February 2020

Master's Degree Thesis

Automatic Localization of Brain ROI from MRI using a Two-Stage Ensemble Hough Convolutional Neural Network

Graduate School of Chosun University

Department of Computer Engineering

Abol Basher

Automatic Localization of Brain ROI from MRI using a Two-Stage Ensemble Hough Convolutional Neural Network

2 단계 앙상블 허프 컨벌루셔널 뉴럴 네트워크를 사용한 뇌 MRI ROI 위치 예측 연구

February 25, 2020

Graduate School of Chosun University

Department of Computer Engineering

Abol Basher

Automatic Localization of Brain ROI from MRI using a Two-Stage Ensemble Hough Convolutional Neural Network

Advisor: Prof. Jung, Ho Yub, Ph.D.

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

Abol Basher

아볼 바셔 석사학위논문을 인준함

위원장 조선대학교 교수 최우열 (인)

위 원 조선대학교 교수 신석주 (인)

위 원 조선대학교 교수 정호엽 (인)

2019년 11월

조선대학교 대학원

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

CNN	Convolutional Neural Network
GH-CNN	Global Hough Convolutional Neural Network
LH-CNN	Local Hough Convolutional Neural Network
Hough-CNN	Hough Convolutional Neural Netwrok
DV	Displacement vector
GDV	Global Displacement Vector
LDV	Local Displacement Vector
GSP	Global Sample Position
LSP	Local Sample Position
GTVP	Global Target Voxel Position
LTVP	Local Target Voxel Position
TGDV	Target Global Displacement Vector
TLDV	Target Local Displacement Vector
PGDV	Predicted Global Displacement Vector
PLDV	Predicted Local Displacement Vector
MRI	Magnetic Resonance Imaging
TIVP	Target Image Voxel Position
ADNI	Alzheimer's Disease Neuroimaging Initiative
GARD	Gwangju Alzheimer's and Related Dementia
ROI	Region of Interest
MSE	Mean Square Error
RMS	root mean square

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Automatic Localization of Brain ROI from MRI using a Two-Stage Ensemble Hough Convolutional Neural Network

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

Artificial Intelligence is contributing in various ways to resolve complicated problems now a days. Artificial neural network has already shown a lot of break through in different disciplines, such as, medical imaging, computer vision, share market. Deep learning based algorithms are using to address various complex problems in medical imaging. Automatic localization of brain regions of interest (ROIs) from magnetic resonance imaging (MRI) scan is a crucially important task to diagnose the various neuro-degenerative diseases, such as, Alzheimer's disease. In this study, a method has been proposed to automatically locate the brain ROI, such as, hippocampus from MRI scan using a two-stage Hough convolutional neural Network (Hough-CNN). The proposed approach is a amalgamation of Hough voting and the deep convolutional neural network (CNN) and it locates the hippocampus in two Phase. In the first phase, the patches are extracted from the whole MRI scan to train the global Hough-CNN except the boundary region. In the second phase, the local patches are generated in the vicinity of the hippocampus, and then the local models are trained using those patches. In the test phase, the extracted patches from the whole hippocampal

region are used to predict the global position of hippocampus. After that, using the global positions of the hippocampus, the local patches are extracted in the vicinity of the target voxel. Using the local patches, the local Hough-CNN estimates the exact location of the hippocampus in the MRI scan. The proposed method is verified using Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Gwangju Alzheimer's and Related Dementia (GARD) cohort datasets.

한 글 요 약

2 단계 앙상블 허프 컨벌루셔널 뉴럴 네트워크를 사용한 뇌 MRI ROI위치 예측 연구

아볼 바셔
지도 교수: 정호엽
컴퓨터공학과
대학원, 조선대학교

인공 지능은 복잡한 문제 해결에 다양한 방식으로 기여하고 있습니다. 인공 신경 네트워크는 이미 의료 이미징, 컴퓨터 비전과 같은 다양한 분야에서 많은 돌파구를 보여주었습니다. Magnetic resonance image (MRI) 스캔으로부터 관심있는 뇌 region of interest (ROI)의 자동 위치 예측은 알츠하이머 병과 같은 다양한 신경 퇴행성 질환을 진단하기 위한 기본 작업 중 하나입니다. 이 연구에서는 2 단계 Hough Convolutional Neural Network (Hough-CNN)를 사용하여 MRI 스캔의 해마와 같은 뇌 ROI를 자동으로 찾는 방법이 제안됩니다. 제안 된 접근법은 Hough 투표와 CNN (deep convolutional neural network)의 융합으로 2 단계 Hough-CNN으로 해마를 찾습니다. 첫 번째 단계에서는 전역 Hough-CNN을 훈련시키기 위해 전체 MRI 스캔에서 three-view-patch (TVP)가 추출됩니다. 두 번째 단계에서는 해마 근처에서 로컬 TVP를 생성한 다음 해당 패치를 사용하여 로컬 모델이 학습합니다. 테스트 단계에서, 전체 해마 영역에서 추출 된 TVP는 해마의 전체 위치를 예측하는 데 사용되며, 해마의 전역 위치를 사용하여, 그 주변 로컬 TVP를 추출합니다. 로컬 TVP를 사용하여

로컬 Hough-CNN은 MRI 스캔에서 해마의 정확한 위치를 다시 추정합니다. 제안 된 방법은 ADNI (Alzheimer ’s Disease Neuroimaging Initiative) 및 Gwangju Alzheimer ’s and Related Dementia (GARD) 코호트 데이터 세트에서 state-of-the-arts 결과를 보여주고 있습니다.

I. INTRODUCTION

Artificial intelligence is being used to address complicated issues in different fields in different way. Deep learning based methods are offering various solutions for many complicated problems, such as, face recognition, object detection, malware detectin, finger print recognition, hand writing recognition, speech recognition, disease detection like alzheimer's, tumour, and cancer cell from multi-modal scanning of medical imaging and in so many other fields. This study mainly focused on medical imaging. The neuro-anatomical structures of human brain are highly complicated in nature and have many regions where the neighbouring structure overlapped each other.

On medical imaging, a lot of researcher are carrying out several studies on different branch of medical issues in recent years. One of the crucial studied field of medical imaging is brain magnetic resonance imaging (MRI)[1]–[4]. Different neurological disorders, such as, Alzheimer's disease, Schizophrenia, Epilepsy etc., are diagnosed using MRI. To perform such diagnosis, the prior knowledge of the particular organ in an MRI is very important. From 3-D MRI space, automatic localization of region of interest (ROI) is very complicated task. The hippocampus is one of the many ROIs, that has received a very high interest by the researcher communities and they are trying to analyze it's shape, size and structural changes along with it's atrophy. It is one of the complex ROIs of the human brain limbic system[2]. The neuro-analyst belived that the hippocampus plays a significant role in the learning process and the memory management of our daily life activities[2]. In addition, the hippocampal shape, structure, and size are the prime biomarker for the Alzheimer's disease detection [1], [2], [5], [6].

Several research communities are working in different ways to perform

accurate localization and segmentation of hippocampus with it's sub-fields automatically [1]–[3], [7]. Several methods[1]–[3], [7], [8] were proposed to localize and segment the hippocampus and it's sub-fields, yet it is a challenging research field due to the anatomical structure of the left and right hippocampi in both hemispheres. In adult brain, the approximate volume of hippocampus is 3.0 to 3.5 cm^3 on each side of the brain [9]. On the other hand, the faint edges and the overlapping intensities with the neighboring structures makes the hippocampal segmentation highly difficult[1]. Because of all this complexities, it will be very helpful for the segmentation process if we have the prior knowledge of the hippocampus in advance. Furthermore, it will provide an confined area of ROI and possibly the segmentation seeding points as well.

In this study, we proposed a method by aggregating the Hough voting [10]–[12] strategy with deep convolutional neural network (CNN)[13], [14]. Deep CNN is known to be a very good feature extractor from the input data. Hough voting and the deep CNN work together to determine the displacement vectors from the random sample points to the target location from the given volumetric MRI scan.

The proposed approach consisted of a two-stage ensemble learning similar to [15] and [16]. Using this method, it is possible to exactly estimate the location of the brain ROIs, such as, hippocampus. The proposed ensemble Hough-CNN consists of multiple global and Local models. The models were amalgamated to form ensemble model in both phase. Then, the amalgamated global and local models were aggregated together to form the final two-stage ensemble Hough-CNN. The global models were trained to learn the global feature maps from the whole MRI scan. The global information of an MRI facilitates the global models to learn the overall structural complexities of multiple organs of the human brain.

The features extracted from the different regions of MRI scan offers an inter-organ variability studies for the global model which help the global model to localized the global position of the target ROI. From the vicinity of the target ROI, multiple patches were extracted to train the local models. The local models learn the locals features of the MRI scan. The differentiable objective looks towards the overlapping organ and the distinguishable features collection from the local area of the target ROI will facilitate the local model to exactly localize the target ROI. In the test phase, the multiple trained global models were ensemble together to predict the displacement vectors. The resultant averaged displacement vectors were added with the global patches centres to have the predicted ROI location in the MRI scan. This predicted location of ROI is considered as the the global position of the ROI in the MRI scan. After that, the predicted global locations of ROI were used to generate the patches for the local models. The local models predicted the displacement vectors from the extracted local patches. The predicted local displacement vectors were added with the global position of the ROI to estimate the exact location of the ROI, which is expected to be inside the ROI. This simple two-stage localization strategy offers us a good model which can localize any ROI, such as, the left and right hippocampi from both hemishpere of brain MRI scan.

A. Contributions

- A two-stage ensemble based method has been proposed to localize the ROIs from MRI scan.
- We design and evaluate different Hough-CNN (GH-CNN and LH-CNN) architecture with different number of layers and filters in each layers. This approach offers an insight view on how different architecture cope with

different variabilities exist in medical MRI scans

- We demonstrate an efficient method to train the deep learning model using small dataset. Each type of network is trained using different amount of data to see the impact in the final localization. We have developed an approach to generate different amount of data from limited amount of MRI scans

- We adopted Keras sequential framework to implement a two-stage ensembled Hough-CNN. It offers automatic data management which allows us to consider as much data as it required to train, validate and test.

B. The Research Objectives

Although several methods have been proposed already, because of the nature of challenging hippocampal complex structure a method of fully automatic localization of hippocampus was not there. Therefore, a fully automatic deep learning based method has been proposed using Hough voting and the deep CNN in this thesis to localize the left and right hippocampi from MRI scan. Our goal is, therefore, to reduce the amount of time required to localize the brain ROIs, such as hippocampus, as well as provide simplified and computationally efficient model which can accurately estimate the left and right hippocampi from an MRI scan.

C. Thesis Layout

The thesis is organized as follows. In Chapter II, we present an overview of the different types of algorithms contributing in medical imaging and other fields.

Then in chapter III, we describe the proposed methodology and the dataset used to validate the proposed two-stage Hough-CNN. Next in Chapter IV, we discuss about the network architecture, loss function employed to train the proposed two-stage ensemble Hough-CNN and representative training procedures. We describe the localization procedures, error calculation and the statistical significant of the proposed method against the manual localization with a quantitative result analysis in Chapter V. We summarize our proposed approach in Chapter VI.

