[UCI]I804:24011-200000278978

2020년 02월
석사학위논문
Deep Learning Methods for Exploring Alzheimer Diseases in Structured Magnetic Resonance Imaging
SAMSUDDIN AHMED

2020년 02월

석사 학위 논문

Deep Learning Methods for Exploring Alzheimer Diseases in Structured Magnetic Resonance Imaging

조선대학교 대학원

컴퓨터공학과

SAMSUDDIN AHMED

February 2020
Master's Degree Thesis

Deep Learning Methods for Exploring Alzheimer Diseases in Structured Magnetic Resonance Imaging

Graduate School of Chosun University

Department of Computer Engineering

Samsuddin Ahmed

Deep Learning Methods for Exploring Alzheimer Diseases in Structured Magnetic Resonance Imaging

Structural Magnetic Resonance Image 영상에서
딥 러닝을 사용한 알츠하이머병 진단 연구

February 25th, 2020

Graduate School of Chosun University

Department of Computer Engineering

Samsuddin Ahmed

Deep Learning Methods for Exploring Alzheimer Diseases in Structured Magnetic Resonance Imaging

Advisor: Prof. Jung, Ho Yub

A thesis submitted in partial fulfillment of the requirements for a Master's degree

October, 2019

Graduate School of Chosun University

Department of Computer Engineering

Samsuddin Ahmed

삼수딘 아흐메드 석사학위논문을 인준함

위원장 조선대학교 교수 **최우열** 

위 원 조선대학교 교수 **정호엽** 

위 원 조선대학교 교수 **신석주** 

2019년 11월

조선대학교 대학원

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

aAD	Asymptomatic ADD
AD	Alzheimer Disease
ADD	Alzheimer Disease Dementia
ADNI	Alzheimer Diseases Neuroimaging Initiative
AI	Artificial Intelligence
AIBL	Australian Imaging Biomarkers and Lifestyle Study of Ageing
DBN	Deep Belief Network
CNN	Convolutional Neural Network
DL	Deep Learning
DML	Deep Metric Learning
FDA	Food and Drug Administration
GARD	Gwangju Alzheimers and Related Diseases Dementia
GUI	Graphical User Interface
LA	Left Amygdala
LASSO	Least Absolute Shrinkage and Selection Operator
LH	Left Hippocampus
LI	Left Insula
MRI	Magnetic Resonance Imaging
ML	Machine Learning
MCI	Mild Cognitive Impairment
mAD	mild ADD
MMSE	Mini-Mental State Exam
MPRAGE	Magnetization Prepared - RApid Gradient Echo
NRCD	National Research Center for Dementia

NIA	National Institute on Aging
NC	Normal Control
NIBIB	National Institute of Biomedical Imaging and Bioengineering
OASIS	Open Access Series of Imaging Studies
PBC	Patch-based Classifier
PCA	Principal Component Analysis
PET	Positron Emission Tomography
RA	Right Amygdala
ROI	Region of Interest
RH	Right Hippocampus
RI	Right Insula
SD	Standard Deviation
sMRI	Structured Magnetic Resonance Imaging
SAE	Sparse Auto Encoder
SVM	Support Vector Machine
TVP	Three View Patch

LIST OF FIGURES

LIST OF TABLES

List of Algorithms

LIST OF SYMBOLS

a_i	Actual offset of i_{th} sample
batchsize	mini-batch size in training and testing
e	exponent
h_x	x coordinate of hippocampus location
h_y	y coordinate of hippocampus location
h_z	z coordinate of hippocampus location
I	Intensity values of sMRI
$\widetilde{I}$	Normalized Intensity values of sMRI
n_x	Cardinality of T_x
mse	Mean square error
n_y	Cardinality of T_y
n_z	Cardinality of T_z
pl_x	x-coordinate of predicted location
pl_y	y-coordinate of predicted location
pl_z	z-coordinate of predicted location
p_i	Predicted offset of i_{th} sample
r_x	x-coordinate of reference point
r_y	y-coordinate of reference point
r_z	z-coordinate of reference point
R	set of reference points
T_x	Set of values consisting of x coordinates of reference points
T_y	Set of values consisting of y coordinates of reference points
T_z	Set of values consisting of z coordinates of reference points
α	Shape constant in X axis of cubic reference frame

β	Shape constant in Y axis of cubic reference frame
γ	Shape constant in Z axis of cubic reference frame
σ	Standard Deviation
μ	Mean
$\forall$	Universal quantifier
$\in$	Membership

한 글 요 약

Structural Magnetic Resonance Image 영상에서 딥 러닝을 사용한 알츠하이머병 진단 연구

삼수 딘 아흐메드
지도 교수: 정 호엽
컴퓨터공학과
대학원, 조선대학교

기계 학습 기술에 기반한 알츠하이머병 (AD) 및 그 prodromal 단계 (aAD, mAD, NC)의 자동 진단은 지속적인 연구가 진행되고 있습니다. 최근 딥 러닝 (DL) 기반 접근 방식은 분류 및 regression에서 최첨단 성능을 보여주며, DL 기반 방법은 AD 연구에 널리 사용되고 있습니다. 다중 모드 진단을 고려한 최첨단 기술은 임상 진단보다 정확도가 더 우수한 것으로 나타났지만, 다중 modality로 데이터를 수집하는 것은 시간이 많이 걸리고 고비용이며, 일부 방식은 방사성 부작용이 있을 수 있습니다. 본 연구는 조선대 국체치매연구단에서 Gwang-ju Alzheimer's research data(GARD) 코호트 데이터 세트와 Alzheimer's Disease Neuroimaging Initiative (ADNI)에서 제공하는 structural magnetic resonance image (MRI) 데이터세트에 국한하였습니다. 연구의 목표는 다음과 같습니다. AD 진단 가능한 특징을 제공하는 두뇌 랜드마크를 식별하고 분석합니다. AD와 그 prodromal AD 진단의 정확도 수준을 높이고, 성능 저하 없이 제한된 컴퓨팅 리소스로 배포할 수 있는CNN을 설계합니다. ROI 기반 알츠하이머병 진단을 위해 우선 GARD에서 제공하는 아틀라스 기법으로 세분화된 영상을 사용하여 AD에 대해 통계적으로 유의한 ROI를 선택했습니다. AD 대 NC 분류 작업은 three-view-patch 기반 컨볼루션 신경망의

앙상블을 학습했습니다. 더 나아가ROI 위치를 예측을 위해 딥 메트릭 학습을 하고, hippocampus, amygdala 및 insula를 기반으로 한 prodromal 단계를 진단하기 위해 ROI 기반 CNN의 앙상블을 학습했습니다. 이 연구를 통해 뇌 MRI의 hippocampus, amygdala 및 insula가 AD에 진단에 결정적인 정보를 제공한다는 것을 관찰하였고, ROI 패치 기반 앙상블 classifier를 사용하여 알츠하이머병 진단 정확성을 state-of-the-arts까지 달성할 수 있었습니다.

ABSTRACT

Deep Learning Methods for Exploring Alzheimer Diseases in Structured Magnetic Resonance Imaging

Samsuddin Ahmed
Advisor: Prof. Jung, Ho Yub, Ph.D.
Department of Computer Engineering
Graduate School of Chosun University

There is ongoing studies for the automatic diagnosis of Alzheimer's disease (AD) and its early stages (aAD, mAD, NC) based on traditional machine learning techniques. Recently deep learning (DL) based approaches are demonstrating state-of-the-art performance in classification and regression. As a result, DL-based methods are becoming popular choice for AD research. The state-of-the-art techniques that consider multimodal diagnosis have been shown to have accuracy better than manual diagnosis. However, collecting data from different modalities is time consuming and expensive, and some modalities may have radioactive side effects. Our study is confined to structural magnetic resonance imaging. Here, we have exploited Gwangju Alzheimer's and Related Dementia (GARD) cohort dataset prepared by National Research Center for Dementia (GARD), Gwangju, South Korea. The objectives of our attempt are as follows: 1) to identify and analyze the brain-landmarks that provide discernible features for AD; 2) to increase the accuracy level of AD and its prodromal stages diagnosis that is comparable to the state-of-the-art methods; and 3) to design simpler CNN that is deployable with limited computing resources without sacrificing

the performance. Achieving the objectives required us to perform following experiments: 1) selecion of statistically significant ROI for AD using the atlas-based segmentation dataset provided by GARD 2) deployment of ensembles of patch based convolutional neural networks on hippocampus features for binary classification tasks 3) utilizing deep metric for ROI localization confining the study only on hippocampus 4) designing ensembles of simpler CNN classifiers for AD and its prodromal stages diagnosis. We have observed that 1) hippocampus provides significant information for AD 2) The ROI does not provides distinctive features for AD in sMRIs modality 3) state-of-the-art diagnosis performance is achievable by deploying patch-based ensemble classifiers based on the significant ROI features.

I. INTRODUCTION

A. Introduction

Recently, Deep learning (DL) has become a powerful and successful approach to lead the era of artificial intelligence (AI). It has achieved state-of-the art performance in classification and regression. The performance of DL-based techniques in image classification [1]–[5] , natural language processing [6], speech recognition [7]–[9], health care [10], [11] are over-human level. As, DL has been showing outstanding performance in all classification and regression [12] tasks, in this study, we have exploited DL approaches for exploring Alzheimer Diseases (ADs) from structural magnetic resonance imaging (sMRI) at the hope that it would potentially assist the radiologist to improve diagnostic accuracy.

Alzheimer disease (AD) is the most predominant neurodegenerative brain disease affecting elderly people worldwide. No cure or effective treatment is currently known for this disease. With the increase of life expectancy, AD is becoming more prevalent among the peoples older than 65 years[13]. Studies are ongoing for the early diagnosis of this disease in order to put a brake on the abnormal degeneration of the brain, to reduce the cost of patient care and to ensure better management. Diagnosis of AD has traditionally relied mainly on cognitive evaluation and clinical observation. Studies [14]–[17], however, indicated that image analysis of neuroimaging scans may be reliable approach for supporting clinical decisions. Attentions had thus been provided to computer aided diagnosis (CAD). Previous studies [14]–[17] have shown that machine learning algorithms were able to classify AD more accurately than experienced clinicians. Considering the outstanding performance DL-based approaches are

becoming the obvious choice for the detection of ADs.

The state-of-the-art approaches either consider the whole brain in a single modality [18] [19] or multimodal [15] datasets to train machine learning models, which have been shown to demonstrate greater accuracy than manual diagnosis.

Investigating more than one data modality is time consuming and expensive. Moreover, modalities such as PET may have radioactive side effects on patients. Here, we consider unimodal imaging for experiment in an attempt to achieve state-of-the-art accuracy and efficiency. Here, we consider sMRI as the modality of our experiments for the following advantages:

1. High degree of imaging flexibility;
2. MRI gives an excellent spatial resolution along with good contrast;
3. No need for ionizing radiation;
4. Useful information about the anatomy of the brain;
5. Lack of pain to patients.

B. Motivation

Our prime motivation of this research is designing a simpler CNN for diagnosing AD and its prodromal stages with significant accuracy and computational efficiency. Our motivations include:

1. Deploy DL methods to identify and analyze the brain-landmarks that provide discernible features for AD;
2. Diagnosis of AD and its prodromal stages with state-of-the-art accuracy;

3. Learning simple CNN model with limited computing resources without sacrificing the performance.

C. Contribution

We have performed following experiments for achieving our objectives.

1. Statistically significant ROIs for AD were selected. We have used atlas-based segmentation dataset of GARD provided by NRCD for statistical tests.

2. Robust ROI localization were performed with HCNN and deep metric based verification. We have confined our experiment only on hippocampus. GARD database was exploited for this purpose and achieved outcomes comparable to state-of-the-art performance

3. Ensembles of patch based convolutional neural networks (PBCNN) were utilized on hippocampus features for performing binary classification between different stages of AD. Here, we have exploited both ADNI and GARD database

D. Methodology

Our complete pipeline is presented in figure 1. At first, the significant ROIs in the sMRI modality were selected by analyzing atlas-based segmented volume measures in GARD data set. We have performed permutation test for finding AD-affected ROIs from 108 different regions. Hippocampus was observed to be the most significant region for AD diagnosis. Then, we have learned hippocampus

features by deploying Siamese network. Contrastive loss function [20] was utilized for learning hippocampus embeddings. At third step, we have localized the hippocampus from sMRI using hough convolutional neural network(HCNN) along with deep metric verification. After that PBCNN was deployed for classifying individual TVP in different classes for different stages of AD. We have tested our models for classifying sMRIs. We have deployed ensemble classifiers for improving the performance.

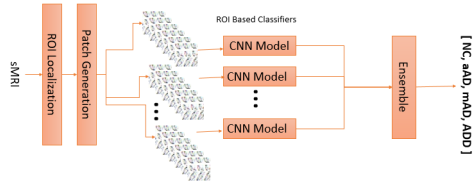


Figure 1: ***The framework for early diagnosis of Alaheimer diseases(AD) in sMRI modality with deep learning(DL) methods***

.

E. Research Outcomes

We have observed that

1. Hippocampus provides most significant information for AD diagnosis. Hippocampus features were observed to provide state-of-the-art results in diagnosing AD and its pro-dromal stages. It provides near to 90% accuracy in classifying AD sMRI from NC classes.

2. The ROI does not provide distinctive features for aAD in sMRIs modality

3. State-of-the-art diagnosis performance is achievable by deploying patch-based ensemble classifiers based on the significant ROI features.

4. Simpler CNN model can be learned by compressing the ensembles without sacrificing the performance.

F. Thesis Layout

The chapter one introduces the entire study. We have briefly described our motivation, contribution, methodology, research outcomes and thesis outline in this chapter.

In chapter two, we have illustrated our data sets under study. We described two data sets namely ADNI and GARD. We have also explored the participants, acquisition protocol and pre-processing techniques of the dataset under study.

In chapter three, brief outline is provided on statistical test for selecting most important structures in sMRI modality. Obtained p-value for sixteen different ROIs is presented here.

In chapter four, we have discussed the learning process of a deep metric for localization of this important biomarker. Elaborate discussion of the structural details of Siamese network is given which is followed by training procedure and testing outcomes.

In chapter five, we have described the localization pipeline of landmark in brain sMRI. We have presented the hippocampus localization process with two step predictions. Far jump hough convolutional neural network(FJHCNN) predicts the rough estimates while short jump hough convolutional neural network(SJHCNN) predicts the fine-tuned predictions.

In chapter six the training and testing of patch-based classifiers are described. We have utilized the patch-based classifiers for designing ensemble model for early diagnosis of AD and its prodromal stages. The results are elaborately described.

In chapter seven, we have compared our method with existing state-of-the-art methods.

In chapter eight, we have concluded our study.

II. DATA SET

A. Introduction

Magnetic Resonance Imaging(MRI) is the de facto modality in brain studies due to its superior image contrast in soft tissue without involving ionizing radiation. MRI image are being widely used to examine other anatomical regions as well [6]. There are a lot of MRI data set for AD detection, such as: ADNI [21], OASIS[22], AIBL[21] etc.

In this study, two different dataset was taken into consideration. One from the ADNI database (adni.loni.usc.edu) and another from GARD database. We will briefly explain the dataset in the following subsections.

B. ADNI Data Collection and Pre-processing

The ADNI was inaugurated in 2003 by the NIA, NIBIB, FDA and private pharmaceutical companies together with nonprofit organizations. This was a $60 million and 5-year long public-private partnership. The rudimentary goal of ADNI was to test whether serial MRI, PET, clinical and neuropsychological assessment and other biological markers can be merged to assess the progression of early AD.

Finding useful markers of very early AD progression is aimed to aid researchers and clinicians to innovate new treatments and assess their effectiveness, as well as reduce time and cost of the clinical trials.

