 Conclusion

In this thesis, we proposed a novel feature hybrid network based on CNN and LSTM which focuses on feature extraction and learning feature relationships to observe AD progression. The dataset in the experiments can assign different significance of each ROI to different stages of AD, ignoring lots of useless information and saving calculation time. The 2D-CNN can combine high-level and low-level features to improve the ability to effectively capture highly discriminative and complex features related to AD, and the extracted feature will be transferred to LSTM modules to learn the relationship between time and brain structural changes by remembering spatial information which is important for AD progression research. The proposed method obtained an accuracy of around 96% which can demonstrate that the proposed architecture has a better performance than other exited methods when dealing with the same dataset due to limited image data. However, there exists some limitations to the proposed method. Firstly, this network should be applied to the processed dataset, affecting the feasibility to a certain extent. Secondly, we cannot observe which features significantly affect AD progression intuitively during the training and testing process.

Bibliography

[1] Huang, Yechong, et al. “Diagnosis of Alzheimer’s Disease via Multi-Modality 3D Convolutional Neural Network.” Frontiers in Neuroscience, vol. 13, no. MAY, 2019, doi:10.3389/fnins.2019.00509.

[2] Drobyshevsky, A., et al. “Diffusion Tensor Imaging of the Developing Rabbit Brain.” Magnetic Resonance in Medicine, vol. 11, no. July, 2004, pp. 2004–2004, https://www.sciencedirect.com/science/article/pii/S1933721307000955.

[3] Zhan, Liang, et al. “Comparison of Nine Tractography Algorithms for Detecting Abnormal Structural Brain Networks in Alzheimer’s Disease.” Frontiers in Aging Neuroscience, vol. 7, no. APR, 2015, pp. 1–19, doi:10.3389/fnagi.2015.00048.

[4] Prasad, Gautam, et al. “Tractography Density and Network Measures in Alzheimer’S Disease.” Proceedings - International Symposium on Biomedical Imaging, 2013, pp. 692–95, doi:10.1109/ISBI.2013.6556569.

[5] Chamberland, Maxime, et al. Tractometry-Based Anomaly Detection for Single-Subject White Matter Analysis. 2020, pp. 1–5, http://arxiv.org/abs/2005.11082.

[6] Reginold, William, et al. “Altered Superficial White Matter on Tractography MRI in Alzheimer’s Disease.” Dementia and Geriatric Cognitive Disorders Extra, vol. 6, no. 2, 2016, pp. 233–41, doi:10.1159/000446770.

[7] Marzban, Eman N., et al. “Alzheimer’s Disease Diagnosis from Diffusion Tensor Images Using Convolutional Neural Networks.” PLoS ONE, vol. 15, no. 3, 2020, pp. 1–16, doi:10.1371/journal.pone.0230409.

[8] Jo, Taeho, et al. “Deep Learning in Alzheimer’s Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data.” Frontiers in Aging Neuroscience, vol. 11, no. August, 2019, doi:10.3389/fnagi.2019.00220.

[9] Ebrahimighahnavieh, Mr Amir, et al. “Deep Learning to Detect Alzheimer’s Disease from Neuroimaging: A Systematic Literature Review.” Computer Methods and Programs in Biomedicine, vol. 187, Elsevier B.V., 2020, p. 105242, doi:10.1016/j.cmpb.2019.105242.

[10] Essemlali, Achraf, et al. “Understanding Alzheimer Disease’s Structural Connectivity through Explainable AI.” Proceedings of Machine Learning Research, vol. 121, 2020, pp. 217–29.

[11] Suk, Heung Il, et al. “Latent Feature Representation with Stacked Auto-Encoder for AD/MCI Diagnosis.” Brain Structure and Function, vol. 220, no. 2, 2015, pp. 841–59, doi:10.1007/s00429-013-0687-3.

[12] Shi, Jun, et al. “Multimodal Neuroimaging Feature Learning with Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer’s Disease.” IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 1, IEEE, 2018, pp. 173–83, doi:10.1109/JBHI.2017.2655720.

[13] Liu, Siqi, et al. “Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer’s Disease.” IEEE Transactions on Biomedical Engineering, vol. 62, no. 4, IEEE, 2015, pp. 1132–40, doi:10.1109/TBME.2014.2372011.