II. Related Works

In this section, we review the different methodologies related to medical imaging and some of the non-medical research mostly related to localization, segmentation, action recognition and object recognition from images and video data. Many methods have been proposed to localize object from images of different modality and disciplines. Brain structural ROIs localization have received a high interest by the research communities. The localization methods can be categorized into many types,such as, manual localization, spatial relation-based localization, atlas and statistical atlas-based localization, statistical shape model-based approaches as well as the artificial intelligence i.e., machine learning and deep learning based methods. Although manual localization considered as the gold standard in medical field, yet the researcher are thriving to build a fully automatic localization methodology. The manual localization of brain structures from MRI scan largely depends on expert radiologist, which can require a very large amount of time [17], [18]. On the contrary, the spatial relation-based method utilized a set of rules that were predefined such as fuzzy sets [19], [20]. The atlas and statistical atlas-based method offer automatic estimation of ROI from MRI scan, but it performs the localization by exploiting an atlas to extrapolate information to the target image using co-registration [21], [22]. In addition, these types of methods can accurately estimate the target ROI's position automatically, however, they usually require a large amount of computation time. To overcome the complexities of atlas-based system, other methods have been proposed, where statistical shape based model is one of the important candidate. The statistical shape model is capable of estimating shape variability from the training population and discern all the shapes existed

in the training data and extract the mean approximate shape [23] . Using the extracted mean shape, the statistical shape model-based approaches localize the approximate position of the ROI.

A pairwise nonrigid coherent point drift registration-based method was proposed in [1] to estimate the hippocampal position. Manually delineated 40 MRI scans were used to validate the proposed approach. They used the strength of assembly based coherent point drift registration and estimated the root means square distance from the random location to the target hippocampus. They have reported the root mean square error of 3.5 mm from the predicted location to the ground truth location.

On the other hand, a knowledge-based localization strategy of brain ROI was reported in [2]. This approach were build based on statistical road map where their intention was to localize the landmark in the brain MRI. They have also localized the hippocampus as representative ROI from the MRI scan using their proposed knowledge-based approach.

In [24], the authors have proposed an automatic segmentation method based on statistical parameter mapping to automate total intra-cranial volume (TIV) measurement from MRI scans. The intra-cranial volume is measured to correlate the hippocampal volume atrophy which is known to have a connection with several neurodegenerative disorders.

An in vivo atlas-based approach has been reported in [25] to segment the hippocampal sub-fields from MRI scan. Hippocampus is known to have one of the most complex structure in human brain. The sub-fields of the hippocampus is highly complicated in nature. J.L. Winterburn et al. proposed an in vivo atlas approach which can segment the hippocampal subfield. This approach has been implemented in Freesurfer 5.3.0.

On the contrary, an Bayesian model of shape and appearance based approach were proposed by [26] to segment the sub-cortical brain. This method has been packaged in FSL First software package.

An neural network based automated epileptic seizures detection were proposed using multi-layer perception in [27].

In [28], 3-D fully convolutional networks based approach has been proposed to segment the sub-cortical regions from brain MRI. 3-D convolution is not computational efficient and it requires very long time to train the network. However, J.Dolz et al. have come up with an idea to confine the limitation of 3-D CNN and segment the sob-cortical organs from MRI scans.

To detect arbitrary shape from image, a method was proposed in [29]. The author used the Hough transform algorithm to extract the features from the input image to identify the shapes.

In [30], A.Yao et al. proposed a method based on Hough transform voting framework to classify and localize human action in videos. A discriminative multi-class codebook were formed from the leaves of the decision trees, where the codebook consist with the action class along with vote for action centres. They used the low-level features such as optical flow and gradients to achieve the state-of-the-art result on several datasets.

J. Gall et al. [10] proposed Hough forest, a method, by improving the general Hough transform and amalgamating with the other models, such as Implicit Shape Model, to detect an object from highly unconstrained images and video frames. A random decision trees extracted from input data are used to form the Hough forest, where the local appearances of images are utilized to construct the leaves of the decision trees. Each leaf of the decision trees are marked as a probabilistic vote in the Hough space, where a set of leaves behaves like an

implicit appearance of codebook. The implicit appearance of codebook were improved to construct the Hough forest and detect the object from images and video frames.

A method were proposed combining generalized Hough transform with a discriminative training in [31] to localize object in medical imaging. H. Ruppertshofen et al. trained their model discriminatingly focusing on common features of the target object in the training images considering a minimal localization error to obtain efficient and robust localization results. They tested their method to localize knee from long-leg radio-graphs, the vertebrae in C-arm CT and the femoral head in whole-body MR.

A multiple objects instance detection methodology were developed by O. Barinova in [32]. In this method, one the one hand, they used the wide applicability and simplicity of Hough transform. On the other hand, they overcome the multiple peak identification problem in Hough images as well as detect the mltiple objects instances without invoking the non-maximum suppression. Using this method, they improve the detection accuracy by a large margin for straight line detection, and the pedestrian detection.

M. Godec et al.[33] extended the concept of Hough Forests to another domain and aggregated the Hough voting-based detection and back-projection method with a rough GrabCut segmentation to track the non-rigid object. They have overcome the limitations of a fixed bounding box and tracked the objects using generalized Hough transform. Under several predefined conditions, such as heavy non-rigid transformation, rotations, changing scales, and the partial occlusions, the proposed method showed robust and accurate tracking result.

On the other hand, A. Tran et al.[34] proposed a method to improve the efficiency and reliability of the Hough forest. Firstly, instead of using the

random patches from the training set, they selected a proportion of patches most relevant to the image content from a geometrical criterion. Secondly,instead of considering the uniformly sampling the parameter space, they utilized a probability map to construct the sample set. Using their proposed approach, training time were reduced drastically without affecting the accuracy and confined the variability of the detectors in a certain range. The proposed method were validated in the context of detecting the car and pedestrian.

Hough forest algorithm works effectively in object detection tasks. To extend the potential of Hough forest in pattern search, C. Henderson et al.[35] proposed a method that was capable of searching the database index and detecting the pattern in a real time. They implemented and detected the pattern from complex and crowded street-scenes in videos which was a challenging task.

Object detection in autonomous video surveillannce systems, Hough forest based models are widely used now a days. Hough forest uses the local patches that vote for the center of the object in image. However, since the voting process is occurred independently by the patches, therefore, it is uncertain that the trees constructed in Hough forests can have the optimal parameters for the entire model. To overcome this issue, a model has been proposed in [36]. The authors introduced weights to each displacement vectors in the training images and the weights were updated during the training process to minimize the global loss function. To accurately localize the instances in the image, the weights were utilized in both training and detection process. The authors validated their method on TUD pedestrian and UIUC car datasets.

To analyze interaction between human and advanced computer system with human behaviour, head pose and facial feature localization can play a vital rule. A method was proposed in [37] based on patches extracted from the image

considering a regular grid. The authors performed a conjugated operation of classification and regression on each patch. The foreground and background were classified in classification phase, whereas, the casting of vote in the Hough space were estimated in the regression phase. Instead of using the random forest, the authors utilized the CNN to extract the features from the input images.

In [38], a method of Latent-Class Hough forests were proposed for 6 DoF pose estimation and object detection from cluttered and occluded images. They amalgamated a scale-invariant patch descriptor with the regression forest considering a template-based split function. The proposed approach used only the positive samples for training, and the leave nodes were considered as the latent variables. The authors validated their method using highly challenging dataset of multiple instance holding 2-D and 3-D clutter as well as background occlusions.

The most relevant study related to our approach has been proposed by F. Milletari et al [39]. They proposed a Hough-CNN based segmentation approach for multi-modal, multi-regional medical images. In [39], they used Hough voting strategy, which allowed them to automatically localize and segment the 26 ROIs of the basal ganglia and the mid-brain from the MRI and transcranial Ultra-sound volumes. To perform the localization and segmentation, CNN architectures have been constructed to see the insights of the network, and how they cope with data variability exist in medical images along with different modalities. Constructed CNN models had different numbers of layers and convolutional kernels in each layer.

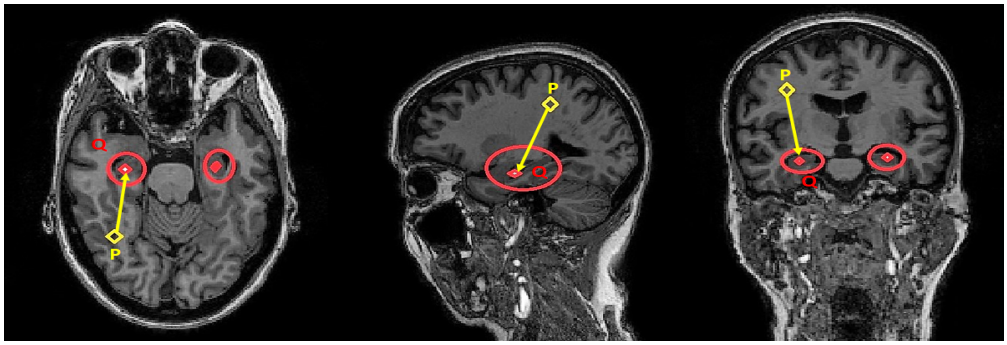
Hough votes consist of a collection of evidences where each evidence refers to a product set of various locations, aspects, and scales. The largest peak of all Hough votes directs towards an instance of an objects. The displacement vectors were estimated by the CNN models and directed towards the ROI based on the

estimated instances in the Hough space domain. In this thesis, we focused on automatic localization of brain ROIs, such as left and right hippocampi, using two-stage ensemble Hough-CNN and we quantitatively evaluated the root mean square error against manually delineated ground truth.

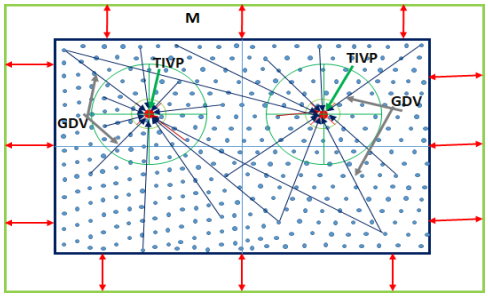
III. Methodology and Datasets

A. Method

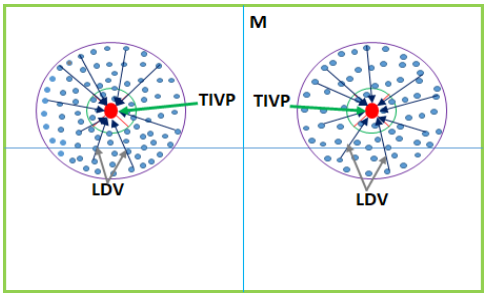
In 3-D MRI space, any specific location or organ's position can be traced if we can estimate the displacement vectors from that specific location or organ's position to any other positions. Using random voxel position as a center, we can generate multiple sample patches with corresponding displacement vectors from the voxel position to the region of interest. The distinguishable anatomical structural context around the ground truth can be used to train the CNN to estimate the displacement vectors. We can use the Hough voting strategy with the amalgamation of deep CNN to calculate the displacement vectors from input image features to the anatomy of interest. From the multiple patches, we estimated multiple displacement vectors, which lead to the target ROI's location. The estimated multiple displacement vectors, then, can be averaged and added with respective random patch centre to determine the ultimate point that is designed to be placed inside the anatomical interest. To determine the accurate location of the anatomy of interest point, two phase Hough-CNN based system have been designed in this study [15]. In the first phase, the global position of the anatomy of interest are estimated using the patches from the whole MRI scan except the boundary region. In the second phase, the patches are extracted in the vicinity of the ROI and those patches are used to train the multiple Hough-CNN models to determine the exact location of the target ROI. The global and local models are amalgamated in the test phase to estimate the accurate location of the anatomy of interest inside the test MRI scan. Using this approach, left and right hippocampi are localized from an MRI scan in this study.