The principal investigator of ADNI is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts and dedication of many co-investigators from a wide range of academic

Table 1: ***Demographic features of ADNI database under study***

Class Label	Number of Scans	Age	Gender(Male/Female)	Education	MMSE
Normal Control(NC)	129	74.3 ± 3.6	57/72	16.41±2.62	29.08±1.08
Mild Cognitive Impairment(MCI)	145	70.23±2.74	83/62	16.25±2.45	28.36±1.61
Alzheimer Demented(AD)	77	71.13±2.54	37/40	23.70±2.13	23.70±2.13

institutions and private corporations, and participants have been recruited from over fifty sites across the United States and Canada. ADNI has been followed by ADNI-GO and ADNI-2 after its initial goal to recruit 800 subjects. To date these 3 protocols have recruited more than 1500 adults, aged 55-91, to participate in the research, consisting of NC aged individuals, persons with early or late MCI, and subjects with early AD. The follow-up duration of each group is specified in the protocols for ADNI-1, ADNI-2, and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. Clinical characteristics along with demographic information of the subjects are described in Table 1.

1. Subjects under Study

From ADNI dataset, we have selected 60 subjects aged between 55 and 92. The chosen participants met the standards defined in the ADNI protocol. There are 351 scans of these 60 subjects. We constructed balanced dataset consisting of 351 scans as follows:

1. 22 NC subjects: 12 males, 10 females; age±SD= 74.3 ±3.6 years, range = 62–91 years; MMSE score = 29.2±1.0, range = 25–30.

2. 18 MCI participants who did not progressed to AD within 18 months: 11 males, 7 females; age±SD = 70.4±3.12 years, range = 56–88 years; MMSE

score = 27.1±1.8, range = 24–30.

3. 20 AD subjects: 9 males, 11 females; age±SD = 74.0 ±5.2 years, range = 54–91 years; MMSE = 23.3±2.0, range = 18–27.

2. MRI Image Acquisition

All participants were scanned with a T1-MRI protocol which is optimized for high quality contrast to noise in a suitable acquisition time. Raw data had an acquisition matrix of $192 \times 192 \times 166$ and voxel size $1.25 \times 1.25 \times 1.2$ mm3. Zero-filled reconstruction resulted in a 256×256 matrix and voxel size of $0.9375 \times 0.9375 \times 1.2$ mm3. The sMRI sequence parameters were with resolution: $0.4 \times 0.4 \times 2.0 mm^3$, repetition time/echo time (TR/TE)8020/50 ms, minimum slices: 24, and minutes acquisition time :8.1 minutes More details on this imaging protocols is available at http://adni.loni.usc.edu/ methods/documents/mriprotocols/.

The raw data for sMRI scans were provided in NII format in the ADNI database. For our experiment, we have done some preprocessing on the data. In the following sub section we will illustrate the steps of preprocessing

3. ADNI Preprocessing

For reconstruction and volumetric segmentation and to extract useful pattern of the data, we used the preprocessing pipeline of the Free-Surfer 5.3.0 [23]software package. The free surfer software performs a sequence of preprocessing tasks with the recon-all processing pipeline on the original sMRI data. The preprocessing includes

1. Motion correction

2. T1-weighted image averaging

3. Registration of volume to the Talairach space

4. Skull striping with a deformable template model

The the pial surface and white surface are generated for hemispheres from the shape of the pons and corpus callosum in the Talairach space. The accurate matching of the morphologically homologous cortical locations across subjects were estimated using the mapping of the atlas on the basis of a cortical surface to a sphere with aligning the cortical patterns. Cortical thickness at every vertices of the cortex are denoted by the average minimum distance between white and pial surfaces. The area of every triangle in the standardized spherical surface tessellation gives the surface area. Equivalently, the registration surface based on the pleat pattern was used to compute the local curvature.

C. GARD Data Collection and Preprocessing

1. Subjects under Study

GARD dataset is based on T1-weighted sMRI baseline scans of 326 samples taken by NRCD at Chosun University Hospital. All the participants are Korean individuals. The ages of the subjects varies from 49 years to 87 years (mean: 72.02 $\pm$ standard deviation:0.60) while more than 88% subjects are over 65 years old. The education level of the subjects varies from illiterate to highly educated(scale 0 to 22).

The scans are labeled with asymptomatic AD (aAD), mild cognitive impairment AD (mAD), AD (ADD) and normal controlled (NC). There are 171 scans for NC class and 81 scans for ADD class. Number of available scans for

Table 2: ***Demographics characteristics of the studied subject from GARD database(the values are mean)***

Diagnosis	#subject	Age(F/M)	Gender(F/M)	Education(0-18;F/M)
ADD	81	71.85/71.46	42/39	5.33/9.51
NC	171	70.27/73.4	88/83	7.52/10.92
aAD	35	72.35/73.27	20/15	6.6/9.6
mAD	39	70.50/74.76	14/25	5.64/9.64

aAD and mAD are 39 and 35 respectively. We have re-termed the ADD-label as AD. We have performed the experiment on all the four classes picking two classes at a time. So, there are six combinations for binary classification.

The written consents was taken either from the participants or from the care givers, sometimes from both for performing the study. The approval from the regional ethics committee was also ensured for conducting the study. A neuro-physiological test was administered by a team of experts for assessing the language, attention, memory, visio-spatial and executive function of the participants. sMRI scans having focal lesions due to medical history such as head trauma, psychiatric causes of the participants were excluded from the dataset.

2. Data Acquisition

The imaging was performed in Chosun university hospital. Contiguous 0.88 mm sagittal MPRAGE images of the whole brain were acquired at Chosun University Hospital. The acquiring machine was a 3T Skyra, Siemens with TR=2300 ms, TE=2.143 ms, T1=900 ms, flip angle=9, field of view=256 $\times$ 256 matrix= 320 $\times$ 320,number of slices=178. T1 weighted MRIs were processed using an automated reconstruction protocol.

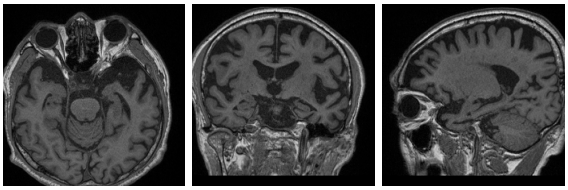
3. Preprocessing

The acquired high resolution structural T1-weighted images were processed using FreeSurfer software package version 5.3.0 [23]. The operating system on which Free surfer was installed was 64-bit CentOS 7. The freesurfer pipeline and methodologies includes a complete automated processing. Here, cortical and subcortical and ROI labeling was performed using the Desikan-Killiany atlas on each subject.

The free-surfer processing pipeline includes:

1. Motion correction;
2. Non-uniform intensity normalization for intensity inhomogeneity correction;
3. Image registration using affine transformation to Talairach space ;
4. Skull-stripping based on combination of watershed algorithm;
5. Removal of non-brain tissues.

After that, the image is intensity normalized. The nonlinear warping of each scan was performed to atlas image which is further utilized for atlas-based segmentation and ROI labeling. In the next step, per hemisphere topologically correct cortical surface representation was generated which lies at WM/GM or WM/CSF interface. The representations are then automatically mapped to a standard spherical coordinate system. This mapping into standard spherical coordinate system allows for automated anatomical parcellation of the cortex into gyral regions. Then, surface parcellation is extended to GM volume that



(a) *(b)* *(c)*

Figure 2: ***Three different views(axial, coronal and sagittal views from left to right) of GARD-sMRI at the voxel location (125,165,151)***

yields regional cortical volumes and parcellation of GM tissue sheet . The entire computation took about 12-18 hours for each scan [24]

D. Dataset Preparation for Patch-Based Model Learning

The axial, coronal and sagittal views of a sample sMRI are shown in Figure 14. We have normalized the intensity values by subtracting the mean intensity and then dividing by the standard deviation to have zero mean and unit variance of the input. The normalization is defined in (1).

$$\widetilde{I}(i_x, i_y, i_z) = \frac{I(i_x, i_y, i_z) - \mu(I)}{\sigma(I)} \tag{1}$$

*H*ere, $I(i_x, i_y, i_z)$ is the intensity of (i_x, i_y, i_z) location before normalization, $\mu(I)$ is the mean intensity and $\sigma(I)$ denotes the standard deviation of the intensity; $\widetilde{I}$ is the normalized intensity of the MRI.

1. TVP Extraction

After intensity normalization, we manually/automatically observed hippocampus locations (h_x, h_y, h_z) on the normalized MRI by using Mango, a multi-image

analysis graphical user interface (GUI)[25] or our developed automatic methods. Six manually observed example locations are presented in fig. 3. Each location and its neighboring points up to α, β and γ pixels in sagittal, coronal and axial direction, respectively, were used as a reference frame for patch generation. Careful selection of these shape constants; i.e., α, β, and γ; ensured that each reference frame lies within the hippocampus region. In our experiment, we have selected α,β and γ from 4 to 8 as the experiment requires. Different values of these constants provides flexible shape of the reference frame to adjust with the shape of ROI. We have randomly chosen n_x, n_y, and n_z number of co-ordinates in sagittal, coronal, and axial directions, respectively. As described in (4), (4), (4) and (5), these co-ordinates are used to generate reference points for TVPs.

$$T_x = rand(h_x - \alpha, h_x + \alpha, n_x) \tag{2}$$

$$T_y = rand(h_y - \beta, h_y + \beta, n_y) \tag{3}$$

$$T_z = rand(h_z - \gamma, h_z + \gamma, n_z) \tag{4}$$

Here, T_x, T_y, T_z represent uniformly distributed integer samples from the specified interval. The number of samples drawn from the interval are denoted by n_x, n_y, and n_z respectively. $rand(h_x - \alpha, h_x + \alpha, n_x)$ returns n_x number of uniformly distributed random integers from the interval $(h_x - \alpha, h_x + \alpha)$. The same explanation follows for next two lines in equation 4.

The reference points were generated by taking all (i, j, k) tuples of $T_x \times T_y \times T_z$ (i.e., the cartesian product). The disjunction of all reference points obtained from each reference frame were used to generate TVPs. The equation (5) summarized the operation. The algorithm 1 concisely describes the reference

point generation process.

$$R = R \cup \{\forall(i,j,k) | i \in T_x, j \in T_y, k \in T_z\} \quad (5)$$

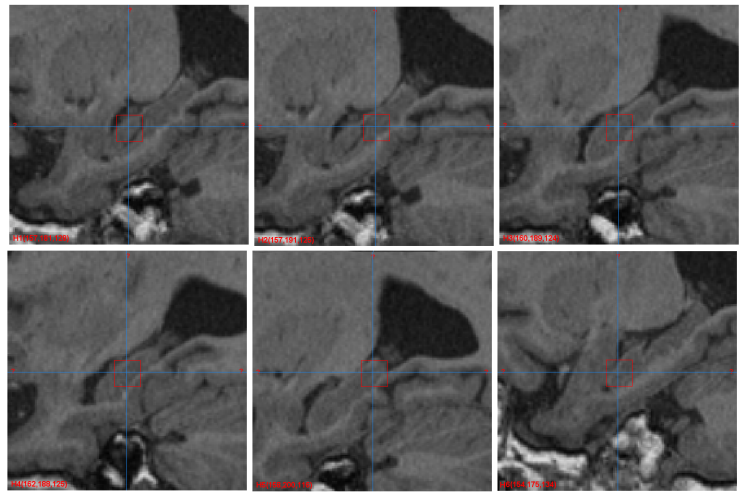


Figure 3: ***An example of manual localization of left (first row) and right(second row) hippocampus in three sMRI (viewed on the sagittal plane) that are used as the center of*** $8 \times 8 \times 8$ ***cubes for generating reference points. The reference points are selected from the cube by a semi-random process. TVPs are generated on the reference points.***

.

2. Data Augmentation

As training a convolution neural network required huge amount of data, we need to augment the data set. We generate additional training data using 1) random translation 2) generating three 32*32 slices in a reference point. This is especially important as there are imbalanced number of AD,aAD, mAD and NC cases.

3. Ground Truth Preparation

Metric Learning: For metric learning we have selected pair of TVPs, say (x_1, x_2). If both of them are associated in the same region then the label is 1(one). If they belongs to different regions then the label is 0(zero).

Localization: We take a reference point by purposive sampling within the range of image co-ordinates. Then we take TVPs centering the reference point. We calculated the offset of hippocampus location from that reference points. This offset were being used as ground truth for localizing the hippocampus in a given image.

Classification: As training and testing of the models were performed with TVPs, the data unit was consist of (x,y) pair, where x is a TVP and y is the one-hot encoding of the class label.

Algorithm 1: Algorithm for generating reference voxel positions from an assumed solid cubic structure to produce TVP.

Input: H=$\{H_1, H_2, H_3, ...H_n\}$: Manual or Model predicted landmark locations; The shape constants of the assumed solid centered at the location feeds α, β, γ

Output: R: a set of reference locations,(x,y,z)

$R = \{\}$

for *each point* $H_r(h_x, h_y, h_z) \in H$ **do**

  $T_x = rand(h_x - \alpha, h_x + \alpha, n_x)$

  // $rand(h_x - \alpha, h_x + \alpha, n_x)$ returns n_x number of uniformly distributed random integers from the interval $(h_x - \alpha, h_x + \alpha)$

  $T_y = rand(h_y - \beta, h_y + \beta, n_y)$

  $T_z = rand(h_z - \gamma, h_z + \gamma, n_z)$

  $R = R \cup \{\forall(i, j, k) | i \in T_x, j \in T_y, k \in T_z\}$

return R

III. LANDMARK SELECTION

Abstract *The clinicians have known that certain regions of interest (ROIs) are related to Alzheimer diseases (ADs). In this experiment we will find which regions of interests are most significant in diagnosing AD and its prodromal stages. To find the significant ROIs we have performed permutation test on the volumetric measures of atlas-based segmented data of GARD datasets. We have found that hippocampus, amygdala, insula, precuneus etc. are most significant ROIs for AD. We also performed some literature review to justify that our findings.*

A. Introduction

The clinicians have known that certain regions of interest (ROIs) are related to Alzheimer diseases (ADs). The studies [26], [27] on the magnitude and spatial pattern of AD acquired on histological or imaging data is useful for CAD based AD-diagnosis from sMRI. There are studies which infer that AD affects specific brain regions more than the whole brain in general. Some studies [28], [29] illustrated that atrophy of the hippocampus due to AD. The entorhinal cortex was also severely affected by this disease [30]–[32], .Ridha et. al. [33] mentioned about expansion of the ventricles due to AD. There are volumetric changes in amygdala [34], [35] and insula[36]–[38] in AD affected sMRI.

After selecting ROIs, AD-related characteristics features from each ROI can be used for classification of the MRI into different stages of AD. The selection of each ROI is the key to ROI-based analysis methods . We have considered the statistical significance of volumetric measurement of 108 ROIs to select the most important ROI. We have found hippocampus, is the most statistically significant region for AD diagnosis. We verified our selection with clinicians assumptions [26], [27] .

B. Materials and Methods

In GARD cohort study, pure volume(P), intracranial volume(V) and cortical thickness(T) of 108 ROIs were assessed. A global mean cortical thickness measurement for each subject was also computed over the whole cortical surface. These measures are of interest in neuro-degenerative diseases diagnosis. The test-retest reproducibility of each quantitative measure was assessed. From these 108 regions, we have selected hippocampus region based on the distinguishing capacity of the measures of the regions. The distinguishing capacity was measured by p-value, which is obtained by applying permutation-test [39] on the given data. The p-value tests the null hypothesis that PVT measures of a specific region from two different groups of MRIs are identical.

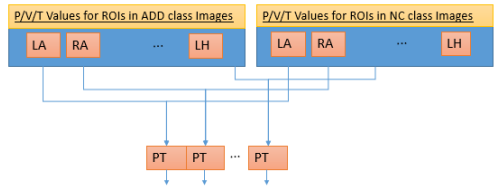


Figure 4: ***ROI selection process; PT is permutation test.***

.

If the p-value is large, there is no reason to conclude that the measurements of the ROI differs for class C_1 to class C_2 (C_1 and C_2 are any two choice from aAD, NC, mAD, ADD and C_1!=C_2). On the contrary, the small overall p-value indicates that the differences we observed are unlikely to be happened from random sampling. Then, we can reject the idea that data for the ROI in C_1 and C_2 groups are identical. Here C_1 and C_2 are considered as two different distributions (class labels) for each of the 108 ROIs.