[14] Hosseini-asl, Ehsan, et al. “ALZHEIMER ’ S DISEASE DIAGNOSTICS BY ADAPTATION OF 3D CONVOLUTIONAL NETWORK” Electrical and Computer Engineering Department , University of Louisville , Louisville , KY , USA .” European Journal of Nuclear Medicine and Molecular Imaging, vol. 30, no. 8, IEEE, 2016, pp. 1104–13.

[15] Liu, Manhua, et al. “Classification of Alzheimer’s Disease by Combination of Convolutional and Recurrent Neural Networks Using FDG-PET Images.” Frontiers in Neuroinformatics, vol. 12, no. June, 2018, pp. 1–12, doi:10.3389/fninf.2018.00035.

[16] Cui, Ruoxuan, and Manhua Liu. “RNN-Based Longitudinal Analysis for Diagnosis of Alzheimer’s Disease.” Computerized Medical Imaging and Graphics, vol. 73, Elsevier Ltd, 2019, pp. 1–10, doi:10.1016/j.compmedimag.2019.01.005.

[17] Yang, Can, et al. “Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory.” Mathematical Problems in Engineering, vol. 2020, 2020, doi:10.1155/2020/2746845.

[18] Hong, Xin, et al. “Predicting Alzheimer’s Disease Using LSTM.” IEEE Access, vol. 7, IEEE, 2019, pp. 80893–901, doi:10.1109/ACCESS.2019.2919385.

[19] Greff, Klaus, et al. “LSTM: A Search Space Odyssey.” IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, 2017, pp. 2222–32, doi:10.1109/TNNLS.2016.2582924.

[20] Yang, Can, et al. “Personalized Channel Recommendation Deep Learning from a Switch Sequence.” IEEE Access, vol. 6, IEEE, 2018, pp. 50824–38, doi:10.1109/ACCESS.2018.2869470.

[21] Yang, Yang, et al. “A Study on Automatic Sleep Stage Classification Based on CNN-LSTM.” ACM International Conference Proceeding Series, vol. 1, 2018, pp. 1–5, doi:10.1145/3265689.3265693.

[22] Livieris, Ioannis E., et al. “A CNN–LSTM Model for Gold Price Time-Series Forecasting.” Neural Computing and Applications, vol. 32, no. 23, Springer London, 2020, pp. 17351–60, doi:10.1007/s00521-020-04867-x.

[23] Bailey, D. L., et al. “Combined PET/MRI: Global Warming—Summary Report of the 6th International Workshop on PET/MRI, March 27–29, 2017, Tübingen, Germany.” Molecular Imaging and Biology, vol. 20, no. 1, Molecular Imaging and Biology, 2018, pp. 4–20, doi:10.1007/s11307-017-1123-5.

[24] Petersen, R. C., et al. “Alzheimer’s Disease Neuroimaging Initiative (ADNI): Clinical Characterization.” Neurology, vol. 74, no. 3, 2010, pp. 201–09, doi:10.1212/WNL.0b013e3181cb3e25.

[25] McCutcheon, Sarah T., et al. “Clinicopathological Correlates of Depression in Early Alzheimer’s Disease in the NACC.” International Journal of Geriatric Psychiatry, vol. 31, no. 12, 2016, pp. 1301–11, doi:10.1002/gps.4435.

[26] Becherer, Nicholas, et al. “Improving Optimization of Convolutional Neural Networks through Parameter Fine-Tuning.” Neural Computing and Applications, vol. 31, no. 8, Springer London, 2019, pp. 3469–79, doi:10.1007/s00521-017-3285-0.

[27] Zhang, Tao, and Mingyang Shi. “Multi-Modal Neuroimaging Feature Fusion for Diagnosis of Alzheimer’s Disease.” Journal of Neuroscience Methods, vol. 341, no. May, Elsevier, 2020, p. 108795, doi:10.1016/j.jneumeth.2020.108795.

[28] Chen, Jianhui, et al. “Learning Incoherent Sparse and Low-Rank Patterns from Multiple Tasks.” *ACM Transactions on Knowledge Discovery from Data*, vol. 5, no. 4, 2012, doi:10.1145/2086737.2086742.