(a) The displacement vector calculation from random patch centre to the target hippocampus locations.



(c) The global displacement vectors (GDV) estimation



(d) The local displacement vectors (LDV) estimation

Figure 1: The displacement vectors estimation from the random patches's centres to the target left and right hippocampi's location. (a) from any particular slice, the extracted patch centres to the target locations are shown for axial, coronal and sagittal slices. (b) The green dots residing inside the MRI are the centres of the global patches. The offsets from the centres to target image voxel positions (TIVP) are calculated to determine the global displacement vectors (GDV) for each patch. (c) Similarly, for the local models, the green dots residing inside the red circle in the vicinity of the TIVP are used to extract patches to training the local models. The offsets from the centres of the green dots to the red dot are the local displacement vectors (LDV) for the local models.

Now, let us consider that any volumatric MRI, $\mathrm{M}_{(\mathrm{XYZ})} : \mathbb{Z}^3 \rightarrow \mathbb{N}$ and three axes are X,Y,and Z in that MRI. If there are any random voxel position $\mathrm{P}_{(\mathrm{a_X},\mathrm{b_Y},\mathrm{c_Z})}$

inside the MRI scan and manually localized voxel position of any particular anatomy of interest, $Q_{(u_X, v_Y, w_Z)}$, then the displacement vectors between the random voxel position P and the manually localized voxel position of any anatomy of interest Q are X_{GT}, Y_{GT} and Z_{GT}.

$$X_{GT} = u_X - a_X \tag{1}$$

$$Y_{GT} = v_Y - b_Y \tag{2}$$

$$Z_{GT} = w_Z - c_Z \tag{3}$$

X_{GT}, Y_{GT}, and Z_{GT} are used as the ground truth while the proposed Hough-CNN are trained. a_X, b_X, and c_X are the patch centres, and using these centres, three view 2-D patches are extracted from MRI scan. The red and yellow dices in Figure 1(a) denote as the voxel location of the target ROI's and the random patch center of the slices of axial, sagittal, and coronal view of an MRI scan, respectively. The blue arrows reaching to the red dice are the displacement vectors of those patches of the specific slices of axial, sagittal, and coronal view from target ROI, i.e., hippocampus in this case. From multiple centres to the target ROI's displacement vectors are shown in Figure 1(b) and (c) imagining that the green dots are the patch centres and the yellow arrow lines are the displacement vectors from those patches, where the red dot is the target ROI.

B. Dataset

In this study, two MRI datasets are considered to verify the proposed approach. Alzheimer's disease neuroimaging initiative (ADNI) dataset consists of 351 MRI scans with three classes: AD, MCI, and NC. The MRI voxels size was 1mm^3 with

the dimension of 256x256x170. The number of subjects participated to generate the scans were 60. On the other hand, Gwangju Alzheimer's and related dementia (GARD) cohort dataset consist of 326 MRI scans with four classes: aAD, mAD, ADD, and NC. The GARD dataset has the dimension of 312x212x240 with a voxel size of 1mm^3. The total of 326 subjects of male and female participated in the scanning process.

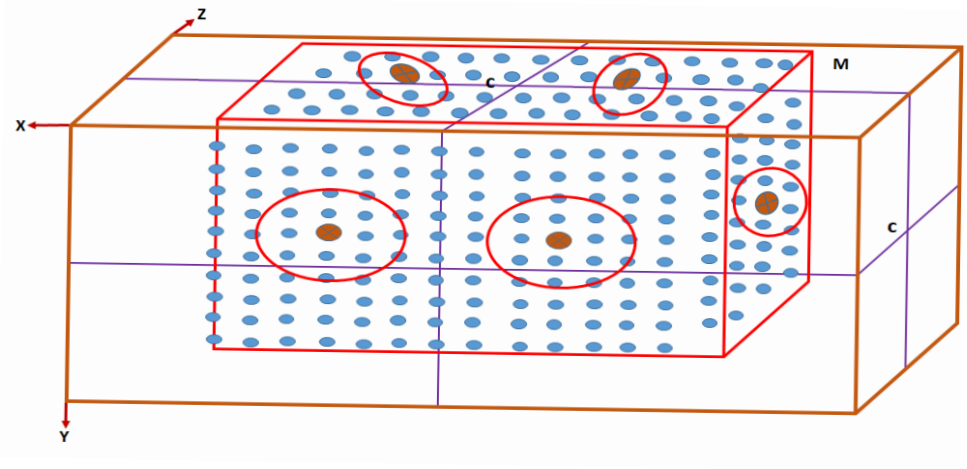


Figure 2: The patches centres residing inside the red cuboid is considered region to generate the patches. The whole cuboid is considered to generat patches for the global model, whereas, only the circular regions used to generate the patches for the local models. The center of the MRI is denoted by C and the circular red dot with a cross is the target left and right hippocampi's locations. X, Y, and Z are the three axes of the 3-D MRI scans.

C. Patch and Label Generation

MRI scan is highly complex and the dimension of the MRI is very large, which is not computationally efficient for the machine that is used to train the

models. Therefore, from 3-D MR imaging space, 2-D patches were extracted considering uniformly distributed random point within the MRI scan. The uniformly distributed random point is shown in Figure 2. The 3-D MRI scan were split into axial, coronal, and sagittal slices. The representative 2-D axial, coronal, and sagittal slices are shown in Figure 3 and 4.

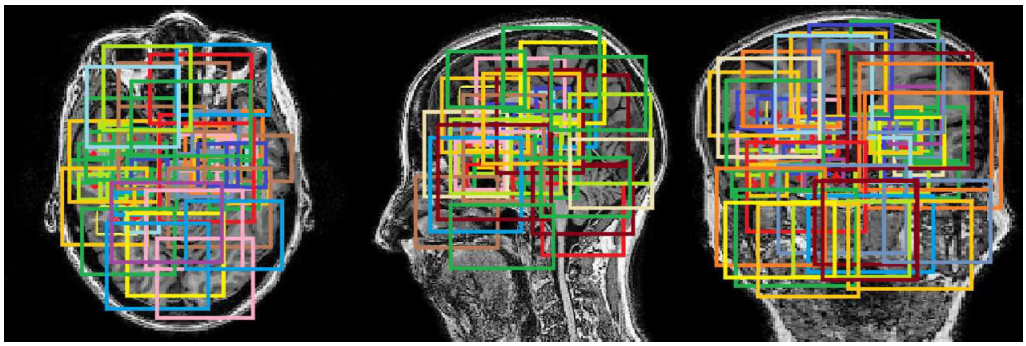


Figure 3: The whole MRI scan except in the boundary region are considered to extract patches using uniformly distributed sample points. From 3-D MRI space 2-D, 96x96 patches are extracted for the global model (GH-CNN). These extracted patches are downscaled to 32x32 and then merged them into 3-channel 32x32 patches

Using those uniformly distributed global points, the generated global patches from MRI slices are shown in Figure 3. The patches were extracted from whole MRI scan except the boundary region. 96x96 sized patches were extracted for the global model. The extracted patches were normalized using mean zero and standard deviation of 1. After that, the 96x96 sized patches were resized into 32x32 sized patches. These global patches contained the global information of the whole MRI along with the hippocampal information.

Similarly, the local patches were extracted using uniformly distributed random local points shown in Figure 4. The generated patches have a size

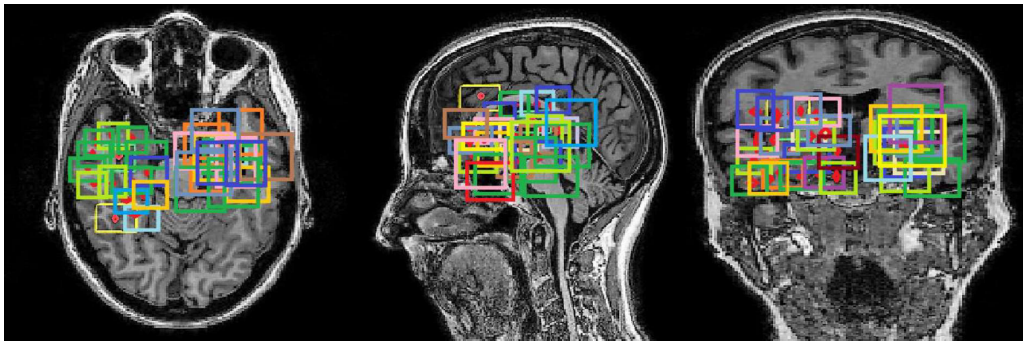


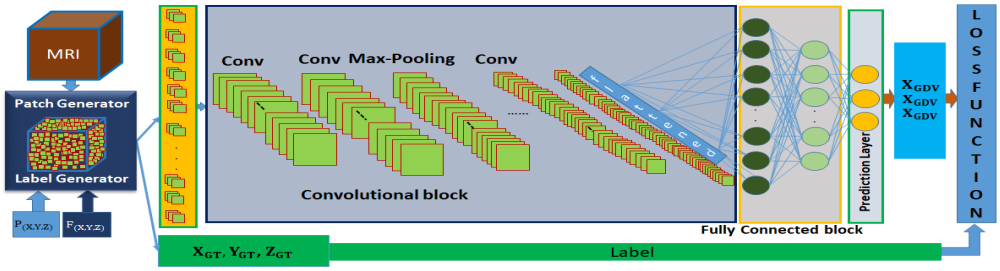
Figure 4: The local patches are generated from the vicinity of the target left and right hippocampi. From the target position of the hippocampus, 3-D 8x8x8 cubic region are selected to reside the local patches centres. The local patches' size of 32x32 are generated from the cubic cuboid and merged them into 3-channel 2-D patches for the local models.

of 32x32. Then, the extracted patches were normalized using mean zero and standard deviation of 1. These generated patches have the local information about the target ROI, such as hippocampus.