Algorithm 2: Algorithm for ROI selection

Input: ROIList={Left Hippocampus, Right Hippocampus, Left insula,...}: Volumetric analysis of GARD cohort GroupList=ADD,aAD,mAD,NC V=Volume measures for each ROI of each MRI

Output: R: a set of reference locations,(x,y,z)

```
p_values = {}
for each region of interest ROI ∈ ROIList do
    G = {}
    for i = 1 to GroupList.length do
        j = i + 1
        while j ≤ GroupList.length do
            G ∪ {(x[i], x[j])}
            j = j + 1
    ind_ROI_P_values_for_diff_groups = { }
    for i = 1 to |G| do
        (X_a, X_b) = G.[i]
        t = one_sided_Permutation_Test(X_a, X_b)
        ind_ROI_P_values_for_diff_groups.append(t)
    p_values.append(ind_ROI_P_values_for_diff_groups)
return p_values
```

Table 3: ***p-value measures for different Region of Interests (ROIs)***

Serial No	Region of Interest(ROI)	p-value	Serial No	Region of Interest(ROI)	p-value
1	Left Hippocampus	0.0001	9	Left Middle Temporal	0.0046
2	Right Hippocampus	0.0002	10	Left Entorhinal Cortex	0.0756
3	Left Insula	0.0014	11	Right Thalamus Proper	0.058
4	Right Amygdala	0.0007	12	Right Middle Temporal	0.051
5	Right Insula	0.0014	13	Left Thalamus Proper	0.0588
6	Left Amygdala	0.0076	14	Right Inferior Temporal	0.0346
7	Left Basal Forebrain	0.4438	15	Right Superior Temporal	0.0536
8	Left Superior Temporal	0.0078	16	Left Isthmus Cingulate	0.0334

C. Results

We have presented 16 ROIs according to the statistical significance. To perform statistical analysis, we consider the quantitative analysis report of GARD dataset. The reliability test of the data performed using Cronbach alpha [40]. We have found the data as reliable with Cronbach alpha, $\alpha = 0.80219$. Then, each category of sMRI volume measures for each of the 108 ROIs was assumed as identical distributions. Then, permutation test for each of the 108 ROIs in the dataset was performed. We sorted the p-values in ascending order and selected the ROI with least p-value considering the lower the p-value the more distinct two distributions are. Table 3 demonstrated first 15 lowest p-valued regions. This gives us the ROIs which strongly rejects our assumption about the volume measures. We keep the same ROI for all binary classification tasks.

D. Conclusion

In this study we have selected most significant ROI by using permutation test of atlas-segmented volume data of GARD database. The hippocampi were found to be most significant ROI. Other important ROIs were observed as amygdala, insula, precuneus, etc.

IV. LEARNING HIPPOCAMPUS EMBEDDING IN sMRI

***Abstract**: Hippocampus is a significant landmark for diagnosing neuro-degenerative diseases. In this study, we have attempted to learn the embeddings of this important bio-marker. The learned embeddings of hippocampus play very significant role while we are performing region of interest based structured magnetic resonance image processing. Along with their limitations in addressing scalability issues, conventional metric learning methods for feature embedding is known to lacking in capturing semantic similarity among the data under study. For these reasons , we have trained deep Siamese convolutional neural network for learning deep metric of hippocampus. We have exploited GARD cohort dataset in our study. The input to the network was pair of three view patches (TVPs) of size* $(32 \times 32 \times 3)$ *which were generated from random locations of the brain including hippocampi regions. The positive TVPs are those which are generated from hippocampus location and the rest are negative examples. We have achieved* 98.72% *accuracy in verifying hippocampus TVPs.*

A. Introduction

Hippocampus, a structure of brain's limbic system, is believed to be playing key role in learning-process and memory [41]. Neuro-degenerative diseases causes atrophy in volume and shape of this important structure [42]. Hippocampus looks likes sea-horses as its name suggests [43]. In coronal section the shape is like a peninsula of gray matter(GM) surrounded by white matter(WM) appearing both the hemispheres. We have depicted hippocampus in the figure 5.

This structure under study is said to be an important biomarker for AD and

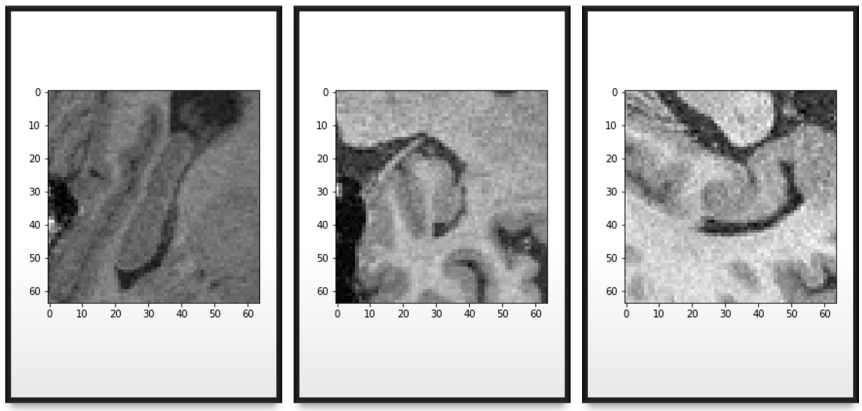


Figure 5: ***Hippocampus in sagittal , axial and coronal view (from left to right)***

.

related diseases. So, this is of great importance to embed hippocampus features and to learn a metric for verifying this important bio-marker. Rather than manual feature extraction methods, we prefer deep metric learning for couple of reasons. Firstly, deep metric learning is capable of finding similarity measures without explicit description of features. Secondly, deep metric learning methods do not require data to be heavily pre-processed. Thirdly, it is very easy to implement and deploy a deep-net framework for wide area of applications ranging from face verification to diseases prediction. Fourthly, zero shot and one shot learning requires very small or no dataset for training the network. Finally, Most of the machine vision problems solved by deep neural networks are showing state of the art performance [44]. For example Face Net [45], Deep Face [46], etc. Machine vision community concentrating in deep distance metric learning since the last few years [47]–[49], and a lot of methods have been devised.

B. Preliminary Concepts

Suppose we are given a dataset X. Two instances of the dataset are x^i and x^j. If we want to measure the similarity or dissimilarity, we need to measure the distance,d of these data points. To measure the distances we use distance metric. Any distance measure needs to have following four properties to be a metric [50].

1. Nonnegativity: $d(x^i, x^j) \geq 0$
2. Symmetry: $d(x^i, x^j) \geq 0$
3. Triangular inequality: $d(x^i, x^j) \leq d(x^i, x^k) + d(x^k, x^j)$
4. Identity of indiscernible: $d(x^i, x^j) = 0$ for $i = j$

The commonly used distance metrics are variants of Chebyshev distance, cosine similarity, bilinear similarity, geodesic distance, etc. But these primitive metrics are sensitive to the scale and dimensions of the features. Furthermore, these cannot use contextual side information for similarity calculation. As a result, for most of the applications which uses these metrics do not provide accurate results.

As an example in fig 6, the conventional metrics are not capable of concluding that the semantically same objects are similar to each other as the semantically different objects are dissimilar. So, we need metric learning algorithms which will incorporate the internal properties of data set as well as consider the user perspectives to find similarity and/or dissimilarity. Facing the limitation of these primitive metrics which do not consider the human perception of similarity/dissimilarity concepts, metric learning algorithms are developed.

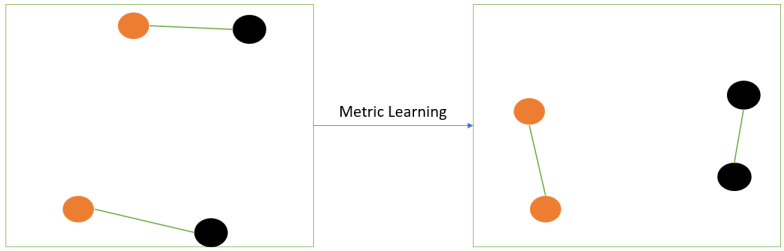


Figure 6: ***same colored bubbles are semantically similar where different colors indicates that bubbles are semantically dissimilar. Metric Learning Algorithm bringing the similar objects nearer while pushing the semantically different object away.***

.

The first metric learning algorithm developed by Xing et all in 2002 [51], basically learns Mahalanobis matrix. The distance is defined by:

$$d(x^i, x^j) = \sqrt{(x^i - x^j)^T M^{-1} (x^i - x^j)} \tag{6}$$

where M^{-1} is the Mahalanobis distance, which is a positive semi-definite matrix that satisfies the metric conditions. The M^{-1} parameterizes the distance. When M^{-1} is identity matrix, the distance is equivalent to Euclidean distance. The Mahalanobis matrix M^{-1} scales the features and utilizes their correlations to compute distances between data more effectively [44]

The main task of conventional distance metric learning algorithms is to learn M^{-1} with the goal of minimizing a constraint cost function. These methods are not powerful enough to capture the nonlinear relationship among data points [52]. Kernel trick is capable of overcoming the problem and are being widely used to implicitly transform the sample data points into a high dimensional feature subspace. Metric learning methods then obtain a metric in the projected feature subspace. In spite of getting feasible solutions, these methods suffer from

the scalability problems as it is difficult to get the explicit nonlinear mapping functions. However, deep metric learning algorithms are good at addressing the nonlinearity and scalability problems which are main problems suffered by conventional metric learning algorithms. The common mechanism of deep based algorithms is to train a deep network for producing an embedding of each input vector so that a loss function related to object distance is minimized. There are several state of the art deep networks that are being used for metric learning. In this study, Siamese network [20] has been used for learning the desired metric.

C. Materials

We have exploited GARD dataset in this study. There are 326 atlas-based segmented sMRI images. We have randomly selected 20 sMRI for training, and 10 for testing. The intensities of the sMRI voxels are normalized so that the mean is zero and variance is one. After normalizing the intensities, we have generated 16 positive TVPs and 16 negative TVPs from each sMRI. From the training TVPs we have randomly selected 32 TVPs (16 positive and 16 negative) for creating a database to be used for evaluating the model with the test set.

D. Methodology

In this study, we have deployed the Siamese network as depicted in figure 7 to learn the deep-metric which can differentiate TVPs of hippocampus from non-hippocampi TVPs. The CNN consist of a pair of networks sharing same weights and loss function. Siamese network learned a function which maps input TVPs into a target space such that the Euclidean distance in the target space approximate the semantic distance between the TVPs. The learning process

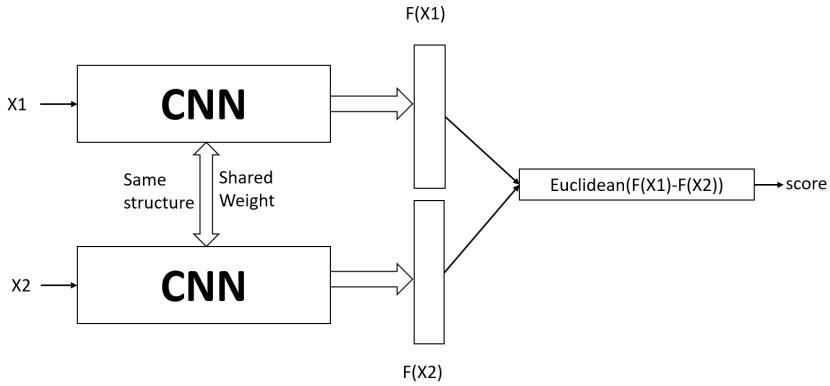


Figure 7: ***Siamese network for metric learning***

.

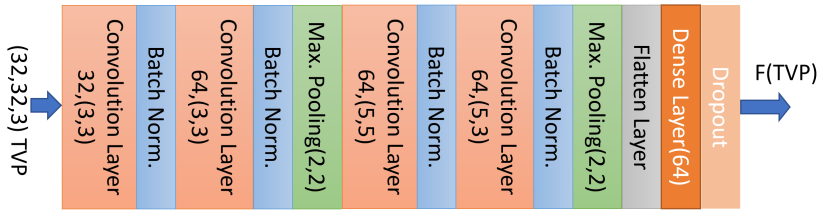


Figure 8: ***One channel in the twin of Siamese network for metric learning***

.

minimizes contrastive [53] loss function which ensures that the similarity metric is small for pair of hippocampus-TVPs and large for distinct-region-TVPs. The CNN works as the mapping function from input to target space. In each channel of the twin,(as depicted in figure 8) there are four convolution layers and one fully connected layer. There is a batch normalization layer after each convolution layer. The last layer of the twin-CNN is the Euclidean distance between the feature embedding of the two different networks.

We have used contrastive loss function for training DML network. The loss

function is defined in equation 7

$$loss(y,\widetilde{y}) = y\widetilde{y}^2 + (1-y)[max(\lambda - \widetilde{y}, 0)]^2 \tag{7}$$

Here, y is the actual distance(0 or 1) and $\widetilde{y}$ is predicted distance between the input pairs. $\lambda(=2)$is used as a distance margin constraint. The constraint defines a radius in target space around euclidean distance. Unlikely pairs have contribution in the loss if their distance is within the defined margin.

E. Experimental Setup

1. Platform

We use the TensorFlow GPU 1.8, keeping Keras as the backend, on top of the Python 3.6 environment. An Intel(R) Xeon (R) CPU E5-1607 v4 @ 3.10 GHz with a 32 GB RAM machine was used. The GPU was NVIDIA Quadro M4000.

2. Dataset Preparation

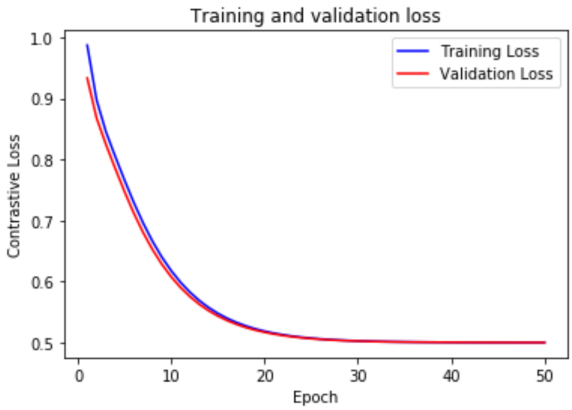
We have generated two different sets of TVPs. The positive samples were randomly produced from 4×4 cube centering at manually labelled Hippocampi locations. The negative samples are produced from other regions of the brain. We have generated 320 TVPs for each set. We have followed pair construction algorithm described in [20]. The pair selection algorithm ensures keeping equal number of similar pairs and dissimilar pairs for both training and testing. Input data consist of a pair of $32 \times 32 \times 3$ TVPs taken from positive and negative samples. A sample for training is $([X_1, X_2], y)$ WhereX_1,X_2 are TVPs and y is the label. y=1 if X_1 is from the same regions as X_2, 0 otherwise. For testing the trained model, we have generated TVPs from the GARD segmented data set. We

have used manually localized hippocampus locations for testing the model on ADNI data set. We kept 16 random TVPS for each class i,e positive and negative class from the training set to form the database. The TVPs in the database are compared with the test TVPs for finding the dissimilarity scores. We have taken the minimum distance score among all the scores for all TVPs in the database and the label of the minimum scored database TVP was considered as the label of test TVP.

3. Training

We have used contrastive loss function for training DML network. The distance margin constraint in the contrastive loss is kept 2. The constraint defines a radius in target space around Euclidean distance. Unlikely pairs have contribution in the loss if their distance is within the defined margin. We have initialized the weight by normal distribution with zero mean and standard deviation 0.01. The biases were initialized for this network from normal distribution with different mean (0.5) and same standard deviation (0.01). For fully connected layer we have initialized biases differently i,e. the mean of the normal distribution was kept zero with standard deviation 0.2. The optimizer used is Adam with a mini-batch size of 32 and initial learning rate is 0.001. The decay of learning rate was kept uniform which is one-tenth if there is no update in the loss for consecutive three epoch. The model has used grid search, to perform hyper-parameter selection. We have trained the presented models for 150 epochs with a batch size of 32. 10-fold cross validation on the training TVPs was performed. The training performance of metric learning network is depicted in figure 9a .