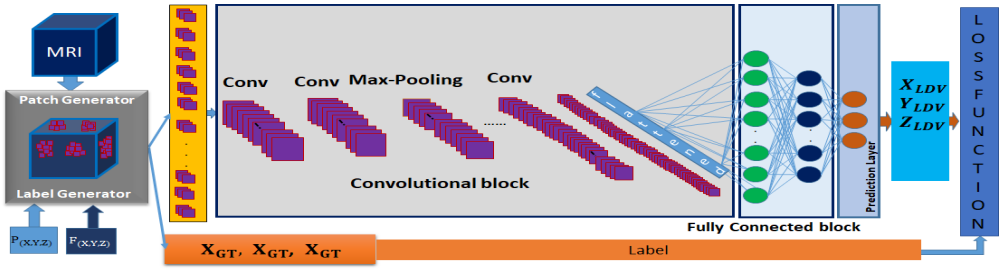
IV. Network Architecture

A. Convolutional Neural Network

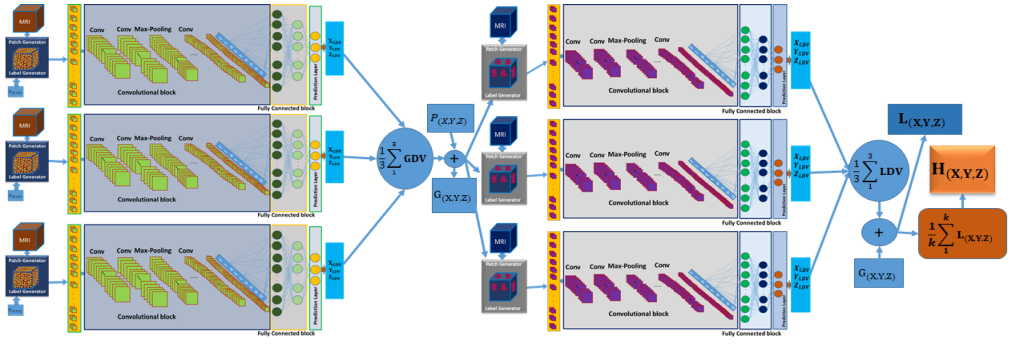
The proposed convolutional neural network [40], [41] model consists of a number of layers which conduct operations on input data (I_{size}). The Convolutional layers ($C_{kernel}^{\#filter}$) perform convolution operation on input images I_{size} with a number of preset kernels. They are usually followed by batch normalization layer [42] and activation function [43], where the normalization layer normalizes the results of convolution and activation function rescales the batch normalized outputs of convolution in a non linear manner. Pooling layers (P_{stride}^{Type}) are used to reduce the dimensionality of the outcomes produced by the previous layers through down-sampling. The type of the pooling can be max-pooling or average pooling. Finally, to extract the high level features, the fully connected layers ($F^{\#filter}$) are employed. The weights are optimized during training through back-propagation[44]. CNNs are good at extracting features without requiring any assistant from the user. The automatic extraction of image features through CNN solved so many complicated issues and helped to analyze complex shapes, orientation, and size from 2-D and 3-D medical and non-medical image data. In addition, the optimal features extraction is a part of the learning process. The data is processed through the layers in a feed-forward manner and the results of the network are compared with the ground-truth through a cost function and the error is back-propagated to update the weights of all the layers. The training continues until it converge. After completing the training, the prediction can be made by using the trained CNN model in a feed-forward manner and results can be reported from the last layer outputs.



(a) GH-CNN



(b) LH-CNN



(c) Two-Stage Ensemble Hough-CNN

Figure 5: The proposed two-stage ensemble Hough-CNN graphical architecture. (a) The generated patches from the whole MRI scan are used to train the global Hough convolutional neural network (GH-CNN). (b) On the other hand, the patches extracted from the vicinity of the MRI scan are used to train the local models. (c) The multiple trained CNN models are amalgamated together to form the ensemble global and local model and then form a two-stage ensemble Hough-CNN to estimate the target ROI locations.

Table 1: GH-CNN and LH-CNN network architectures used in the ADNI dataset.

Model Name	Network Architecture	
GH-CNN	GM_{lh}	$I_{32}, C_3^{16}, C_3^{16}, P_2^m, C_3^{32}, P_2^m, C_3^{32}, P_2^m, C_3^{64}, C_3^{128}, F^{128}, F^{128}, F^3$
	GM_{rh}	$I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3$
LH-CNN	LM_{lh}	$I_{32}, C_3^{32}, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, F^{512}, F^{256}, F^3$
	LM_{rh}	$I_{32}, C_3^{32}, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, F^{256}, F^{128}, F^3$

$I_{\text{sample size}} = \text{Network input}, C_{(\text{kernal size})}^{(\#\text{filter})} = \text{Convolutional layer},$
$P_2^m = \text{Max Pooling with stride } 2, F^{(\#\text{filter})} = \text{Fully connected layer}.$

B. Network Architecture Overview

Different types of network design topologies were proposed[16], [27], [45], [46] and [28] to analyze the brain ROIs. To localize the ROI, we have proposed a multi phase ensemble [15], [16], [47] based network design topology in this study. We have designed two different types of Hough-CNN architecture to address the localization of left and right hippocampi, a significant brain ROI which is known to be related with different neuro-degenerative disorders, from an MRI scan. In addition, Hippocampus is also known to have connection with our memory and learning process. We have constructed global Hough convolutional neural network (GH-CNN) and local Hough convolutional neural network (LH-CNN). GH-CNN is responsible for estimation of the global positioning of the anatomy

Table 2: GH-CNN and LH-CNN network architectures used in the GARD cohort dataset.

Model Name	Network Architecture	
GH-CNN	*$\mathrm{GM_{lh}}$	$\mathrm{I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3}$
	$\mathrm{GM_{rh}}$	$\mathrm{I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, P_2^m, C_3^{256}, C_3^{512}, F^{512}, F^{128}, F^3}$
LH-CNN	$\mathrm{LM_{lh}}$	$\mathrm{I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, C_3^{256}, F^{256}, F^{128}, F^3}$
	$\mathrm{LM_{rh}}$	$\mathrm{I_{32}, C_3^{16}, C_3^{32}, P_2^m, C_3^{64}, P_2^m, C_3^{128}, C_3^{256}, F^{256}, F^{128}, F^3}$

$$\mathrm{I_{sample\ size}} = \mathrm{Network\ input}, \mathrm{C^{(\#filter)}_{(kernal\ size)}} = \mathrm{Convolutional\ layer},$$
$$\mathrm{P_2^m} = \mathrm{Max\ Pooling\ with\ stride}\ 2, \mathrm{F^{(\#filter)}} = \mathrm{Fully\ connected\ layer}.$$

*One of $\mathrm{GM_{lh}}$ does not have a batch normalization layer, instead, it has dropout layer after first fully connected layer (f-1)(25 %) and second fully connected layer(f-2)(35 %)

of interest, whereas, the LH-CNN estimates the local position of the respective ROI localized by the GH-CNN.

C. Global Hough Convolutional Neural Network

The global models extract the global features from whole MRI scan and predict the global displacement vectors (GDV). Multiple GH-CNN models are aggregated together to form the ensemble GH-CNN model. The GH-CNN consists of 6 convolutional layers with 3 fully connected layers. All convolutional layers have the same kernel size (3x3) but the number of filters are different in different convolutional blocks. Each convolutional block is followed by batch

normalization layers and relu activation function. Three pooling layers are used to design the GH-CNN. The pooling layers are placed after 2^{nd}, 3^{rd}, and 4^{th} convolutional layers. The fully connected layers are also followed by a batch normalization layer and relu activation except the last fully connected layer. The last fully connected layer is only followed by a batch normalization layer. The GH-CNN architecture is shown in Figure 5(a) and the parameter detail is shown in Table 1 and 2

Algorithm 1 Global Model Training Procedures

1	Data: Input MRI volume, $\mathrm{M}_{(\mathrm{XYZ})}$,
2	Output: Trained GH-CNN model
3	Initialize the number of smaple k;
4	Initialize the global sample centres $GSP_{(XYZ)}$, where
5	$\mathrm{X} = \{\mathrm{X}_1, \mathrm{X}_2, \cdots \mathrm{X}_\mathrm{k}\}$,
6	$\mathrm{Y} = \{\mathrm{Y}_1, \mathrm{Y}_2, \cdots \mathrm{Y}_\mathrm{k}\}$,
7	$\mathrm{Z} = \{\mathrm{Z}_1, \mathrm{Z}_2, \cdots \mathrm{Z}_\mathrm{k}\}$,
8	Generate the global samples, $\mathrm{S}_\mathrm{g} = \{\mathrm{S}_1, \cdots \mathrm{S}_\mathrm{k}\}$ using (X,Y,Z)
9	Get the target ROIs Voxel location, $\mathrm{GTVP}_{\mathrm{XYZ}}$
10	Calculate the global displacement vector, $\mathrm{TGDV}_{(\mathrm{XYZ})} = \mathrm{GTVP}_{(\mathrm{XYZ})} - GSP_{(XYZ)}$
11	Initialize the number of epochs e;
12	For i : 1 to e do
13	Network input: S_g; Label: $\mathrm{TGDV}_{(\mathrm{XYZ})}$
14	Predict, $\mathrm{PGDV}_{(\mathrm{XYZ})}$
15	Calculate loss, $\mathrm{L} = \frac{1}{\mathrm{k}}\left((\mathrm{TGDV}_{(\mathrm{XYZ})} - \mathrm{PGDV}_{(\mathrm{XYZ})})^2\right)$;
16	Find the best $\mathrm{PGDV}_{(\mathrm{XYZ})}$ such that $\mathrm{L} \to 0$;

D. Local Hough Convolutional Neural Network

The neighboring area of target ROI contains the most important and distinguishable features. Therefore, the local models, which are trained to extract feature to localize the target ROI, plays the most important role. To localize the hippocampus from 3-D MRI scan, we proposed a local models which is responsible to estimate the exact location. Multiple local models are amalgamated to construct the ensemble LH-CNN architecture.

The LH-CNN structure is little different than the GH-CNN. The number of convolutional layers used in LH-CNN is not the same as GH-CNN. For GARD dataset, LH-CNN consists of 5 convolutional layers with 3 fully connected layers. The kernel size (3x3) is same as the GH-CNN. But the number of filters used in convolutional layers are different in different convolutional layers. The convolutional layers are aggregated with the batch normalization layers and relu activation function. Two max-pooling layers are used to construct the LH-CNN. The max-pooling layer is added after 2^{nd}, and 3^{rd} convolutional layers. The fully connected layers are amalgamated with relu activation function except the last fully connected layer. The relu activation are followed by a batch normalization layers. The last fully connected layer are followed only by a batch normalization layer. Similar network structure were used for ADNI Dataset as well, except that 4 convolutional layers were considered to build the model.

The local 3-channel 2-D 32x32 patches generated from the neighboring area of the hippocampus are used to train the LH-CNN. The detail patch generation techniques are explained in previous chapter. The local model learns the feature maps from the input patches and predicts local displacement vector (LDV) for each patch. After that, the averaged predicted LDV is calculated and added with

global model estimated hippocampus position. The resultant value of the two-stage ensemble Hough-CNN is the final predicated location of the hippocampus. Using this voxel location, three view-planes of axial, sagittal, and coronal slice of the hippocampus are displayed.