The reason of the better validation performance than the training is that in



(a)

Figure 9: ***Training and validation loss of deep metric learning network***

validation time the dropout layer and regularizers in different layers are turned off. Another reason is that training loss is calculated as an average over all batch-wise losses in each epoch. On the other hand, the validation loss is calculated at the end of each epoch. So, in our case the validation loss is lower than the training loss.

F. Results

For interpreting the score, we consider the multiplicative inverse of the Euclidean distances yielded by the model. The additive factor 1 (one) prevents divide by zero error. If the model output is y for any pair (X_1, X_2), we have transformed y to Y according to equation 8.

$$Y = \frac{1}{1+y} \tag{8}$$

Here, y is the un-normalized Euclidean distance of (i.e., dissimilarity between) two TVPs in the target space and Y is the normalized similarity score in

the range [0,1]. This transformation makes sure the range of similarity is between the interval [0,1] while do not altering the inverse relation between similarity and dissimilarity. For testing the DML network we consider the accuracy in verifying whether the TVPs are containing the hippocampus or not. If the minimum score found with positive class database, we consider the test TVP was taken from hippocampus region, And if the minimum score is found for negative class database then the TVP under observation is considered non-hippocampi. Table 4 presents class label verification results along with confidence score of eight TVPs from different sMRI scans from GARD data set. ANd table 5 represents the same for ADNI data.

The accuracy of the model is calculated based on equation .

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Here, TP is the number of TVPs that are drawn from hippocampus regions and matched with positive class database, TN is the number of TVPs that are drawn from non-hippocampus regions and matched with negative class database, FP is the number of TVPs drawn from non-hippocampi region but matched with positive class database, FN is the number of TVPs drawn from hippocampus regions but matched with negative class database. The total accuracy we have achieved is 98.72% in finding the similar TVPs. We have provided 8 test TVPs with similarity score. The distance scores are normalized to get the similarity scores.

G. Conclusion

The deep distance metric learning algorithms so far are application dependent. Each deep network performs well for one or some applications, while

Table 4: ***Similarity scores (normalized) for eight different TVPs of different sMRI from GARD dataset. The similarity indicates that the presented TVP is similar to the related class in the stored database with given confidence score.***

MRI ID	**Center of TVP**	**Actual Region**	**Verified Region**	**Confidence Score**
14071906	77,127,82	Hippocampus	Hippocampus	0.97
14051804	92,157,84	Hippocampus	Hippocampus	0.89
14080210	66,159,95	Non-hippocampus	Non-hippocampus	0.96
17101603	82,141,80	Hippocampus	Hippocampus	0.86
14051110	122,76,145	Non-hippocampus	Non-hippocampus	0.93
17092001	145,88,133	Non-hippocampus	Non-hippocampus	0.87
15031904	124,72,153	Non-hippocampus	Non-hippocampus	0.91
15031904	95,152,90	Hippocampus	Hippocampus	0.82

Table 5: ***Similarity scores (normalized) for five different TVPs of different sMRI from ADNI data set. The similarity indicates that the presented TVP is similar to the related class in the stored database with given confidence score.***

MRI ID	**Center of TVP**	**Actual Region**	**Verified Region**	**Confidence Score**
112538	75,130,88	Hippocampus	Hippocampus	0.85
137298	82,152,79	Hippocampus	Hippocampus	0.83
124940	46,159,105	Non-hippocampus	Non-hippocampus	0.82
132779	84,133,81	Hippocampus	Hippocampus	0.81
112391	111,72,155	Non-hippocampus	Non-hippocampus	0.92

performance deteriorate for other applications. Despite their above-mentioned limitation, the proposed Siamese network architecture provides robust accuracy in learning hippocampus features. We have observed 98.72% accuracy in verifying hippocampus TVPs by the proposed Siamese model. This achievement has further application in sMRI processing.

V. ROBUST LANDMARK LOCALIZATION

***Abstract**: Accurate landmark-localization in 3D structured magnetic resonance imaging of human brain is a challenging problem. In this study, we have proposed deep-metric aware cascaded Hough convolutional neural network approach for localizing hippocampus in brain sMRI. The process combines three steps: 1) rough estimation of the hippocampi location with the aid of a trained HCNN model which accepts three view patches of size* $128 \times 128 \times 3$ *from the spherical surface centering the middle voxel of the sMRI. The output of HCNN is the offsets of the hippocampi from the center of TVP. 2) The second step is the verification of the locations by using a pair of deep metric learning network. 3) The last step is to find the fine-grained locations by using a HCNN which takes TVPs from the coarse-grained locations and yields offsets. From these offsets we computed the final hippocampus locations. We observed that our method consistently produces accurate hippocampus locations with mean localization error 1.04 mm and 1.37 mm for left and right hippocampus respectively. We have performed our analysis on GARD dataset.*

A. Introduction

Landmark localization in 3D medical imaging may be categorized into atlas-based approaches and learning based methods. First one requires one or multiple atlas with predefined regions and suffers from computing time and sometimes accuracy. The second method deploys machine learning approaches to devise a model and are demonstrating superior performance now-a-days. These types of methods either depends on classification or regression tasks. Classification models extract patches from a voxel and classifies it as a landmark or not while regression-based methods develop regression model to predict 3D displacement

from a local voxel to a target voxel by learning the non-linear relationship between these two voxel features. The last one created the hope of using any local patch for estimating a potential landmark position from given local image patches. There are methods which considers global information along with local correlations. Jung et. al. proposed two stage hough convolutional neural network [54] considering both local and global correlation. But global predictions are not verified whether the model providing the expected landmark offsets. In this study we are proposing deep metric based verification strategy to perform robust localization.

B. Methods

Our localization involves three tasks.

1. Rough Estimation: Prediction of offset from the level one network which takes large scale TVP input. We name it far jump hough convolutional neural network(FJCNN). As our CNN predicts the offsets just like hough forest by learning the geometric shape of the hippocampus we name it hough-CNN.

2. Confirmation: Similarity measure with aid of learned metric whether the TVP generated from predicted offset is significant, i,e whether the offset is really indicating the hippocampus location. If not the previous step is repeated with next random TVP. The random reference point is generated from the spherical surface.

3. Fine Tuning: Second level CNN in the cascaded structure predicts the offset from the average predictions from the neighboring TVPs of the location

generated from the previous prediction.

1. Far Jump Hough Convolutional Neural Network

The network is depicted in figure 10. FJHCNN architecture was chosen by inspiring from [55]. The network consist of eight convolution layers, four max pooling layers, five batch normalization layer. The flatten layer follows two dense layers. The first dense layer is followed by a dropout of 0.25. The input to this model is TVPs of size $128 \times 128 \times 3$. A sphere of radius 8 was assumed and random locations were produced from the surface of the sphere to produce TVPs. The network predicts initial estimates of the offsets from the center of TVPs. Batch normalization and dropout layers limits the chance of over-fitting. The TVPs used in this level contain almost all the 2D-slice in axial, coronal and sagittal views excluding the boundaries. So, the FJHCNN reasonably capable of predicting the offsets. But if the hippocampus is not present in the TVPs it is hard to learn the geometric correlation with its neighboring structures. We have used leaky relu as activation function for all the convolution layer with $\alpha = 0.3$ In the first fully connected layer we have used tanh as the activation. In the last layer we have used leaky relu with $\alpha = 0.9$ which is almost same as linear activation.

2. Verification with Deep Metric

The FJHCNN provides a pair of offsets which are associated with left and right hippocampi locations. After computing the locations from theses offsets ,TVPs of size $32 \times 32 \times 3$ are from the locations. Then, separate verification is conducted for left and right hippocampi. The verification data cosist of hippocampii-TVPs from diverse classes(ADD,NC,aAD,mAD) and genders. The siamese net tells

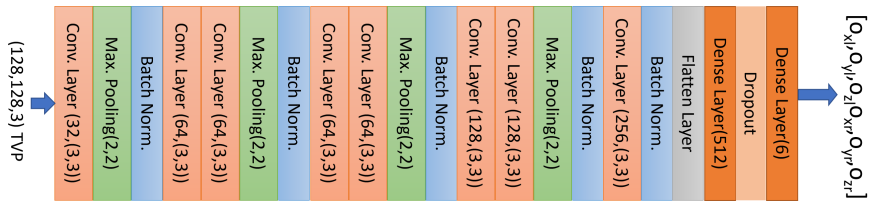


Figure 10: ***Far Jump hough convolutionl neural network(FJHCNN). The network roughly predicts the offsets of the left and right hippocampus from the center of Three view Patch(TVP)***

.

the significance of the proposed locations by yielding similarity score aand association with the class label(hippocampus/ non-hippocapus). If the associated class label is hippocampus then the predictions of FJHCNN are queued for further processing. If associated class label is non-hippocampus then next run of FJHCNN were conducted to get new predictions.

3. Short Jump Hough Convolutional Neural Network

The network is depicted in figure 11. The network consist of five convolution layers, two max pooling layers, three batch normalization layers. The flatten layer follows two dense layers. The first dense layer is followed by a dropout of 0.25. The input to this model is TVPs of size $32 \times 32 \times 3$. A cube of size $8 \times 8 \times 8$ was assumed centering on the FJHCNN-predicted voxel locations. From the cube, Gaussian random voxel positions were generated by algorithm refreferencePointGenerationAlorithm. This network predicts fine-tuned estimates of the offsets from the center of TVPs. Batch normalization and dropout layers limits the chance of over-fitting. The TVPs used in this level considers local correlation between landmark location and its neighboring voxels.

The models learn the geometric correlation with its neighboring structures. The activation function used in this network follows same pattern of FJHCNN.

Figure 11: ***Short Jump hough convolutionl neural network(SJHCNN). Two different network of this kind predicts fine tuned offsets of the left and right hippocampus from the center of three view patch(TVP)***

.

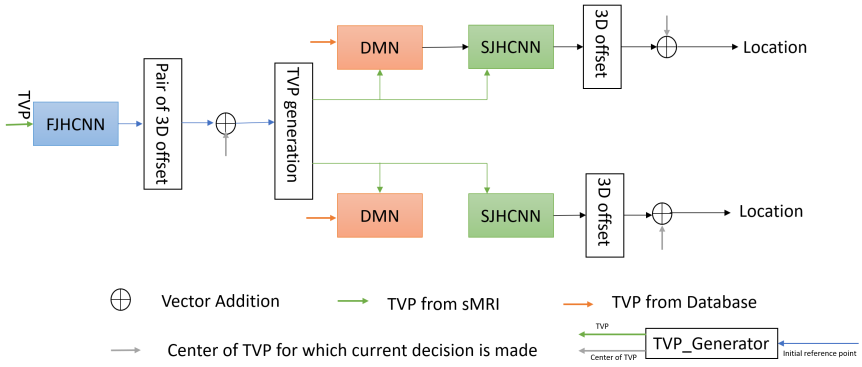


Figure 12: ***Hippocampii localization using CNN and deep metric learning***

.

C. Experimental Setup

1. Platform

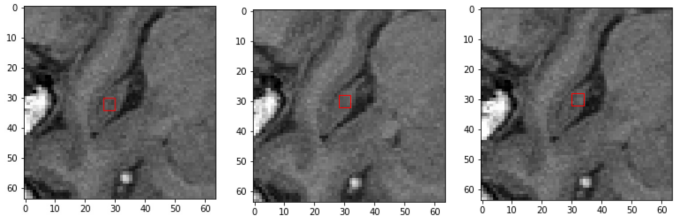
We use the TensorFlow GPU 1.8, keeping Keras as the backend, on top of the Python 3.6 environment. An Intel(R) Xeon (R) CPU E5-1607 v4 @ 3.10 GHz with a 32 GB RAM machine was used. The GPU was NVIDIA Quadro M4000.

2. Training Dataset Preparation

We have performed the localization task on GARD dataset. Data preparation for different tasks involved in the whole process is described below.

For the first level CNN, we have assumed a sphere of radius 8 centering at the middle voxel of the MRI scan. Then, we have generated TVPs , from the voxels on the surface of the sphere. We have selected the spherical surface to ensure that TVPs generated from some voxels might contain hippocampus. For each TVP there is an associated offset which indicates the distance of the hippocampus in each direction from the center of TVP. The size of the TVPs are $128 \times 128 \times 3$. The ground truth is $a_{lh}(x,y,z), a_{rh}(x,y,z)$ which represent the offset of hippocampii from reference point $r(x,y,z)$. Input to first level CNN a TVP of size $128 \times 128 \times 3$)

For the second level CNN, we have manually localized hippocampus locations with the aid of expert physician. Sample locations are presented in figure 13. We have considered a cube of size $8 \times 8 \times 8$ centering the manually localized voxel. Then, we have generated TVPs and offsets from the voxels inside the cube by using algorithm 1. The size of the TVPs are $32 \times 32 \times 3$. The ground truth are $(a_{lh}(x,y,z)$ or $a_{rh}(x,y,z))$ which represent the offsets of hippocampii from reference points $(r_{lh}(x,y,z)$ or $r_{rh}(x,y,z))$.



(a) *(b)* *(c)*

Figure 13: ***An example of ground-truth locations of hippocampus in three sMRI (viewed on the sagittal plane).***

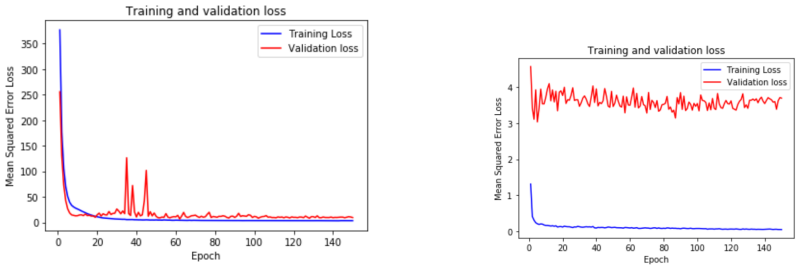
3. Training

We have initialized the weight by normal distribution with zero mean and standard deviation 0.01. The biases were initialized for this network from normal distribution (mean: 0.5 and standard deviation: 0.01). For fully connected layer we have initialized biases differently i,e. the mean of the normal distribution was kept zero with standard deviation 0.2. Adam optimizer was used with a mini-batch size of 32 and initial learning rate 0.001. Other parameters were kept in default settings of the original paper of Adam.

We have used heterogeneous learning rate for different layers of the model. The decay of learning rate was kept uniform which is one-tenth if there is no update in the loss for consecutive three epoch.

The models were trained for 150 epochs.

For FJHCNN and SJHCNN networks we have used mean squared error as the loss function described in the equation 10. The FJHCNN network receives a TVP and a pair of offsets $(a_x^{lh}, a_y^{lh}, a_z^{lh})$ and $(a_x^{rh}, a_y^{rh}, a_z^{rh})$. It yield a pair of offsets $(p_x^{lh}, p_y^{lh}, p_z^{lh})$ and $(p_x^{rh}, p_y^{rh}, p_z^{rh})$.



(a) *(b)*

Figure 14: ***Training loss of (a) Far Jump HCNN (b) Short Jump HCNN***

The SJHCNN network receives a TVP and an offset $(a_x^{lh}, a_y^{lh}, a_z^{lh})$ or $(a_x^{rh}, a_y^{rh}, a_z^{rh})$ for left or right hippocampus respectively. It yield an offset $(p_x^{lh}, p_y^{lh}, p_z^{lh})$ or $(p_x^{rh}, p_y^{rh}, p_z^{rh})$ for left or right hippocampus respectively.

$$mse = \frac{\sum_{i=1}^{batchsize}(a_i - p_i)^2}{batchsize} \tag{10}$$

Here, $batchsize = 32$ is the number of input-output pair in a batch and a_i and p_i are tuples representing actual offsets and predicted offsets.