Algorithm 2 Local Model Training Procedures

1	Data: Input MRI volume, $\mathrm{M_{(XYZ)}}$,
2	Output: Trained LH-CNN model
3	Initialize the number of smaple k;
4	Initialize the local sample centres $LSP_{(XYZ)}$, where
5	$\mathrm{X} = \{\mathrm{X_1}, \mathrm{X_2}, \cdots \mathrm{X_k}\}$,
6	$\mathrm{Y} = \{\mathrm{Y_1}, \mathrm{Y_2}, \cdots \mathrm{Y_k}\}$,
7	$\mathrm{Z} = \{\mathrm{Z_1}, \mathrm{Z_2}, \cdots \mathrm{Z_k}\}$,
8	Generate the local samples, $\mathrm{S_l} = \{\mathrm{S_1}, \cdots \mathrm{S_k}\}$ using (X,Y,Z)
9	Get the target ROIs Voxel location, $\mathrm{LTVP_{XYZ}}$
10	Calculate the local displacement vector, $\mathrm{TLDV_{(XYZ)}} = \mathrm{LTVP_{(XYZ)}} - LSP_{(XYZ)}$
11	Initialize the number of epochs e;
12	For i : 1 to e do
13	Network input: $\mathrm{S_l}$; Label: $\mathrm{TLDV_{(XYZ)}}$
14	Predict, $\mathrm{PLDV_{(XYZ)}}$
15	Calculate loss, $\mathrm{L} = \frac{1}{\mathrm{k}} \left((\mathrm{TLDV_{(XYZ)}} - \mathrm{PLDV_{(XYZ)}})^2 \right)$;
16	Find the best $\mathrm{PLDV_{(XYZ)}}$ such that $\mathrm{L} \rightarrow 0$;

E. Loss Function

In case of optimization algorithm, the function that determine the difference in between the predicted output and the true output is known as a loss function or a cost function. To train the GH-CNN and LH-CNN, we used square error as a loss function. Now, if the predicted displacement vectors for an anatomy of interest in 3-dimensional space are $\mathrm{X_{PT}}$, $\mathrm{Y_{PT}}$, and $\mathrm{Z_{PT}}$, whereas, the target displacement vector are $\mathrm{X_{GT}}$, $\mathrm{Y_{GT}}$, and $\mathrm{Z_{GT}}$ considering axial, sagittal and coronal view, then the projected generalized loss function can be written as follows.

$$\mathrm{MSE} = \frac{1}{\mathrm{k}*\mathrm{q}}\left(\sum_{\mathrm{j}=1}^{\mathrm{j}=\mathrm{k}*\mathrm{q}}\right) \left(\frac{1}{3}\left((\mathrm{X_{GT_j}}-\mathrm{X_{PT_j}})^2+(\mathrm{Y_{GT_j}}-\mathrm{Y_{PT_j}})^2+(\mathrm{Z_{GT_j}}-\mathrm{Z_{PT_j}})^2\right)\right) \tag{4}$$

Now, for the GH-CNN, if the global displacement vectors are $\mathrm{X_{GDV}}$, $\mathrm{Y_{GDV}}$, and $\mathrm{Z_{GDV}}$, where the target displacement vectors are $\mathrm{X_{GT}}$, $\mathrm{Y_{GT}}$, and $\mathrm{Z_{GT}}$ considering axial, sagittal and coronal view, then the loss function for global models can be written as follows.

$$\mathrm{MSE_{GH-CNN}} = \frac{1}{\mathrm{k}*\mathrm{q}}\left(\sum_{\mathrm{j}=1}^{\mathrm{j}=\mathrm{k}*\mathrm{q}}\right) \left(\frac{1}{3}\left((\mathrm{X_{GT_j}}-\mathrm{X_{GDV_j}})^2+(\mathrm{Y_{GT_j}}-\mathrm{Y_{GDV_j}})^2+(\mathrm{Z_{GT_j}}-\mathrm{Z_{GDV_j}})^2\right)\right) \tag{5}$$

Similarly, for the LH-CNN, if the local displacement vectors are $\mathrm{X_{LDV}}$, $\mathrm{Y_{LDV}}$, and $\mathrm{Z_{LDV}}$, whereas, the target displacement vectors are $\mathrm{X_{GT}}$, $\mathrm{Y_{GT}}$, and $\mathrm{Z_{GT}}$ considering axial, sagittal, and coronal view, then the loss function for local models can be written as follows.

$$\mathrm{MSE}_{\mathrm{LH-CNN}} = \frac{1}{\mathrm{k}*\mathrm{q}}\left(\sum_{\mathrm{j}=1}^{\mathrm{j}=\mathrm{k}*\mathrm{q}}\right)$$
$$\left(\frac{1}{3}\left((\mathrm{X}_{\mathrm{GT_j}}-\mathrm{X}_{\mathrm{LDV_j}})^2+(\mathrm{Y}_{\mathrm{GT_j}}-\mathrm{Y}_{\mathrm{LDV_j}})^2+(\mathrm{Z}_{\mathrm{GT_j}}-\mathrm{Z}_{\mathrm{LDV_j}})^2\right)\right) \quad (6)$$

F. Training

For ADNI dataset, 231 MRI scans were used to generate the training patches, whereas, 56 MRI scans are used to generate the validation patches. On the other hand, 196 MRI scans were used to generate the training patches for the GARD cohort dataset. For validation, 65 MRI scans are used to extract patches. After generating the patches for GH-CNN and LH-CNN with their respective labels, we run the model for training. All the models (24) were trained separately. We have used the Adam optimizer[48] with square error cost function to observe the training progress. The adam optimizer's parameter setting were kept as default except the learning rate. We used 1e-5 to 1e-2 as a learning rate to train the 24 models. The training were continued total of 250 epochs for each global model, whereas, it was 200 epochs for each local model. The representative training and validation curves for the GH-CNN and LH-CNN of the GARD cohort dataset were shown in Figure 6 and 7, respectively. For the GH-CNN, the training time was an average of 10 hours, whereas, for the local model, it took on average of 6 hours to train the models. We performed 5-fold cross validation for both datasets.

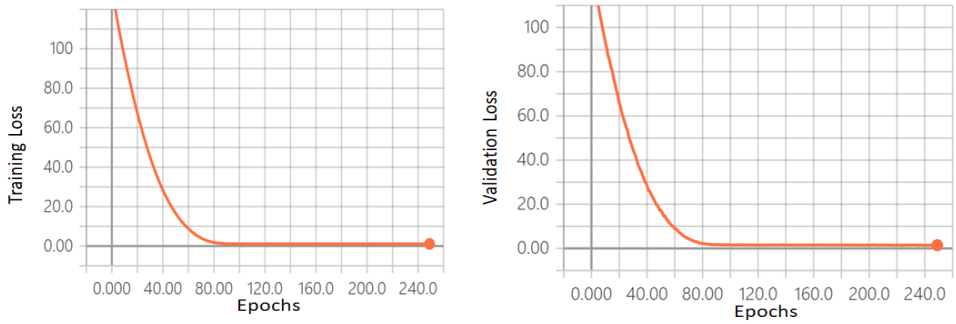


Figure 6: The representative training and validation loss curve of the GARD cohort dataset for the GH-CNN: Right Hippocampus.

The cross-validation results of the GARD and ADNI cohort dataset are shown in Table 3 to Table 6 and Table 7 to Table 10, respectively. All the training were performed on HP Workstation Intel Xeon Processor (3.10 GHz) with 32 GB RAM along with NVIDIA Quadro MD4000 GPU (8GB).

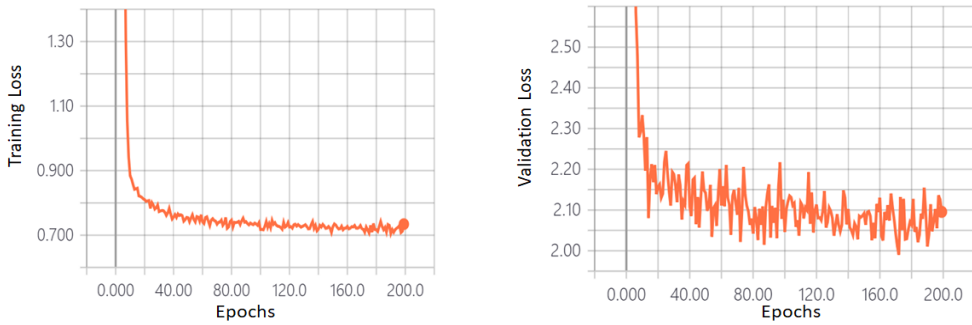


Figure 7: The representative training and validation loss curve of the GARD cohort dataset for the LH-CNN: Right Hippocampus.

Table 3: GARD cohort dataset: Left hippocampus (LH-CNN):

5-fold cross validation

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	0.7372	0.7447	0.7365	0.7359
Validation MSE	2.1765	2.0254	1.6226	1.8735
Testing MSE	2.9105	2.9269	2.9060	3.0835

Table 4: GARD cohort dataset: Left hippocampus ($\mathrm{GH-CNN}^{*}$):

5-fold cross validation

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	1.1158	1.0648	1.0728	1.0509
Validation MSE	1.0743	1.2216	1.0420	1.1747
Testing MSE	1.2307	1.2008	1.2163	1.2421

Table 5: GARD cohort dataset: Right hippocampus (LH-CNN):

5-fold cross validation

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	0.1915	0.1959	0.1961	0.1915
Validation MSE	1.8579	1.7032	1.3898	2.1815
Testing MSE	1.8435	1.8159	1.8443	1.9543

Table 6: GARD cohort dataset: Right hippocampus($GH-CNN^*$):
5-fold cross validation.

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	1.1020	1.0907	1.0834	1.0989
Validation MSE	1.2734	1.3316	1.1550	1.4988
Testing MSE	1.3435	1.3748	1.3694	1.4312

Table 7: ADNI dataset: Left hippocampus (LH-CNN):
5-fold cross validation

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	0.1904	0.1944	0.1909	0.1935
Validation MSE	1.8078	2.2081	2.4217	2.0039
Testing MSE	2.3643	2.6135	2.3467	2.5086

Table 8: ADNI dataset: Left hippocampus ($GH-CNN^*$):
5-fold cross validation

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	0.9827	0.8994	0.9385	0.9760
Validation MSE	1.4610	2.3999	2.0405	1.3011
Testing MSE	1.7051	1.6782	1.7966	1.8073

Table 9: ADNI dataset: Right hippocampus (LH-CNN):
5-fold cross validation.

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	0.6116	0.6192	0.6174	0.6208
Validation MSE	1.5231	1.6899	2.0639	1.4907
Testing MSE	1.8935	1.9638	2.1877	1.8086

Table 10: ADNI dataset: Right hippocampus ($GH-CNN^{*}$):
5-fold cross validation.

Type	Fold-1	Fold-2	Fold-3	Fold-4
Training MSE	0.7625	0.7501	0.7539	0.7659
Validation MSE	1.0095	0.9184	1.1342	0.7891
Testing MSE	1.2212	1.2016	1.2596	1.1607

*GH-CNN Model labels were down-scaled by 3 as the input image were down scaled by 3.
Therefore, all the GH-CNN model MSE values are needed to be up-scaled by 3 to get the actual MSE.

V. Localization Procedures

A. ROI Localization Proccedure

In this phase, the trained model were aggregated to construct the two-stage ensemble Hough network. Three trained global models were amalgamated to form the global ensemble network. Similarly, three trained local models were added together to form the local ensemble network. After that, the global ensemble model and the local ensemble model were aggregated together to form the single two-stage ensemble Hough-CNN.

From ADNI and GARD datasets, 56 and 65 MRI scans are used to test the two-stage ensemble Hough-CNN, respectively. The generated global patches from test MRI scans were used as input to the global ensemble network to predict the global displacement vectors. Multiple patches were generated from each MRI scan. Therefore, three global models predict multiple displacement vectors for multiple patches generated from each MRI scan. The predicted displacement vectors were added with the patch centres, which ultimately lead to the ROI locations. The predicted multiple ROI locations were averaged to determine the global position of the target ROI. After that, the predicted voxel location of the target ROI were compared with the manually localized voxel location to determine the global error.

The multiple global models predicted locations of the ROI were used to generate the patches for the local models. The generated patches were used as a input to the local ensemble network to predict the exact location of the ROI. The local models predict multiple displacement vectors for the multiple patches generated from each MRI scan. The predicted displacement vectors were added with the averaged multiple global ensemble model predicted ROI locations. Then,

the predicted ROI's voxel locations were averaged to generate the final voxel location of the target ROI.

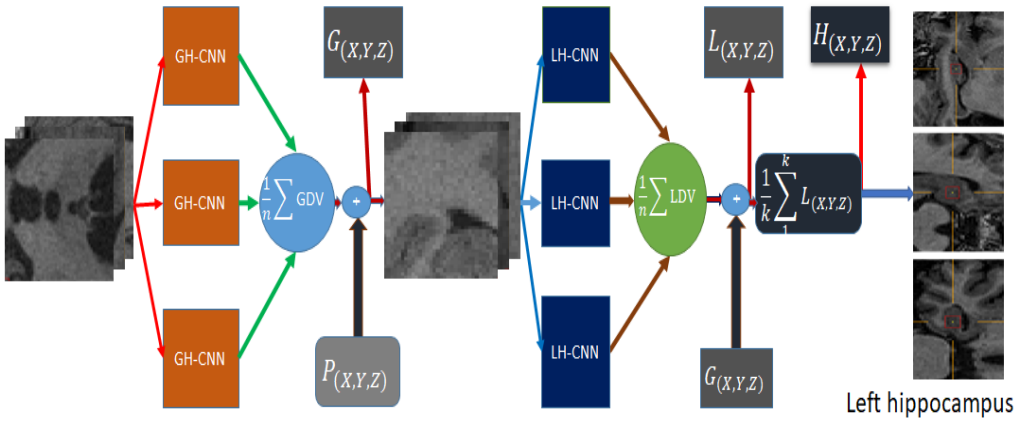


Figure 8: The localization of left hippocampal position from MRI scan using the proposed two-stage ensemble Hough-CNN. The ensemble view of the two-stage Hough-CNN in the test phase is shown here. In the test phase, 3-view patches from the axial, coronal, and sagittal slices of the MRI scan are used to estimate the global displacement vectors (GDV) and then estimated GDV are added with respective random sample patches' centres. The resultant values are the global position of the target ROI. Using these global position of the target ROI, the local patches are generated and then used as a input to the local models to estimate local displacement vector (LDV). The estimated LDV are added with the global position of the target ROI and then averaged to calculate the exact location of the ROI. Here, n is the number of models used in each phase and k is the number of patches generated from each MRI scan.

Now, let us consider that the ensemble global model predicted displacement vectors are $\mathrm{V}_{(\mathrm{X_{GDV}},\mathrm{Y_{GDV}},\mathrm{Z_{GDV}})} \in \mathbb{R}^3$ and the generated uniformly distributed random patches' centres are $\mathrm{P}_{(\mathrm{X,Y,Z})} \in \mathbb{N}^3$. Then, the global model predicted ROI's voxel locations, $G_{(X,X,Z)}$, can be calculated using the following

expression.

$$G_{(X,X,Z)} = \mathrm{P}_{(\mathrm{X,Y,Z})} + \mathrm{V}_{(\mathrm{X_{GDV}},\mathrm{Y_{GDV}},\mathrm{Z_{GDV}})} \tag{7}$$

The ensemble global model predicted ROI's voxel locations, $G_{(X,X,Z)}$ will be used to generate the patches for the local ensemble model. Consider that the ensemble local model predicted the displacement vectors are $\mathrm{U}_{(\mathrm{X_{LDV}},\mathrm{Y_{LDV}},\mathrm{Z_{LDV}})}$. On the other hand, The patches centres for local ensemble model are $G_{(X,X,Z)}$. Therefore, the ensemble local model predicted ROI's voxel location, $H_{(X,Y,Z)}$ can be calculated using the following expression.

$$H_{(X,Y,Z)} = \frac{1}{\mathrm{k}} \sum_{1}^{\mathrm{k}} \left(G_{(X,Y,Z)} + \mathrm{U}_{(\mathrm{X_{LDV}},\mathrm{Y_{LDV}},\mathrm{Z_{LDV}})} \right) \tag{8}$$

Here, k denotes the number of patches generated from each MRI scan.The whole process is shown in Figure 8. The two-stage localization procedure is illustrated in **Algorithm 3**.

B. Error Calculation

In ADNI and GARD dataset, the voxel volume size is $1\mathrm{mm}^3$. Therefore, the distance between the two voxels is 1mm. In 3-D MRI space, the root means squared difference between the manually localized target ROI's voxel position and the ensemble Hough-CNN model predicted voxel location is the prediction root mean square error (Euclidean distance) of the proposed automated method. Now, we can express the predicted root mean square error as follows.

$$\mathrm{PE_{rms}} = \sqrt{(H_{(X,Y,Z)} - \mathrm{F}_{(\mathrm{X,Y,Z})})^2} \tag{9}$$

Where $\mathrm{PE_{rms}}$ denoted as the predicted root mean square error. $\mathrm{H}_{(\mathrm{X,Y,Z})}$ and $\mathrm{F}_{(\mathrm{X,Y,Z})}$ are the two-stage ensemble Hough-CNN predicted ROI's voxel location

and the manually localized ROI's voxel locations, respectively. The estimated root mean sqaure error (Euclidean distance) of the left and right right hippocampi of the ADNI and GARD datasets are shown in Table 12 and 13. The average runtime to calculate the displacement vectors, which lead to the hippocampal position in an MRI, is 2 seconds.

C. Statistical Analaysis

To show the agreement of estimation between the manual and the proposed automatic Hough-CNN based method, we have performed the statistical test on the predicted result. Bland-Altman [49] mean-difference plot are used to test the substitution of any two methods on measuring the same object. The suitability of new method as a proxy with the accepted old method is validated by the statistical point of view using Bland-Altman plot. We have used the Bland-Altman plot to verify our proposed approach with manual delineation. The scatter plots are shown in Figure 10 (a-d). We considered the X, Y, Z axes positions as the axial, sagittal and coronal view voxel locations. After that, we compared them with the manual delineation of same axes to estimate the differences and plotted them against the mean values of the respective axes. The scatter plots were drawn considering the 95% line of agreement (LOA). Most of the data for left and right hippocampi of ADNI and GARD cohort datasets were confined by the 95% LOA, which indicates the similar estimation of the voxel locations by the both methods.

Furthermore, we have performed the correlation estimation between our method and the manual tracing of the hippocampus location in 3-D MR image space. Using the measured locations of the left and right hippocampi of MRI scans of ADNI and GARD cohort datasets, We have generated the plots utilizing

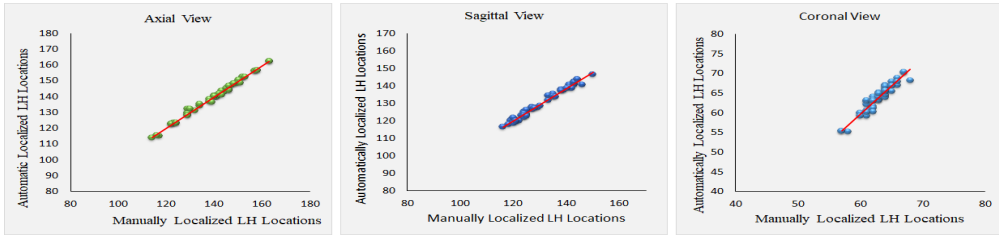
the SPSS 16.0 software package. The correlation of the manual results and the automatically predicted results are shown in Figure 9(a, b), and (c, d) for the left and right hippocampi of the ADNI and GARD cohort datasets, respectively. The measured p-value (P $<$0.001) in the test indicated that both the methods were statistically significant. In addition, the automatic localization method can be used a proxy for hippocampal position estimation from a 3-D MRI.

The estimated results were further analyzed to assess the suitability of the proposed method as a replacement for the manual delineation. We computed the Pearson squared correlation coefficient (R^2), where the high R^2 suggests that the automatic method is highly significant and can be used as proxy. We have reported the estimated R^2 as well as the slop of regression β values for the left and right hippocampi of the ADNI and GARD cohort datasets in Table 11. From Table 11, we can see that R^2 and β values are very close to 1, which indicate that the proposed method can be a replacement of the manual delineation for brain ROI localization from the MRI scan.

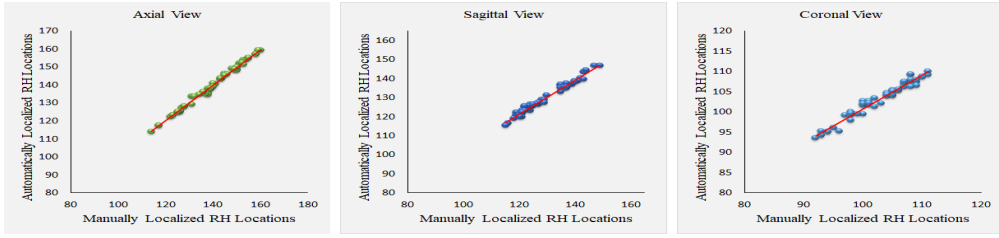
Table 11: Comparison of the proposed method and manual delineation: squared Pearson's correlation coefficient (R^2) along with the slop of regression (β), both with the 95% confidence intervals.