The training performance of both the networks are depicted in figure 14a and 14b

D. Performance Analysis

The output of the networks were added with the center of TVPs for which the offset is predicted by the network. Then, we have found the location of the landmark. To calculate the error we have consider the euclidean distance between the manually identified locations and the calculated locations. The predicted locations are calculated using equation 11:

Table 6: ***Test error in millimetre(mm) for hippocampus localization.***

MRI Scan	Left Hippocampus		Right Hippocampus	
	Average Error of 1st Level CNN Model	Error of 2nd Level CNN Mode	Average Error of 1st Level CNN Model	Error of 2nd Level CNN Mode
14050407	3.63	1.74	2.29	1.36
14062105	3.52	0.54	3.87	1.75
15031902	2.92	0.61	2.25	1.09
15031904	3.86	1.63	1.86	0.63
16050301	3.72	1.26	3.55	1.72
16061303	3.19	0.68	2.11	1.36
14092401	3.67	0.89	3.92	1.82
14092707	2.85	1.57	3.67	1.69
15031905	3.51	1.13	3.57	0.84
15032504	3.02	0.32	2.93	1.31
Average	3.39	1.04	3.00	1.357
standard deviation	0.34	0.48	0.76	0.37

$$pl_x = p_x + r_x \tag{11a}$$

$$pl_y = p_y + r_y \tag{11b}$$

$$pl_z = p_z + r_z \tag{11c}$$

The error was calculated using equation 12

$$error = \sqrt{(al_x - pl_x)^2 + (al_y - pl_y)^2 + (al_z - pl_z)^2} \tag{12}$$

For localizing left hippocampus the average error of the first level CNN network was 3.39 ± 0.34 mm while the error is 3.00 ± 0.76 for right hippocampus. The average error for the second level CNN for left hippocampus was 1.04 ± 0.48 mm and for right hippocampus it was 1.357 ± 0.37 mm. Table 6 shows the detail results of our experiment.

E. Conclusion

Siamese network along with cascaded HCNN provides robust localization performance . Our proposed pipeline demonstrated 1.04 ± 0.48 mm error for localizing left hippocampus and 1.357 ± 0.37 mm error in localizing right hippocampus. The performance of the model is seemed to be invariant to geometric transformation of the sMRI scan.

VI. PATCH-BASED CLASSIFIERS FOR ALZHEIMER DISEASE DIAGNOSIS

***Abstract**: The state-of-the-art techniques either consider multimodal or unimodal-sMRI for diagnosing AD and its prodromal stages. Despite their over human-level accuracy, collecting data for these methods are time consuming, expensive and some modalities are harmful as these may have radioactive side effects. This study is confined to hippocampus regions in structural magnetic resonance imaging (sMRI). The objectives of this attempt are as follows: to increase the accuracy level that is comparable to the state-of-the-art methods; to overcome the over-fitting problem, and; to analyze the performance of hippocampal features for AD diagnosis. To achieve the objectives, at first, we localized the hippocampus, then, incorporate ensembles of simple convolutional neural networks (CNNs) for classifying different stages of AD. We deployed a patch-based classification approach as it is fast to train, simple to implement, and flexible enough for easy deployment. We have performed our experiment on the Gwangju Alzheimer's and Related Dementia (GARD) cohort dataset prepared by the National Research Center for Dementia (GARD), Gwangju, South Korea. We localized the left and right hippocampus and fed three view patches (TVPs) to the CNN after the preprocessing steps. We achieve* 90.38% *accuracy.*

A. Introduction

Recently, patch-based techniques are widely being used in medical imaging. Its applications areas span from segmentation, noise removal, super-resolution, anomaly detection, disease diagnosis to image synthesis and many more. From the inspiration of these wide range of applications, we have used three view

patches (TVPs) from the ROI of the sMRI for diagnosing AD and its prodromal stages. The TVP-representation of the ROI is in between the whole sMRI and localized voxels representations. This representation scheme was observed to be successful in metric learning and localization as presented in previous experiments of us. Here we use the same representation for classification of sMRI scans into different class labels.

In our classification task, we have used three view patches(TVPs) for following reasons:

1. TVPs are in between of whole sMRI-based global representation and individual voxel based local information.

2. It is well known that CNNs are highly susceptible to the sample size. The more samples we have from each class, the more accurate the CNN performs. Classification accuracy is subject to the discriminating features among the available classes [56]. The availability of discriminating features of a class depends on the number of samples from the class. The main problem of AD diagnosis is the scarcity of data. We have a limited number of samples from each class. This scarcity of data may lead to an over-fitted model. Therefore, we deployed a patch-based classifier which facilitates generating a sufficient number of patches for training.

3. Patch-based processing assists us designing simpler CNN model.

4. Successful deployment of patch-based classifiers in [57], [58] and [57], [59]–[62] inspired us attempting the experiments with TVPs.

B. Patch-based Classifiers

The proposed pipeline is depicted in figure 15. Our framework consists of three individual models for generating decision scores on individual patches, followed by a score aggregator and final classifier.

After collecting data, we performed the preprocessing tasks, as stated in the previous studies. Then, we performed localization of the left and right hippocampus, which we consider as the ROI for our experiment. Then, from the ROI, we generated TVPs of size $32 \times 32 \times 3$ or $64 \times 32 \times 3$. TVPs along with the labels of MRI are considered as a data unit to feed into the CNN. The CNN is trained to predict individual TVPs as one of the given classes. Number of TVPs generated for a ROI is the sample size for training and validation of a patch-based classifier. So, we can generate large number of samples for training that leads to mitigate data scarcity problem. We trained three individual models with these generated patches. The CNNs are trained to predict individual TVPs as between two given classes.

The input to the CNN is a $32 \times 32 \times 3$ TVP. The output is softmax score of $[C_1\ C_2]^T$ which is a classification score of the TVP. C_1 and C_2 may be $aAD, mAD, NC, ADD\ and\ C_1 != C_2$. The kernel sizes of the convolutions layers were less than or equal to 7×7 to extract detail information about the ROI under consideration. The activation function in the convolution layers were rectified linear unit (ReLU) [63]. Pooling operation [64] down samples the patches by a factor of specified stride. To enforce normally distributed output of each convolution layer, we use batch normalization [65] before these layers.

The output of the last convolution layer is the feature embedding of the ROI under observation. These features are flattened and then feed to the fully

connected layers for classification purpose. We deployed dropout [66] of 0.25 in the first fully connected layer. The last layer activation was softmax [67]. The loss function was cross-entropy. Xavier initialization [68] technique was used for initializing weights. For optimization, we used Adam optimizer [69] with its default settings.

At first, we designed the classifier for left hippocampus. Then, the same bare architecture was trained for right hippocampus classification. As the input size is different for both hippocampi classifier (i.e., LHRH model), we had to tweak the architecture of the related model. The performance of each model was measured individually. These three models were then added together, and a SoftMax classifier was used for the final distinction.

The proposed pipeline is depicted in Figure 15. Our framework consists of three individual models for generating decision scores on individual patches, followed by a score aggregator and final classifier.

1. CNN for the single Hippocampus

The model for the single (left or right) hippocampus classification is presented in figure 17. We tried different structures and hyper-parameters. We determined the proposed network after several trials. There are three convolution layers and two fully connected layers in the model. Each convolution layer and fully connected layer are preceded by batch normalization excluding first and last layers. First and second convolutions are followed by the average pooling layer. The inputs are normalized previously. Before the last fully connected layer, we used a dropout of 0.25, which converges the training process faster and increases the accuracy. The output of the last convolution layer is the feature embedding of the hippocampus

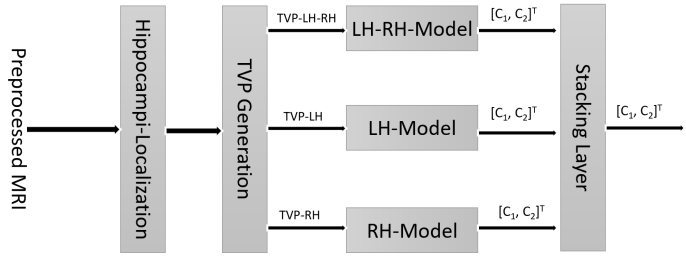


Figure 15: ***Input to the localization phase is the preprocessed sMRI. The output of this phase is a pair of 3D locations for left and right hippocampi. TVP Generation phase uses these locations as center of a cube of size*** $8 \times 8 \times 8$ ***for generating TVPs. The size of the TVPs produced from the left hippocampus (TVPLH) and right hippocampus (TVPRH) is*** $32 \times 32 \times 3$***; the size of merged TVP from both the hippocampi (TVPLHRH) is*** $64 \times 32 \times 3$***; left hippocampus classifier (LH-Model) and right hippocampus classifier (RH-Model) are pretrained CNN models that take TVPs as input and yields a softmax score for each TVP; both hippocampi classifier(LHRH-Model) is another pretrained CNN model that takes TVPLHRH of size*** $64 \times 32 \times 3$ ***and yields a softmax score; The scores are summed up and normalized by a softmax classifier in stacking layer to obtain the final label.***

.

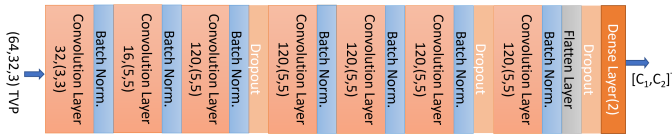


Figure 16: ***Convolutional neural network(LHRH Model) for classifying merged TVP(LHRHTVP)***

.

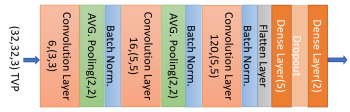


Figure 17: ***Convolutional neural network(LH Model or RH Model) for classifying LHTVP or RHTVP***

.

region under study(i,e., left or right hippocampus). These features are further fed to the fully connected layers to classify C_1 versus C_2. Adding a dropout of 0.25 in the first fully connected layer improved the accuracy.

We used softmax as the last layer activation and cross-entropy as the loss function. The Adam optimizer [69] and Xavier initialization [68] were used. The exponential decay rate for first and second moment estimates are 0.9 and 0.999 respectively. The architecture and structural details of the proposed CNN are noted in Table 17. Total number of trainable parameter in the network is 105,826. The architecture and structural details of the proposed CNN for the right hippocampus classification are noted in figure 17. Total number of parameters in the network is 100,197 among which 99,925 parameters are trainable.

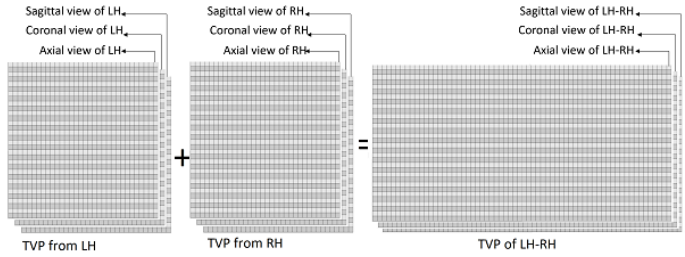


Figure 18: ***To prepare training and testing samples (of size*** $64 \times 32 \times 3$***) for the patch-based left and right hippocampus classifier (LHRH model), each TVP (of size*** $32 \times 32 \times 3$***) from the left hippocampus is merged with the corresponding TVP (of size*** $32 \times 32 \times 3$***) of the right hippocampus.***

.

2. CNN for the both hippoicampi Classification

The architecture of the proposed CNN for the classification of both hippocampi is shown in Figure 16. There are seven convolution layers. Each follows batch normalization and/or drop out. It takes input of size $64 \times 32 \times 3$. We merged the TVPs of size $32 \times 32 \times 3$ from the left and right hippocampus to generate these input patches. We illustrate the merging operation in Figure 18. The output of the seventh convolution layer is the feature embedding of the hippocampus region. These features are further fed to the fully connected layers to classify AD versus NC. The total number of parameters for this network is 409,666. The Adam optimization [69] and Xavier initialization techniques [68] were used in this model.

3. sMRI Classification

We have considered individual sMRI classification with the aid of algorithm 3.

Algorithm 3: Algorithm for Ensemble Decisions

Input: Data: MRI Image Volume; R_L, R_R: two sets of reference locations,(x,y,z) for left and right hippocampus, respectively.

Output: $DS(DS[C_1], DS[C_2])$: Decision scores of an MRI

Data: Let $N = |R_L| = |R_R|$ be the number of patches sampled from a hippocampus of an MRI. LHMODEL(), RHMODEL(), LHRHMODEL() returns the decision scores for individual TVPs as $(s[C_1], s[C_2])$

1 $tvplh = TVP\_Generator(R_L)$

// $TVP\_Generator(R_L)$ returns $32 \times 32 \times 3$ TVP centering at the locations $\in R_L$

2 $tvprh = TVP\_Generator(R_R)$

3 $tvplhrh = merged\_TVP\_Generator(R_L, R_R)$

// $merged\_TVP\_Generator(R_L, R_R)$ returns TVPs of size $64 \times 32 \times 3$ generated from pairs of TVPs of size $32 \times 32 \times 3$ centering at the pair of locations (l,r). Here $l \in R_L$, $r \in R_R$ and the function for mapping the corresponding locations is, $F : R_L \rightarrow R_R$ is one to one and onto.

4 $s_l = LHMODEL(tvplh)$

5 $s_r = RHMODEL(tvprh)$

6 $s_{\text{lr}} = LHRHMODEL(tvplhrh)$

7 $score_{[C_1]} = \sum_{m \in \{l,r,lr\}} \sum_{i=1}^{|N|} \mathbf{s}_m^{(i)}[C_1]$

8 $score_{[C_2]} = \sum_{m \in \{l,r,lr\}} \sum_{i=1}^{|N|} \mathbf{s}_m^{(i)}[C_2]$

9 $DS[C_1] = \frac{e^{score_{[C_1]}}}{e^{score_{[C_1]}} + e^{score_{[C_2]}}}$

10 $DS[C_2] = \frac{e^{score_{[C_2]}}}{e^{score_{[C_1]}} + e^{score_{[C_2]}}}$

11 return DS

For each of the TVPs $(x_1, x_2, x_3, ...x_n)$ if the model predictions are $(y_1, y_2, y_3, ..., y_n = Y_1)$ where each $y_i; i \in 1, 2, 3, ..., n$ represents two different scores $y_i^{C_1}$ and $y_i^{C_2}$ in favor of class C_1 and C_2.

$$Y^{C_1} = \sum_{i=1}^{n} \mathbf{y}_i^{C_1} \tag{13a}$$

$$Y^{C_2} = \sum_{i=1}^{n} \mathbf{y}_i^{C_2} \tag{13b}$$

$$Y_{mri}^{C_1} = \frac{e^{Y^{C_2}}}{e^{Y^{C_1}} + e^{Y^{C_2}}} \tag{14a}$$

$$Y_{mri}^{C_2} = \frac{e^{Y^{C_2}}}{e^{Y^{C_1}} + e^{Y^{C_2}}} \tag{14b}$$

Here, equation 13 aggregates all the decision scores in favor of C_1 and C_2 class label respectively . These scores are softmax normalized in equation ref 14. The scores for the class label of aAD and mAD are also determined in the same way. Equation 14 determines the label of MRI.

C. Experimental Setup

1. Data set Preparation for Patch-based CNN Classifiers (PBCNNC)

For any given sMRI X_1 with label Y_1 , at first, we have performed the localization of the ROIs (here hippocampus) in the pre-processed sMRI. Then, from the localized regions, we have generated TVPs $(x_1, x_2, x_3, ...x_n)$ with label$(y_1, y_2, y_3, ..., y_n = Y_1)$. We have done this for all sMRI in the data set for

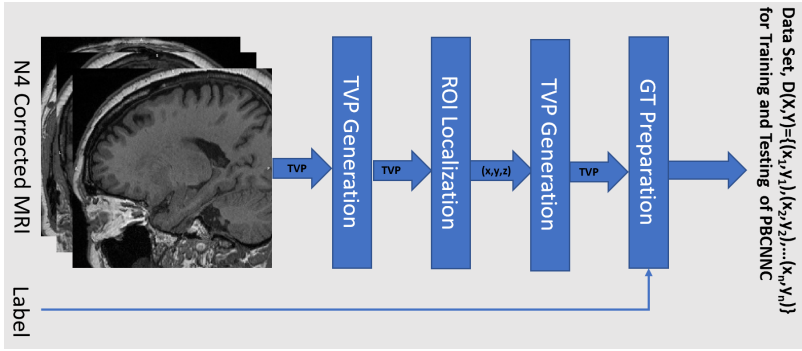


Figure 19: ***Ground truth preparation for patch-based convolutional neural network classifiers(PBCNNC)***

.

training and testing. The process of preparing the dataset for PBCNNC is depicted in figure 19

2. Data set separation

From the ADNI dataset, we consider only those subjects whose disease status remains the same over different MRI scans. We have selected a total of 60 subjects from our dataset. For each subject, there are different MRI scans. We separate the training, testing and validation set in such a way that the conjunction of any two sets, keeping the subject ID of MRI scans as the key, yields the null set. This ensures the prevention of data leakage. We also ensure that MRI scans from each class are uniformly distributed among the three sets to address the class imbalance problem. We keep 60% of MRI scans as the training set, 20% for the test set and 20% for the validation set. We augmented the data of each class by applying shearing, re-scaling and zooming of the patches.