Dataset	R^2			β		
	Axial	Sagittal	Coronal	Axial	Sagittal	Coronal
ADNI: LH (56)	0.993	0.980	0.893	0.996	0.979	0.945
ADNI: RH (56)	0.993	0.979	0.963	0.996	0.990	0.981
GARD: LH (65)	0.771	0.939	0.890	0.878	0.969	0.943
GARD: RH (65)	0.852	0.947	0.935	0.923	0.973	0.967

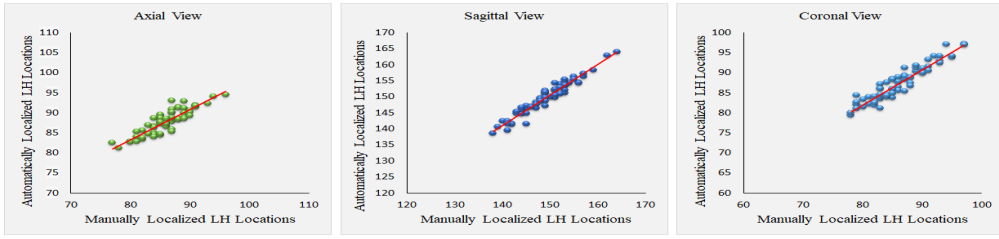
LH = Left Hippocampus, RH = Right Hippocampus



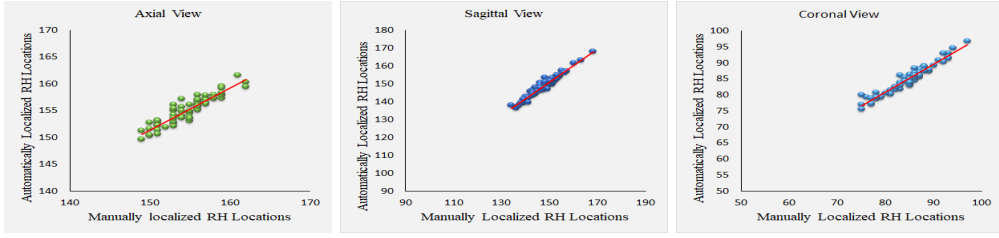
(a)



(b)

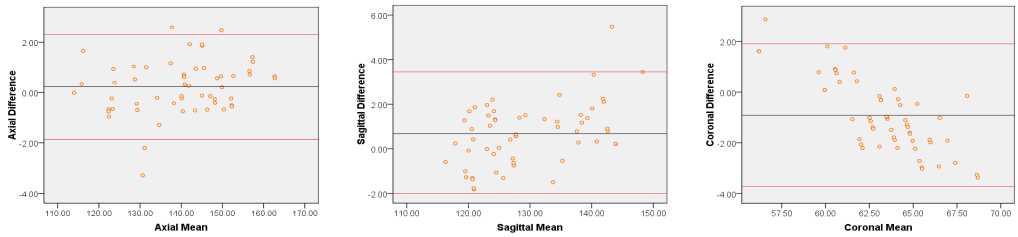


(c)

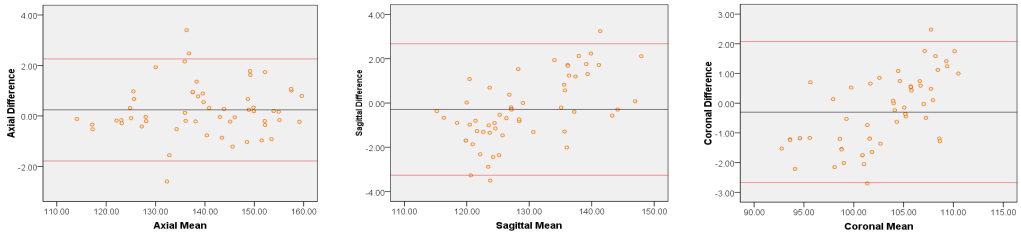


(d)

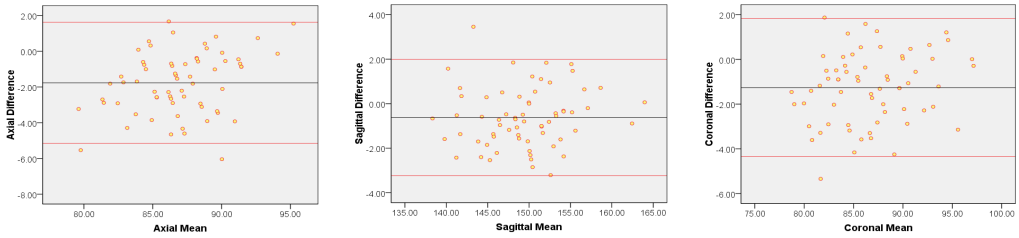
Figure 9: The left and right hippocampi of ADNI (56 MRI scans) (a, b) and GARD cohort (65 MRI scans) (c, d) datasets: scatter plots of estimated voxel locations by our proposed approach and the manually delineated positions with linear line of best fit (not forced through the origin).



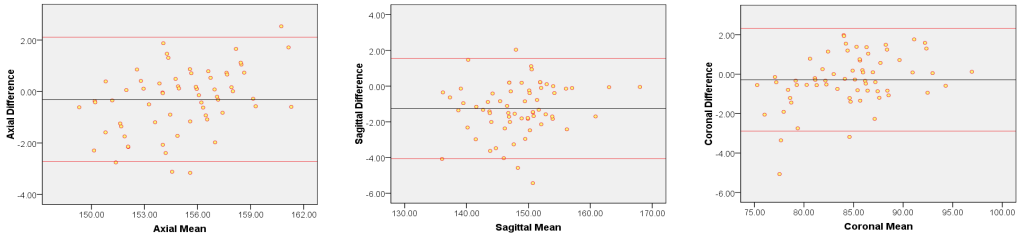
(a)



(b)



(c)



(d)

Figure 10: The left and Right hippocampi of ADNI (56 MRI scans) (a, b) and GARD cohort (65 MRI scans) (c, d) datasets: scatter plots of left and right hippocampi's positions estimation from MRI scans with 95% limits of agreement (LOA) confined by the red lines.

Algorithm 3 Hippocampus Localization

1	Data: Input MRI volume, $\mathrm{M_{XYZ}}$,
2	Output: Estimated locations, $H_{(X,Y,Z)}$,
3	Initialize the number of samples k,
4	Generate Sample Center, $P_{(X,Y,Z)}$, Where
5	$\mathrm{X} = \{\mathrm{X_1}, \mathrm{X_2}, \cdots \mathrm{X_k}\}$,
6	$\mathrm{Y} = \{\mathrm{Y_1}, \mathrm{Y_2}, \cdots \mathrm{Y_k}\}$,
7	$\mathrm{Z} = \{\mathrm{Z_1}, \mathrm{Z_2}, \cdots \mathrm{Z_k}\}$,
8	Generate global samples, $\mathrm{S_g} = \{\mathrm{S_1} \cdots \mathrm{S_k}\}$,
9	For i : 1 to k do
10	$\quad \mathrm{V_{(X_{GDV}, Y_{GDV}, Z_{GDV})}} \leftarrow \mathrm{S_g}$ to global trained model;
11	$\quad G_{(X,Y,Z)} = P_{(X,Y,Z)} + \mathrm{V_{(X_{GDV}, Y_{GDV}, Z_{GDV})}}$;
12	Use $\mathrm{G_{(X,Y,Z)}}$ as a random reference points,
13	Generate samples, $\mathrm{S_l} = \{\mathrm{S_1} \ldots \mathrm{S_k}\}$,
14	For i : 1 to k do
15	$\quad \mathrm{U_{(X_{LDV}, Y_{LDV}, Z_{LDV})}} \leftarrow \mathrm{S_l}$ to local trained model;
16	$\quad L_{(X,Y,Z)} = G_{(X,Y,Z)} + \mathrm{U_{(X_{LDV}, Y_{LDV}, Z_{LDV})}}$;
17	$H_{(X,Y,Z)} = \frac{1}{\mathrm{k}} \sum_{1}^{\mathrm{k}} \left(L_{(X,Y,Z)}\right)$;

D. Result

From ADNI dataset, 56 T1-weighted MRI scans of 8 subjects were used in the test phase to localize the left and right hippocampi.The automatic localized voxel positions were compare with the expert manual outlining. The estimated root mean square error between the predicated location and the manual outlining were

2.3187 mm and 2.0440 mm for the left and right hippocampi, respectively.

Table 12: Prediction error in the ADNI MRI dataset. The boldface numbers are the final error.

Dataset	Hippocampus	Model Name	Individual model's prediction error (mm)	Aggregated prediction error (mm)
ADNI MRI Dataset (343)	Left hippocampus	GM-1	3.2477	3.2745
		GM-2	3.2802	
		GM-3	3.3346	
		LM-1	2.3536	**2.3187**
		LM-2	2.3967	
		LM-3	2.2432	
	Right hippocampus	GM-1	2.9094	2.9741
		GM-2	3.0279	
		GM-3	3.0273	
		LM-1	2.0256	**2.0440**
		LM-2	2.1322	
		LM-3	2.0544	

GM = Global Model, LM = Local Model

Similarly, from GARD dataset, 65 T1-weighted MRI scans of 65 subjects were considered in the test phase to localize the left and right hippocampi. The left and right hippocampi of test MRI scans were manually localized previously

by the expert operator. The estimated root mean square error between our proposed approach and the manual outlining were 2.3267 mm and 2.2519 mm for the left and right hippocampi, respectively.

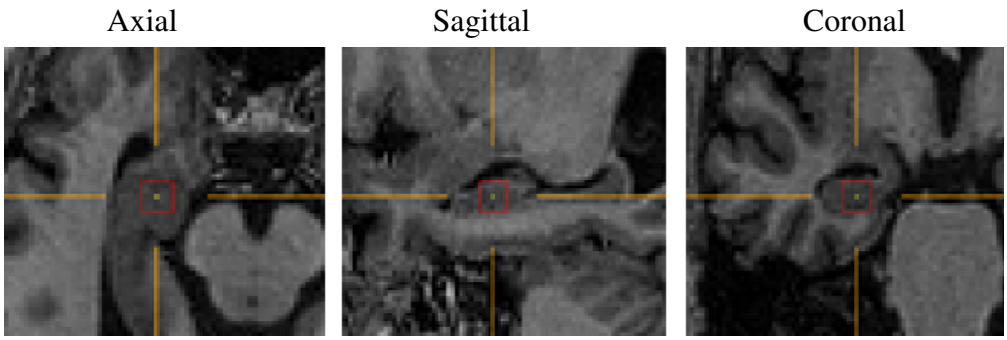
Table 13: Prediction error in the GARD cohort dataset. The boldface numbers are the final error.

Dataset	Hippocampus	Model Name	Individual model's prediction error (mm)	Aggregated prediction error (mm)
GARD cohort Dataset (326)	Left hippocampus	GM-1	3.6243	3.5228
		GM-2	3.4936	
		GM-3	3.5016	
		LM-1	2.3075	**2.3267**
		LM-2	2.3355	
		LM-3	2.3701	
	Right hippocampus	GM-1	3.7468	3.7108
		GM-2	3.6941	
		GM-3	3.7192	
		LM-1	2.3035	**2.2519**
		LM-2	2.2363	
		LM-3	2.2641	

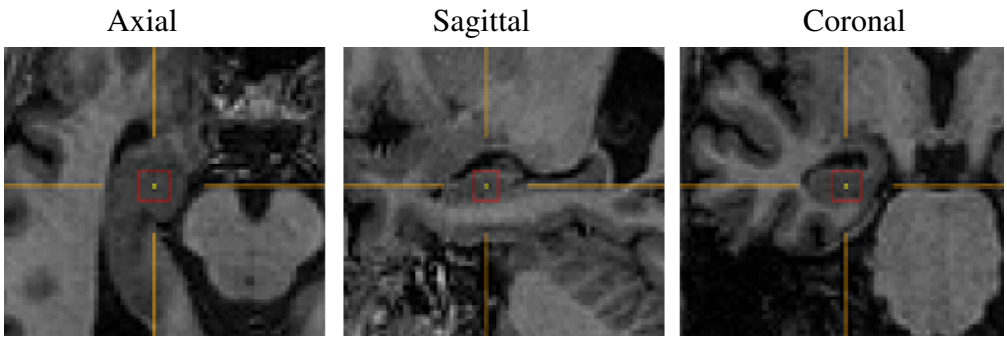
GM = Global Model, LM = Local Model

The representative localized three-view plane images of axial, sagittal and

coronal slices of the left and right hippocampi with the minimum and maximum root mean square error between the proposed approach and the manual outlining of ADNI and GARD cohort datasets are shown in Figure 11 to Figure 18.