All the MRIs in the GARD dataset are baseline MRI scans, so we did not need

to separate the MRIs according to patients. We divided the dataset into training, validation and a test set according to the procedure that we followed for the ADNI dataset separation. We also applied the same data augmentation techniques to the GARD dataset.

3. Platform

We use the TensorFlow GPU 1.8, keeping Keras as the backend, on top of the Python 3.6 environment. An Intel(R) Xeon (R) CPU E5-1607 v4 @ 3.10 GHz with a 32 GB RAM machine was used. The GPU was NVIDIA Quadro M4000.

4. Training

For patch-based classification, we trained different architectures with different hyper-parameters. The presented models were trained for 20 epochs with a batch size of 32. We followed the 60-20-20 approach for using sample patches for training, validation and testing. We started the training with a learning rate of 0.001. If the validation loss stopped improving for 3 consecutive epochs, we reduced the learning rate by a factor of 10. It was observed that the learning rates were between 0.001 to 0.0001. The default parameter settings were used for the optimizers, regularizers and constraints.

We used 3-fold cross-validation for training the PBCs. Every other settings are as like [58]. At first we have trained the bare model for AD/NC classification. Then, we re-trained the model for AD/aAD. The AD/aAD model was retrained for AD/mAD classification task. This model was retrained for classifying mAD vs aAD . Then, we retrained the previous model for diagnosing aAD from NC.

D. Results

We have evaluated 6 different models trained for six different classification tasks for each of the TVP type(left, right and merged). The evaluation outcomes are summarized in table 7. The reported results were found by feeding a balanced number of samples from each pair of classes.

1. Evaluation Metric

For evaluating the model performance we have considered equations 15, 16 17, 18

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

$$precision = \frac{TP}{TP+FP} \tag{16}$$

$$recall = \frac{TP}{TP+FN} \tag{17}$$

$$f1\,score = 2 \times \frac{precision \times recall}{precision + recall} \tag{18}$$

Here, TN, TP, FN and FP are acronyms for the number of model-predicted true negative, true positive, false negative and false positive samples, respectively.

For evaluating each model, we used an individual TVP as a sample. We generated TVPs from each ROI from test MRIs . We feed TVPs to the patch-based classifiers to obtain the decision scores for each individual TVP. If the obtained score is greater than 0.5, we labeled the MRI as class 1; otherwise, we labeled it as class 2.

Table 7: ***Classification performance of left and right Hippocampi based features on GARD database***

ROI	AD/NC	AD/mAD	AD/aAD	mAD/aAD	mAD/NC	aAD/NC
LH	84.31	78.26	82.61	73.33	79.07	54.76
RH	82.35	78.43	80.39	72.55	79.07	53.49
LHplusRH	88.24	83.33	87.50	80.00	81.40	56.29
Ensemble	90.38	87.50	82.61	80.00	82.35	57.14

2. Left Hippocampus Region-based Classifiers:

For the left hippocampus features, we have observed 90.73% accuracy in classifying AD over NC. The precision and recall for this ROI is 90.17% and 90.90% respectively. The f1-measure performance was 90.47%. The figure 20 demonstrates the classification performance of all pair of classes. We observed that left hippocampus atrophy demonstrated superior distinctive features for diagnosing AD and its pro-dromal stages than others regions. We have also observed that left hippocampus provides very little or no decisive features for aAD diagnosis over NC. The diagnostic accuracy of mAD subjects from AD, aAD and NC classes are 84.77%, 73.51% and 78.15% respectively.

3. Right Hippocampus Region-based Classifiers:

The right hippocampus model accurately diagnosed 86.75% of the MRIs . 86.12% was correctly diagnosed as AD-affected, and a total of 86.78% of the AD diagnosed MRIs are actually AD affected. This region was observed to classify mAD MRI form aAD and NC with a nearly 70.20% and 78.81% accuracy,respectively; though the classification performance between aAD and

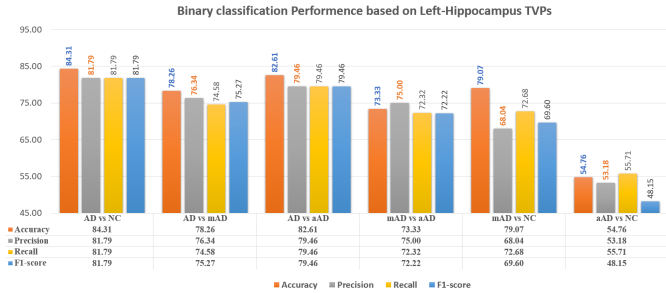


Figure 20: ***AD/NC, AD/mAD, AD/aAD, mAD/aAD, mAD/NC and aAD/NC classification performance based on test TVP generated from left hippocampus(LH)***

.

NC is 54.30%.Right hippocampus also provided useful information to classify AD vs mAD with yielding 82.12% of accuracy. AD vs aAD classification performance was observed to be 84.56%.

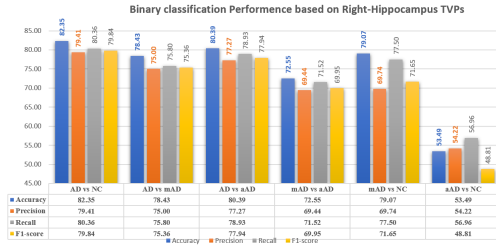


Figure 21: ***AD/NC, AD/mAD, AD/aAD, mAD/aAD, mAD/NC and aAD/NC classification performance based on test TVP generated from right hippocampus(RH)***

.

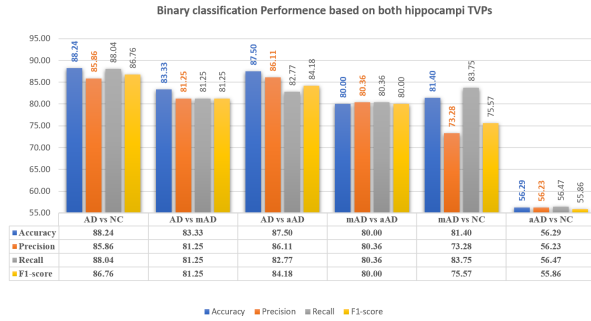


Figure 22: ***AD/NC, AD/mAD, AD/aAD, mAD/aAD, mAD/NC and aAD/NC classification performance based on test TVP generated from both the hippocampi***

.

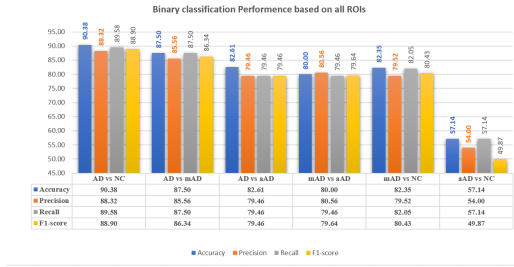


Figure 23: ***AD/NC, AD/mAD, AD/aAD, mAD/aAD, mAD/NC and aAD/NC classification performance based on ensemble classifiers***

.

4. Both Hippocampi Region-based Classifiers:

The diagnostic accuracy from the merged-TVPs of both the hippocampi was observed to be better than single hippocampus-based experiments. The reason may be that the combined TVPs provide more distinctive information.

This model accurately diagnosed 88.24% of the test sMRIs.By this model 86.12% was correctly diagnosed as AD-affected, and a total of 86.78% of the

AD diagnosed MRIs were actually AD affected. This model was observed to be classify mAD MRI form aAD and NC with a nearly 80.00% and 81.40% accuracy,respectively; though the classification performance between aAD and NC is 56.29%. Combined TVPs also provided useful information to classify AD vs mAD with yielding 83.33% of accuracy. AD vs aAD classification performance was observed to be 87.50%.

5. Ensemble Classifiers:

After training and evaluation all 18 different models we have stacked the outcome of each 3 models which shares the same objective. We have found that ensembling the scores of three different models improves the diagnostic accuracy.

From the ensemble model, we achieved 90.38% accuracy in classifying AD over NC. The precision and recall of this are 88.32% and 89.58% respectively. The f1-measure performance was 88.90%. The figure 23 demonstrates the classification performance of all pair of classes. We observed that the ensemble model demonstrated superior performance in classifying AD and its pro-dromal stages. We have also observed that hippocampus provides very little or no decisive features for aAD diagnosis over NC. The diagnostic accuracy of mAD subjects from AD, aAD and NC classes are 87.50%, 80.0% and 82.35% respectively.

E. Conclusion

This study provides an efficient framework for AD diagnosis from brain sMRI. We have observed that the hippocampus, which is known to be one of the most affected clinically studied biomarkers for AD detection, is providing

significant features to diagnose AD and its prodromal stages. For the two different hippocampi in the brain, we have deployed two patch-based classification models. However, deployment of another model for classifying both hippocampi increases the performance. We then designed ensemble models for an improved classification outcome. We designed the CNN classifiers based on TVPs on the semi-randomly generated locations in the vicinity of the hippocampus region localized by metric-aware hough convolutional neural network. This approach facilitated generation of the necessary data for training and testing. After sufficient training, we combined the models to obtain the expected accuracy (90.38% for GARD), which is comparable to the models designed in the sMRI modality.

VII. RELATED WORKS

A. Introduction

There are different approaches for analyzing sMRI to detect AD which varies from traditional machine learning techniques to modern deep learning methods [70]. In traditional machine learning methods, handcrafted features from voxels, regions or patches are extracted and then a classifier is used to classify the label of the sMRI [71]. The main problems are i) handcrafted features may not well co-ordinated with classifiers which may leads to sub-optimal diagnostic performance, ii) specific regions/patches may not provide adequate information which may characterize the global brain structural information. The deep learning methods address the first problem by automatically extracting features despite processing the whole brain which is computationally expensive. Accepting the second limitation as true, we have considered the heuristics from clinicians as well as volumetric analysis obtained from atlas-based segmentation study in GARD database about the more responsible regions for AD diagnosis. The primary goal of us to investigating the sMRI as a dignostic modality for aAD. We also have verified the suspicions of clinicians about the significance of the stated pre-defined regions in diagnosing AD by using deep learning methods. In our study, convolutional neural networks based classifications models showed state-of-the-art AD-diagnostic performance based on automatically extracted features from predefined ROIs.

The studies [26], [27] on the magnitude and spatial pattern of AD acquired on histological or imaging data is useful for CAD based AD-diagnosis from sMRI. The prior knowledge about AD affects brain regions , such as atrophy of the hippocampus [28], [29], the entorhinal cortex [30]–[32], expansion of the

ventricles [33], and volumetric changes in amygdala [34], [35], insula [36]–[38] are considered for our experiments.

Three view patches of these important regions containing discriminatory AD-related information are then used for training the CNN. In this study, we propose a simple 2D-CNN models to extract discriminating information from three view patches of specific predefined-regions, then a softmax classier is used for binary classifications of each TVP.

The structural variation between the hippocampus, one of the structures of the medial temporal lobe, of AD and healthy individuals studied intensively and found to go through severe structural changes in AD individuals.

Therefore, features of this region is sometimes used directly for classifying the subjects into normal and diseased classes. Regions of interest (ROIs) based approaches to measure anatomical atrophies of brain MRI are (becoming) popular for automatic AD diagnosis. After selecting ROI, AD-related characteristics features from each ROI can be used for classification of the MRI into AD/NC/MCI. The selection of each ROI is the key to ROIs based analysis methods. We have considered the statistical significance of volumetric measurement of 108 ROIs to select the top most important ROIs. We have found hippocampus, amygdala, insula, isthmus cingulate as more statistically significant. This study confine to hippocampus feature-based classification only.

B. Comparison

There are studies based on the conventional machine learning techniques which focused on developing models to detect anatomical and functional disorders due to AD in human brain [72]–[78]. These methods have primarily relied

on manually designed features, which heavily depend on professional expertise, require repeated trials, and tend to be time-consuming and subjective processes. However, as the cause of AD is not completely understood, designing robust analysis methods for effective hand-crafted features using medical experts' knowledge is a challenging task. In contrast, deep learning is capable of automatically learning input features from a large set of training data. Many previous studies were conducted to further explore CNN architectures dedicated to generating robust AD features.

Gupta et. al. [79] used cross-domain features to represent MRI data. They deployed a stacked autoencoder (SAE) to learn a set of filters from natural image database and then applied a CNN to obtain a more effective feature representation for AD classification. Despite being very simple, they showed high classification performance in comparison with contemporary approaches. Liu et al. [80] also proposed an SAE-based multimodal neuroimaging feature learning algorithm from a region of interest (ROI) for AD diagnosis. This framework uses a zero-masking strategy for data fusion to extract complementary information from multiple data modalities.

Brosch et al. [81] learned a low-dimensional manifold of brain volumes with a deep belief networks (DBN) algorithm to detect the modes of variations that correlate to demographic and disease parameters for AD. Their primary contributions are following: 1) they introduced a much more computationally efficient training method for DBNs that allows training on 3D medical images with a resolution up to $128 \times 128 \times 128$, and 2) they demonstrated that DBNs can learn a low-dimensional manifold of brain volumes that can detect modes of variations.

Payan et al. [82] used a sparse auto-encoder to learn feature embedding and

then feed these embeddings to a convolution neural network for AD classification. The authors built a learning algorithm that is able to discriminate between healthy brains and diseased brains using sMRI images as input. They investigated a class of deep artificial neural networks and a specific combination of sparse auto-encoders and CNNs. The main novelty of their approach is to use 3D convolutions on the whole MRI image. Li et al. [83] proposed a robust multitask deep learning framework using a dropout [66] and stability selection technique to improve the ROI feature representation for AD/MCI diagnosis.

Shi et al. [15] developed a robust deep learning framework for multimodal AD diagnosis from sMRI and PET scans. They applied principal component analysis (PCA) to obtain features and then utilized a stability selection technique together with the LASSO method [14] to select the most effective features. The selected features were then feed to the deep learning structure. Unsupervised training was performed for initializing model weights in the deep structure and then fine-tuned by AD patient labels. During the fine-tuning phase, the dropout layer was deployed to enhance the model's generalization capacity. Finally, the feature representations was used for classifying AD/MCI by a support vector machine (SVM).

Visual inspection of neuro-imagery is not effective in identifying minor structural and metabolic changes as it is susceptible to the limitation of human eye and others related factors like subjectivity and experience and expertise of the clinicians. Automatic methods have been shown to be equally or more effective than clinicians in diagnosing AD from neuroimages [84]

Chincarini et al. [85] considered the textural and statistical features of the perirhinal cortex, hippocampus, entorhinal cortex and parahippocampal gyri. A random forest classifier was deployed for analyzing the features of each region to

extract the relevant ones, which were subsequently processed with supprot vector machine(SVM) for prediction of AD conversion. They have used ADNI database and their experiment showed 97% AD vs NC and 92% mAD vs NC classification accuracy.

Tang et al. [86] used shape diffeomorphometry of the left and right amygdala, thalamus, caudate, putamen, hippocampus, globus pallidus, and lateral ventricle for prediction of AD progression using LDA. The authors experimented on ADNI dataset. The accuracy was 74.77% for stable mild cognitive impairment(sMCI) vs progressive mild cognitive impairment(pMCI).

Gerardin et al., 2009 [87] presented a method to automatically classify between AD patients from mild cognitive impairment (MCI) and elderly controls(NC) subjects. The authors used spherical harmonics (SPHARM) coefficients to represent the shape of the hippocampi. SPHARM coefficients are then feeded to the SVM for classification. A bagging strategy is used to select the most relevant features for classification. They used Centre Hospitalo-Universitaire (CHU) of Caen dataset. Their method shows 94% AD vs NC classification accuracy. For MCI vs NC classification their method obtained 83% accuracy.

Li et al., 2007 [88] pinpointed hippocampal regional changes by surface-based anatomic mesh modeling method that match homologous hippocampal surface points between individuals. They have performed 100 times 3-fold CV and 100 times leave-one-out cross validation (LOOCV). They have found that the classification results demonstrated no distinct changes by using different patch sizes.

The key technique in [84] is the use of cross–domain features to represent MRI data. They deployed a sparse autoencoder for learning a set of filters from

Table 8: ***Comparison with state of the art methods***

Method	Subjects AD/mAD/aAD/NC	Dataset	Modality	ROI	Method	AD/NC	AD/mAD	AD/aAD	mAD/aAD	mAD/NC	aAD/NC
Wang et. al[100]	18/-/-/26	ADRC	sMRI	Hippocampus	LR	81.10	-	-	-	-	-
Li et. al[101]	19/-/-/20	internal	sMRI	Hippocampus	SVM	94.90	-	-	-	-	-
Gerardin et. al.	23/23/-/25	HCUDC	sMRI	Hippocampus	SVM	94.00				83	
shen et. al.[103]	99/-/-/138	ADNI	sMRI	Hippocampus	Bagged SVM	88.30					
Sorensen et al	101/233/-/169	ADNI	sMRI	Hippocampus	SVM	91.20				76.40	
Chincarni et al	144/302*/-/189	ADNI	sMRI	BSR	SVM	97				92	
Gupta et al	200/411/-/232		sMRI	WB voxel-level	Sparse AE + CNN	94.7	88.1			86.4	
Payan et al [82]	755/755/-/755		sMRI	WB Voxel level	Sparse AE and CNN	95.4	95.0			92.1	
Hosseini Asi et al	70/70/-/70		sMRI	WB Voxel level	AE and CNN	97.6	95.0			90.8	
chen et al	94/121/-/123			wb voxel level	SAE	89.0				81.7	
Liu et al	180/374/-/204		sMRI	WB Region level	SAE	82.6				72.0	
Liu et al	85/168/-/77		sMRI and PET	WB Region Level	SAE	91.4	82.1				
henrichs et al[200]	183			voxel features	LPBoost+SVM	.82					
salvatore	137/-/-/162			region feature	LR+ensembled SVM	0.76					
oikkalainen	115/-/-/88			region features	LR+ensemble SVM						
Proposed Approach	81/39/35/171	GARD, ADNI	sMRI	ROI	**Ensemble CNN**	90.38	87.50	82.61	80.00	**82.35**	**57.14**

natural images and then applied convolution operation to extract features from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Using this new representation, the authors classified MRI instances into three categories: AD, MCI and HC. By focusing on an expressive bassis set for putative biomarkers, they achieved comparatively high diagnostic accuracy.

In the multimodal stacked deep polynomial approach (MMSDPN) [15], as reported in the paper, the authors used MRI, PET and cerebrospinal fluid (CSF) data to achieve 97.13% accuracy with 4.44% variation. MCI vs NC 87.24% $\pm$ 4.52%

The authors in [83] used the dropout technique to prevent overfitting in their multitask learning approach. Their approach demonstrated 91.4% accuracy on MRI, PET and CSF data. Both of the approaches were studied on the ADNI data (51 AD patients and 52 NC patients from each of the mentioned modalities).

The authors in [80] added a weight decay to regularize the objective function as a way to prevent over-fitting. They experimented on 77 NC and 85 AD scans in both MRI and PET data from ADNI. Their model showed 82.59% accuracy with 5.33 variation in the sMRI modality, which is 91.40% $\pm$ 5.56% in the multimodal data. Payan et al. [82] studied the MRI modality with 755 scans for each of the classes. This approach also added a weight decay term similar to [80] to regularize the objective function in order to prevent over-fitting. The method demonstrated 95.39% accuracy in classifying AD versus NC. A feature selection method was deployed along with an l-norm penalty on weights to prevent over-fitting in [19]. The method achieved 82% accuracy with spatial augmentation on MRI and PET images. The reported accuracy without spatial augmentation on the data was 77%. This method was studied on 149 PET and 183 MRI images from ADNI. To avoid the over-fitting problem, Salvatore et al. [18] performed feature

extraction and feature selection tasks separately for training-validation data and testing data. They reported 76% test accuracy on ADNI sMRI. The number of samples for AD was 137 and 162 for NC.

The method in [79] obtained 93.80% accuracy on ADNI MRI data. They did not explicitly discuss overfitting. This method learned a set of bases from natural images by deploying SAE and then used these bases to learn MRI features. The number of scans was 200 for AD and 232 for NC. VoxCNN in [89] reported $79.0\% \pm 0.08\%$ accuracy. By using the residual neural network (ResNet), the approach [89] obtained $80.0\% \pm 0.07\%$ accuracy. Here, the labeled sMRI scans included 50 AD and 61 NC from ADNI. Here, the overfitting problem was addressed by pre-training.

In this paper, the proposed patch-based ensembles of simple models demonstrate significant performance. We used only small patches (32×32) from the hippocampus of the brain MRIs and achieved comparable accuracy. Our patch generation reduces the scarcity of training data for generalization. Using the ensemble technique also contributed to building a robust model while avoiding the over-fitting problem. It helps us to avoid obtaining an over-capacity network regarding the training time.

C. Conclusion

The proposed work provides an pair-wise classification accuracy of different stages of AD. We have observed that, selected predefined ROI (hippocampus)in sMRI has little or no information for aAD diagnosis. For the selected structure in the brain, we had to deploy six different patch-based classification models for each binary classification problem. We achieved good accuracy for

AD/NC, AD/mAD, AD/aAD, mAD/aAD and mAD/NC classification compare to other state-of-the-art methods. Our analysis on aAD is new to the machine learning/deep learning based AD research community.

VIII. CONCLUSION

Irreversible brain disorder AD is one of the major causes of death of elderly individuals. The world-wide effort for early diagnosis of this diseases is also significant. Designing efficient method would beefs up the diagnosis process. Deployment of deep learning methods would provide improved accuracy in early diagnosis of this disease. The state-of-the-art methods utilizes either multi-modal data or whole brain information in single modality for AD diagnosis.

Our approach uses single modality (i,e, sMRI) of data and deployed ROI-based CNN for efficient performance. We have achieved comparable performances in this regard(90.38%). We have also contributed in selecting region of interest which provides discernible features for AD and its pro-dromal stages. We have utilized atlas-based segmented data from GARD for this purpose. The hippocampus localization procedure with deep metric verification also provided state-of-the art results. This deep-metric aware robust ROI localization technique provides hope of avoiding time consuming processes for AD diagnosis.

PUBLICATIONS

A. Journals

1. S. Ahmed, K. Y. Choi, J. J. Lee, *et al.*, "Ensembles of patch-based classifiers for diagnosis of alzheimer diseases", *IEEE Access*, vol. 7, pp. 73 373–73 383, 2019. DOI: 10.1109/ACCESS.2019.2920011. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2920011.

B. Conferences

1. S. Ahmed, A. Basher, A. Reja, *et al.*, "A brief Review on Deep Metric Learning", in *Proceedings (number unknown)International Conference of Next Generation Computing*, 2018.

2. S. Ahmed and H. Y. Jung, "Alzheimer's disease and normal control MRI classification over hippocampus region", in *Proceedings 4th International Conference of Next Generation Computing*, 2018.

3. H. Y. J. Samsuddin Ahmed, "Learning hippocampus embeddings by siamese convolutional neural network", *1st Conference on the Smart Media Apllication*, no. 1, 2019.

C. Workshops

1. S. Ahmed and H. Y. Jung, "MRI Preprocessing for Alzheimer Diseases Prediction", in *Proceedings 8th Workshop on Convergent and Smart Media Systems*, 2018.

2. S. Ahmed and H. Y. Jung, “Alzheimer Diseases MRI Classification over right Hippocampus Features”, in *Proceedings 9th Workshop on Convergent and Smart Media Systems*, 2018.

References

[1] Y. Xian, Z. Akata, G. Sharma, Q. Nguyen, M. Hein, and B. Schiele, "Latent embeddings for zero-shot classification", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] S. Zhu and R. Zanibbi, "A text detection system for natural scenes with convolutional feature learning and cascaded classification", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[3] Y. Wang, J. Choi, V. Morariu, and L. S. Davis, "Mining discriminative triplets of patches for fine-grained classification", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] X. Chang, Y.-L. Yu, Y. Yang, and E. P. Xing, "They are not equally reliable: Semantic event search using differentiated concept classifiers", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[5] V. N. Murthy, V. Singh, T. Chen, R. Manmatha, and D. Comaniciu, "Deep decision network for multi-class image classification", in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[6] J. Burstein, C. Doran, and T. Solorio, Eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, 2019, ISBN: 978-1-950737-13-0. [Online]. Available: `https://www.aclweb.org/anthology/volumes/N19-1/`.

[7] C. Xu, W. Rao, X. Xiao, E. S. Chng, and H. Li, "Single channel speech separation with constrained utterance level permutation invariant training using grid LSTM", in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*, 2018, pp. 6–10. DOI: 10.1109/ICASSP.2018.8462471. [Online]. Available: https://doi.org/10.1109/ICASSP.2018.8462471.

[8] K. Tan, J. Chen, and D. Wang, "Gated residual networks with dilated convolutions for supervised speech separation", in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*, 2018, pp. 21–25. DOI: 10.1109/ICASSP.2018.8461819. [Online]. Available: https://doi.org/10.1109/ICASSP.2018.8461819.

[9] Z. Fan, Y. Lai, and J. R. Jang, "SVSGAN: singing voice separation via generative adversarial network", in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*, 2018, pp. 726–730. DOI: 10.1109/ICASSP.2018.8462091. [Online]. Available: https://doi.org/10.1109/ICASSP.2018.8462091.

[10] M. Chen, Y. Zhang, M. Qiu, N. Guizani, and Y. Hao, "SPHA: smart personal health advisor based on deep analytics", *IEEE Communications Magazine*, vol. 56, no. 3, pp. 164–169, 2018. DOI: 10.1109/MCOM.2018.1700274. [Online]. Available: https://doi.org/10.1109/MCOM.2018.1700274.

[11] M. J. Cardoso, A. Feragen, B. Glocker, E. Konukoglu, I. Oguz, G. B. Unal, and T. Vercauteren, Eds., *International Conference on Medical Imaging with Deep Learning, MIDL, 2019, 8-10 July 2019, London, United Kingdom*, vol. 102, Proceedings of Machine Learning Research, PMLR, 2019. [Online]. Available: http://proceedings.mlr.press/v102/.

[12] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539. [Online]. Available: https://doi.org/10.1038/nature14539.

[13] D. L. Weimer and M. A. Sager, "Early identification and treatment of alzheimers disease: Social and fiscal outcomes", *Alzheimers & Dementia*, vol. 5, no. 3, 215–226, 2009. DOI: 10.1016/j.jalz.2009.01.028.

[14] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of alzheimer's disease and mild cognitive impairment", *NeuroImage*, vol. 55, no. 3, pp. 856 –867, 2011, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2011.01.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811911000267.

[15] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of alzheimers disease", *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, 173–183, 2018. DOI: 10.1109/jbhi.2017.2655720.

[16] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li, "A robust deep model for improved classification of ad/mci patients", *IEEE Journal of*

Biomedical and Health Informatics, vol. 19, no. 5, 1610–1616, 2015. DOI: 10.1109/jbhi.2015.2429556.

[17] A. Gupta, M. S. Ayhan, and A. S. Maida, "Natural image bases to represent neuroimaging data", in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13, Atlanta, GA, USA: JMLR.org, 2013, pp. III–987–III–994. [Online]. Available: http://dl.acm.org/citation.cfm?id=3042817.3043047.

[18] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer's disease: A machine learning approach", *Frontiers in Neuroscience*, vol. 9, p. 307, 2015. DOI: 10.3389/fnins.2015.00307. [Online]. Available: https://10.3389/fnins.2015.00307.

[19] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, "Spatially augmented lpboosting for AD classification with evaluations on the ADNI dataset", *NeuroImage*, vol. 48, no. 1, pp. 138–149, 2009. DOI: 10.1016/j.neuroimage.2009.05.056. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2009.05.056.

[20] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification", in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*, 2005, pp. 539–546. DOI: 10.1109/CVPR.2005.202. [Online]. Available: https://doi.org/10.1109/CVPR.2005.202.

[21] *Alzheimer's disease neuroimaging initiative*. [Online]. Available: http://adni.loni.usc.edu/.

[22] *Oasis brains*. [Online]. Available: http://www.oasis-brains.org/.

[23] M. Reuter, *Freesurfer*. [Online]. Available: https://surfer.nmr.mgh.harvard.edu/.

[24] B. Kannappan, "A large-scale neuroimaging study on brain morphometry in normal aging and prediction of alzheimer's disease", PhD thesis, Department of Life Sciences , Graduate School of Chosun University, Gwangju, Republic of Korea, 2019.

[25] *Mango software*. [Online]. Available: http://ric.uthscsa.edu/mango/.

[26] J.-C. Baron, G. Chetelat, B. Desgranges, G. Perchey, B. Landeau, V de la Sayette, and F. Eustache, "In vivo mapping of gray matter loss with voxel-based morphometry in mild alzheimer's disease", *Neuroimage*, vol. 14, no. 2, pp. 298–309, 2001 ; DOI: 10.1006/nimg.2001.0848.

[27] G. B. Frisoni, C Testa, A Zorzan, F Sabattoli, A Beltramello, H Soininen, and M. P. Laakso, "Detection of grey matter loss in mild alzheimer's disease with voxel based morphometry", *Journal Neurology Neurosurgery & Psychiatry*, vol. 73, no. 6, 657–664, 2002. DOI: 10.1006/nimg.2001.0848.

[28] B. H. and E. Braak, "Neuropathological stageing of alzheimer-related changes", *Acta Neuropathologica*, 1991. DOI: 10.1007/BF00308809. [Online]. Available: https://doi.org/10.1007/BF00308809.

[29] L. E. Wisse, G. J. Biessels, S. M. Heringa, H. J. Kuijf, D. H. L. Koek, P. R. Luijten, and M. I. Geerlings, "Hippocampal subfield volumes at 7t in early alzheimer's disease and normal aging", *Neurobiology of Aging*, vol. 35, no. 9, pp. 2039–2045, 2002. DOI: 10.1016/j.neurobiolaging.2014.02.021.

[30] D. Devanand, R. Bansal, J. Liu, X. Hao, G. Pradhaban, and B. S. Peterson, "Mri hippocampal and entorhinal cortex mapping in predicting conversion to alzheimers disease", *NeuroImage*, vol. 60, no. 3, 1622–1629, 2012. DOI: 10.1016/j.neuroimage.2012.01.075.

[31] T. Stoub, E. Rogalski, S. Leurgans, D. Bennett, and L. Detoledo-Morrell, "Rate of entorhinal and hippocampal atrophy in incipient and mild ad: Relation to memory function", *Neurobiology of Aging*, vol. 31, no. 7, 1089–1098, 2010. DOI: 10.1016/j.neurobiolaging.2008.08.003.

[32] L. Detoledo-Morrell, T. R. Stoub, S. Leurgans, D. Fleischman, D. A. Bennett, and D. A. Turner, "P3-086 rate of entorhinal and hippocampal atrophy in incipient and very mild ad: A 6-year mri study", *Neurobiology of Aging*, vol. 25, 2004. DOI: 10.1016/s0197-4580(04)81238-7.

[33] B. H. Ridha, V. M. Anderson, J. Barnes, R. G. Boyes, S. L. Price, M. N. Rossor, J. L. Whitwell, L. Jenkins, R. S. Black, M. Grundman, and et al., "Volumetric mri and cognitive measures in alzheimer disease", *Journal of Neurology*, vol. 255, no. 4, 567–574, 2008. DOI: 10.1007/s00415-008-0750-9.

[34] S. J. Kile, W. G. Ellis, J. M. Olichney, S. Farias, and C. Decarli, "Alzheimer abnormalities of the amygdala with klüver-bucy syndrome

symptoms", *Archives of Neurology*, vol. 66, no. 1, 2009. DOI: 10.1001/archneurol.2008.517.

[35] L. P. D. Val, J. L. Cantero, and M. Atienza, "Atrophy of amygdala and abnormal memory-related alpha oscillations over posterior cingulate predict conversion to alzheimer's disease", *Scientific Reports*, vol. 6, no. 1, 2016. DOI: 10.1038/srep31859.

[36] K. H. H. S. H. Moon Y. Moon W. J, "Regional atrophy of the insular cortex is associated with neuropsychiatric symptoms in alzheimer's disease patients", *European Neurology*, vol. 71, pp. 223–229, 7 2014. DOI: 10.1159/000356343. [Online]. Available: https://doi.org/10.1159/000356343.

[37] A. L. Foundas, C. M. Leonard, S. M. Mahoney, O. F. Agee, and K. M. Heilman, "Atrophy of the hippocampus, parietal cortex, and insula in alzheimer's disease: A volumetric magnetic resonance imaging study.", *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*, vol. 10, no. 2, pp. 81–89, 1997.

[38] A. L.Foundas, K. F.Eure, and B. Seltzer, "Conventional mri volumetric measures of parietal and insular cortex in alzheimer's disease", *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 20, pp. 1131–1144, 7 1996. DOI: 10.1016/S0278-5846(96)00101-7. [Online]. Available: https://doi.org/10.1016/S0278-5846(96)00101-7.

[39] P. T. Reiss, M. H. H. Stevens, Z. Shehzad, E. Petkova, and M. P. Milham, "On Distance-Based Permutation Tests for Between-Group Comparisons ", *Journal of the International Biometric Society*, vol. 66, no. 2, pp. 636–643, 2010. DOI: 10.1111/j.1541-0420.2009.01300.x. eprint: https:

//onlinelibrary.wiley.com/doi/full/10.1111/j.1541-0420.2009.01300.x. [Online]. Available: https://doi.org/10.1111/j.1541-0420.2009.01300.x.

[40] K. S. Taber, "The use of cronbach's alpha when developing and reporting research instruments in science education", *Research in Science Education*, vol. 48, no. 6, pp. 1273–1296, 2018, ISSN: 1573-1898. DOI: 10.1007/s11165-016-9602-2. [Online]. Available: https://doi.org/10.1007/s11165-016-9602-2.

[41] M.-R. Siadat, H. Soltanian-Zadeh, and K. V. Elisevich, "Knowledge-based localization of hippocampus in human brain mri", *Computers in Biology and Medicine*, vol. 37, no. 9, 1342–1360, 2007. DOI: 10.1016/j.compbiomed.2006.12.010.

[42] A. M. Kälin, M. T. M. Park, M. M. Chakravarty, J. P. Lerch, L. Michels, C. Schroeder, S. D. Broicher, S. Kollias, R. M. Nitsch, A. F. Gietl, and et al., "Subcortical shape changes, hippocampal atrophy and cortical thinning in future alzheimers disease patients", *Frontiers in Aging Neuroscience*, vol. 9, 2017. DOI: 10.3389/fnagi.2017.00038.

[43] J. Milos and P. Mihovil, "A note on the sea-horse in the human brain", *Translational Neuroscience*, vol. 1, no. 4, pp. 335–337, 2010, ISSN: 2081-6936. DOI: 10.2478/v10134-010-0041-8. [Online]. Available: https://doi.org/10.2478/v10134-010-0041-8.

[44] P. Moutafis, M. Leng, and I. A. Kakadiaris, "An overview and empirical comparison of distance metric learning methods", *IEEE Transactions on Cybernetics*, vol. 47, no. 3, 612–625, 2017. DOI: 10.1109/tcyb.2016.2521767.

[45] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering", in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 815–823. DOI: 10.1109/CVPR.2015.7298682. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7298682.

[46] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification", in *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, 2014, pp. 1701–1708. DOI: 10.1109/CVPR.2014.220. [Online]. Available: https://doi.org/10.1109/CVPR.2014.220.

[47] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: attention-based convolutional neural network for modeling sentence pairs", *TACL*, vol. 4, pp. 259–272, 2016. [Online]. Available: https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/831.

[48] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a "siamese" time delay neural network", in *Proceedings of the 6th International Conference on Neural Information Processing Systems*, ser. NIPS'93, Denver, Colorado: Morgan Kaufmann Publishers Inc., 1993, pp. 737–744. [Online]. Available: http://dl.acm.org/citation.cfm?id=2987189.2987282.

[49] Y. Bengio and Y. LeCun, Eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*, 2015. [Online]. Available: https://iclr.

cc/archive/www/doku.php\%3Fid=iclr2015:accepted-main.html.

[50] J. Han and M. Kamber, *Data mining: concepts and techniques*, 3rd ed. Elsevier, 2012.

[51] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information", in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, ser. NIPS'02, Cambridge, MA, USA: MIT Press, 2002, pp. 521–528. [Online]. Available: http://dl.acm.org/citation.cfm?id=2968618.2968683.

[52] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review", *Neurocomputing*, vol. 187, pp. 27–48, 2016. DOI: 10.1016/j.neucom.2015.09.116. [Online]. Available: https://doi.org/10.1016/j.neucom.2015.09.116.

[53] *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, IEEE Computer Society, 2006, ISBN: 0-7695-2597-0. [Online]. Available: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=10924.

[54] A. Basher, K. Y. Choi, J. J. Lee, B. Lee, B. C. Kim, K. H. Lee, and H. Y. Jung, "Hippocampus localization using a two-stage ensemble hough convolutional neural network", *IEEE Access*, vol. 7, pp. 73 436–73 447, 2019. DOI: 10.1109/ACCESS.2019.2920005.

[55] H. Fan and E. Zhou, "Approaching human level facial landmark localization by deep learning", *Image Vision Comput.*, vol. 47, pp. 27–35, 2016. DOI: 10.1016/j.imavis.2015.11.004. [Online]. Available: https://doi.org/10.1016/j.imavis.2015.11.004.

[56] L. Deng and D. Yu, "Deep learning: Methods and applications", *Foundations and Trends in Signal Processing*, vol. 7, no. 3-4, pp. 197–387, 2014. DOI: 10.1561/2000000039. [Online]. Available: https://doi.org/10.1561/2000000039.

[57] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. M. Baltruschat, R. Werner, and A. Schlaefer, "Skin lesion classification using cnns with patch-based attention and diagnosis-guided loss weighting", *CoRR*, vol. abs/1905.02793, 2019. arXiv: 1905.02793. [Online]. Available: http://arxiv.org/abs/1905.02793.

[59] K. Roy, D. Banik, D. Bhattacharjee, and M. Nasipuri, "Patch-based system for classification of breast histology images using deep learning", *Comp. Med. Imag. and Graph.*, vol. 71, pp. 90–103, 2019. DOI: 10.1016/j.compmedimag.2018.11.003. [Online]. Available: https://doi.org/10.1016/j.compmedimag.2018.11.003.

[60] A. Prochazka, S. Gulati, S. Holinka, and D. Smutek, "Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition", *Comp. Med. Imag. and Graph.*, vol. 71, pp. 9–18, 2019. DOI: 10.1016/j.compmedimag.2018.10.001. [Online]. Available: https://doi.org/10.1016/j.compmedimag.2018.10.001.

[61] P. Poudel, A. Illanes, M. Sadeghi, and M. Friebe, "Patch based texture classification of thyroid ultrasound images using convolutional neural network", in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2019, Berlin, Germany, July 23-27, 2019*, 2019, pp. 5828–5831. DOI: 10.1109/EMBC.2019.8857929. [Online]. Available: https://doi.org/10.1109/EMBC.2019.8857929.

[62] H. Yuan, "Robust patch-based sparse representation for hyperspectral image classification", *IJWMIP*, vol. 15, no. 3, pp. 1–15, 2017. DOI: 10.1142/S021969131750028X. [Online]. Available: https://doi.org/10.1142/S021969131750028X.

[63] K. Hara, D. Saitoh, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning", in *2015 International Joint Conference on Neural Networks, IJCNN 2015, Killarney, Ireland, July 12-17, 2015*, 2015, pp. 1–8. DOI: 10.1109/IJCNN.2015.7280578. [Online]. Available: https://doi.org/10.1109/IJCNN.2015.7280578.

[64] S. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, and H. Cheng, "Classification of alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling", *J. Medical Systems*, vol. 42, no. 5, 85:1–85:11, 2018. DOI: 10.1007/s10916-018-0932-7. [Online]. Available: https://doi.org/10.1007/s10916-018-0932-7.

[65] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France,*

6-11 July 2015, 2015, pp. 448–456. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html.

[66] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?id=2670313.

[67] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks", in *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, 2016, pp. 507–516. [Online]. Available: http://proceedings.mlr.press/v48/liud16.html.

[68] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010*, 2010, pp. 249–256. [Online]. Available: http://proceedings.mlr.press/v9/glorot10a.html.

[69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980.

[70] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri", *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, pp. 1–1, 2018, ISSN: 0162-8828. DOI: 10.1109/TPAMI.2018.2889096.

[71] S. L. Goldenberg, G. Nir, and S. E. Salcudean, "A new era: Artificial intelligence and machine learning in prostate cancer", *Nature Reviews Urology*, vol. 16, no. 7, 391–403, 2019. DOI: 10.1038/s41585-019-0193-3.

[72] D. Shen, C. Wee, D. Zhang, L. Zhou, and P. Yap, "Machine learning techniques for AD/MCI diagnosis and prognosis", pp. 147–179, 2014. DOI: 10.1007/978-3-642-40017-9_8. [Online]. Available: https://doi.org/10.1007/978-3-642-40017-9_8.

[73] Y. Wang, M. Liu, L. Guo, and D. Shen, "Kernel-based multi-task joint sparse classification for alzheimers disease", *2013 IEEE 10th International Symposium on Biomedical Imaging*, 2013. DOI: 10.1109/isbi.2013.6556786.

[74] S.-T. Yang, J.-D. Lee, T.-C. Chang, C.-H. Huang, J.-J. Wang, W.-C. Hsu, H.-L. Chan, Y.-Y. Wai, and K.-Y. Li, "Discrimination between alzheimers disease and mild cognitive impairment using som and pso-svm", *Computational and Mathematical Methods in Medicine*, vol. 2013, 1–10, 2013. DOI: 10.1155/2013/253670.

[75] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of alzheimers disease", *NeuroImage*, vol. 65, 167–175, 2013. DOI: 10.1016/j.neuroimage.2012.09.065.

[76] A. Ortiz, J. M. Górriz, J. Ramírez, and F. Martínez-Murcia, "Lvq-svm based cad tool applied to structural mri for the diagnosis of the alzheimer's disease", *Pattern Recognition Letters*, vol. 34, no. 14, 1725–1733, 2013. DOI: 10.1016/j.patrec.2013.04.014.

[77] J. Escudero, J. P. Zajicek, and E. Ifeachor, "Machine learning classification of mri features of alzheimers disease and mild cognitive impairment subjects to reduce the sample size in clinical trials", *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011. DOI: 10.1109/iembs.2011.6091962.

[78] R. K. Lama, J. Gwak, J.-S. Park, and S.-W. Lee, "Diagnosis of alzheimer's disease based on structural mri images using a regularized extreme learning machine and pca features", *Journal of Healthcare Engineering*, vol. 2017, 1–11, 2017. DOI: 10.1155/2017/5485080.

[79] A. Gupta, M. S. Ayhan, and A. S. Maida, "Natural image bases to represent neuroimaging data", in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13, Atlanta, GA, USA: JMLR.org, 2013, pp. III–987–III–994. [Online]. Available: http://dl.acm.org/citation.cfm?id=3042817.3043047.

[80] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham, and ADNI, "Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease", *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1132–1140, 2015, ISSN: 0018-9294. DOI: 10.1109/TBME.2014.2372011.

[81] T. Brosch and R. Tam, "Manifold learning of brain mris by deep learning", *Advanced Information Systems Engineering Lecture Notes in Computer Science*, 633–640, 2013. DOI: 10.1007/978-3-642-40763-5_78.

[82] A. Payan and G. Montana, "Predicting alzheimer's disease - A neuroimaging study with 3d convolutional neural networks", in *ICPRAM 2015 - Proceedings of the International Conference on Pattern Recognition Applications and Methods, Volume 2, Lisbon, Portugal, 10-12 January, 2015.*, 2015, pp. 355–362.

[83] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, and J. Li, "A robust deep model for improved classification of ad/mci patients", *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, 1610–1616, 2015. DOI: 10.1109/jbhi.2015.2429556.

[84] A. Gupta, M. Ayhan, and A. Maida, "Natural image bases to represent neuroimaging data", in *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, ser. JMLR Workshop and Conference Proceedings, vol. 28, 2013, pp. 987–994. [Online]. Available: http://jmlr.org/proceedings/papers/v28/gupta13b.html.

[85] A. Chincarini, P. Bosco, P. Calvini, G. Gemme, M. Esposito, C. Olivieri, L. Rei, S. Squarcia, G. Rodriguez, R. Bellotti, P. Cerello, I. D. Mitri, A. Retico, and F. Nobili, "Local MRI analysis approach in the diagnosis of early and prodromal alzheimer's disease", *NeuroImage*, vol. 58, no. 2, pp. 469–480, 2011. DOI: 10.1016/j.neuroimage.2011.05.083. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2011.05.083.

[86] X. Tang, D. Holland, A. M. Dale, L. Younes, and M. I. Miller, "Baseline shape diffeomorphometry patterns of subcortical and ventricular structures in predicting conversion of mild cognitive impairment to alzheimers disease", *Journal of Alzheimers Disease*, vol. 44, no. 2, 599–611, 2015. DOI: 10.3233/jad-141605.

[87] E. Gerardin, G. Chetelat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehericy, L. Garnero, and et al., "Multidimensional classification of hippocampal shape features discriminates alzheimers disease and mild cognitive impairment from normal aging", *NeuroImage*, vol. 47, no. 4, 1476–1486, 2009. DOI: 10.1016/j.neuroimage.2009.05.036.

[88] S. Li, F. Shi, F. Pu, X. Li, T. Jiang, S. Xie, and Y. Wang, "Hippocampal shape analysis of alzheimer disease based on machine learning methods", *American Journal of Neuroradiology*, vol. 28, no. 7, 1339–1345, 2007. DOI: 10.3174/ajnr.a0620.

[89] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain MRI classification", in *14th IEEE International Symposium on Biomedical Imaging, ISBI 2017, Melbourne, Australia, April 18-21, 2017*, 2017, pp. 835–838. DOI: 10.1109/ISBI.2017.7950647. [Online]. Available: https://doi.org/10.1109/ISBI.2017.7950647.

ACKNOWLEDGEMENTS

My heart-felt gratitude to my honorable supervisor Professor Ho Yub Jung. His intensive care and hands on instructions for reading a journal paper and writing a report and then refining the ideas by his unique method of vibrating disciple's neurons made me capable of accomplishing this seem-to-impossible work. From the very beginning of my study here, his continuous support at odds and evens help me focus on the objectives of the study. I would like to thank NRCD and ADNI for making the databases available. I would also like to show my gratitude to my course teachers for their awesome deliveries at lecture hall, challenging assignments and projects which help me learning state-of-the-art techniques of deep learning and medical imaging. I am grateful to my beloved kid, whom I could not touch yet, my respected parents & elder sister, affectionate siblings and my better half for their patience and prayers. Finally, I would like to thank Chosun family, specifically, officials from International Office and Graduate School Office for all kinds of their support during this endeavor.