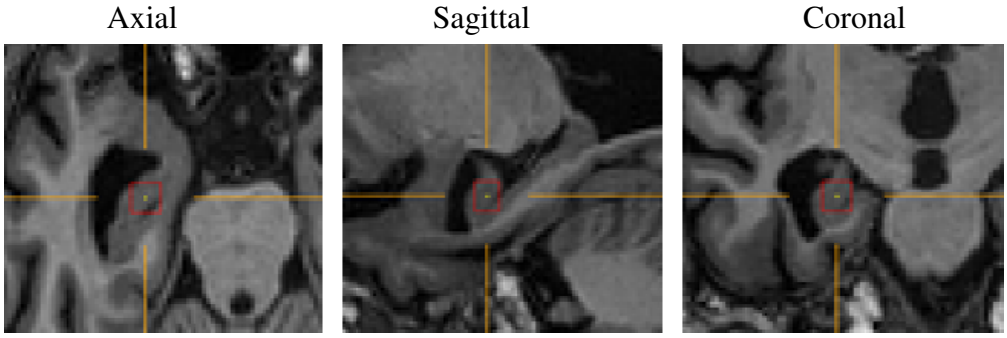


(a) Predicted hippocampus location

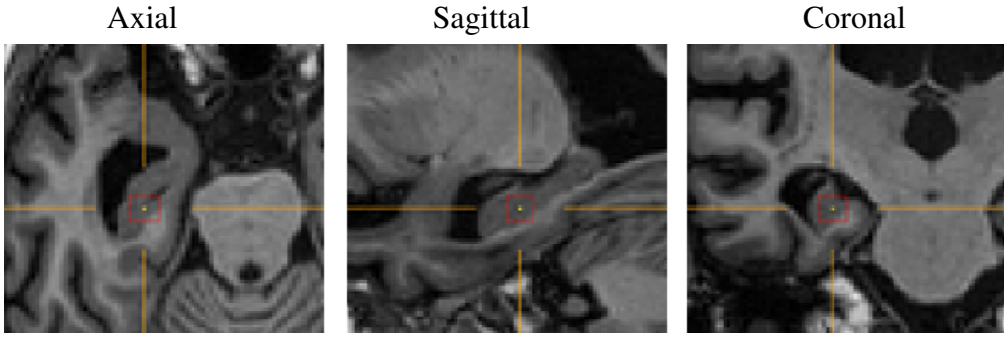


(b) Ground truth hippocampus location

Figure 11: ADNI Dataset (BEST CASE): Left hippocampus. Minimum RMS error is 0.7134 mm.

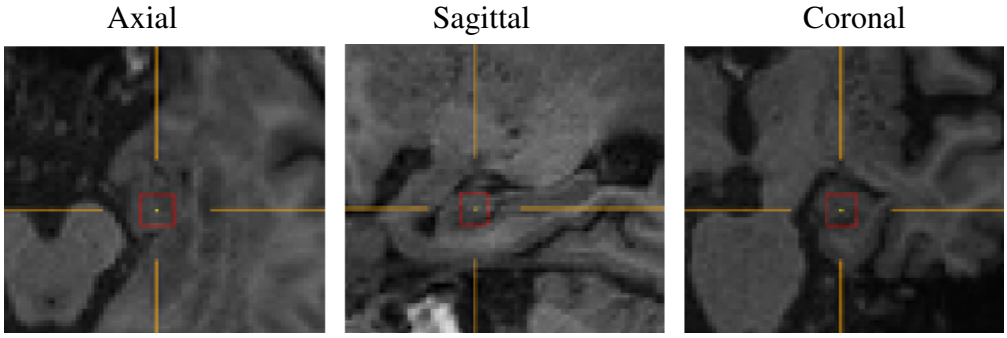


(a) Predicted hippocampus location

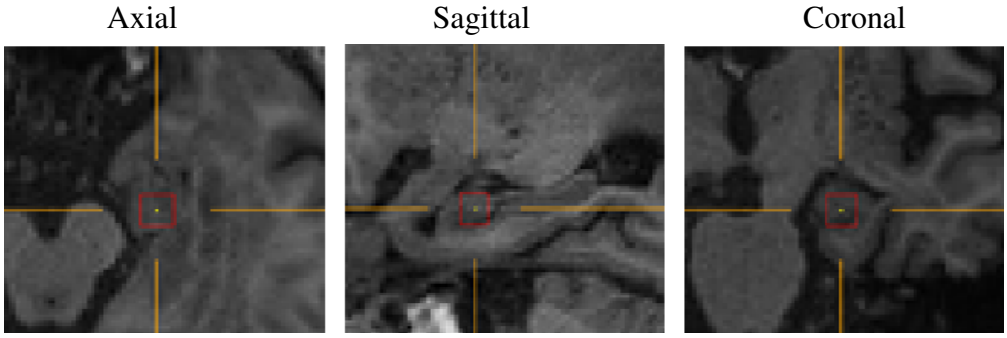


(b) Ground truth hippocampus location

Figure 12: ADNI Dataset (WORST CASE): Left hippocampus. Maximum RMS error is 5.5622 mm.

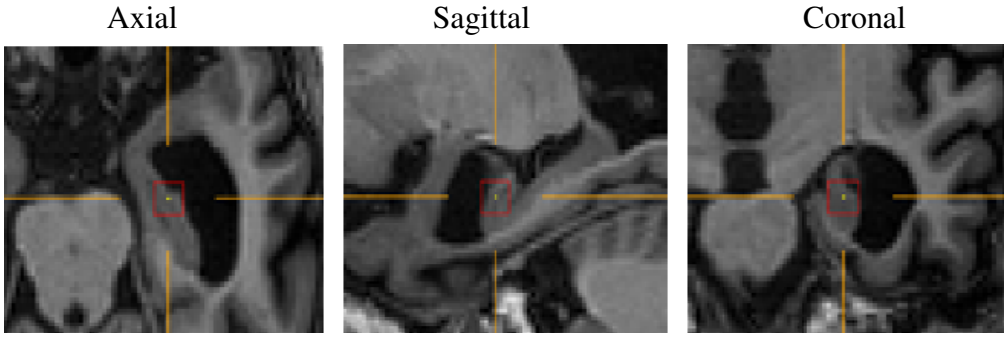


(a) Predicted hippocampus location

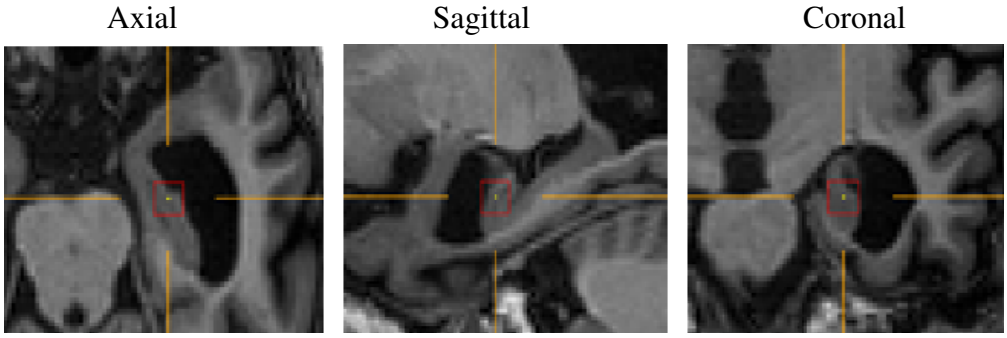


(b) Ground truth hippocampus location

Figure 13: AND Dataset (BEST CASE): Right hippocampus. Minimum RMS error is 0.1785 mm.

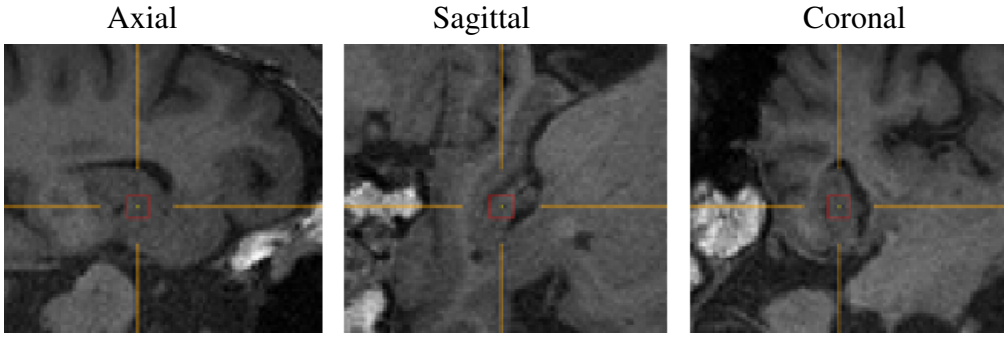


(a) Predicted hippocampus location

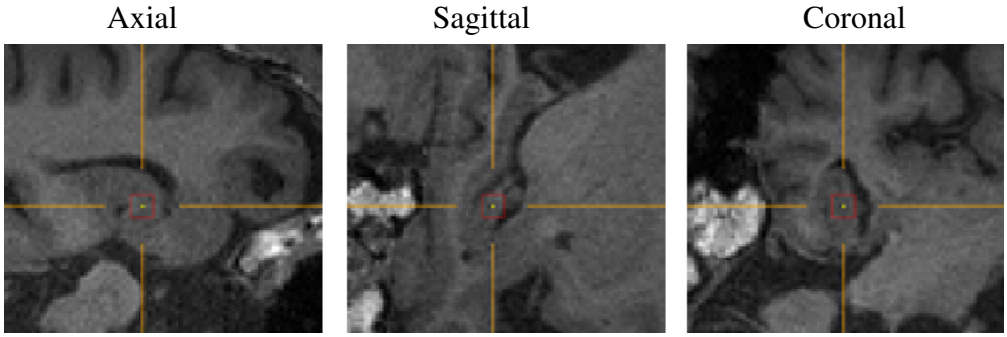


(b) Ground truth hippocampus location

Figure 14: ADNI Dataset (WORST CASE): Right hippocampus. Maximum RMS error is 4.3902 mm.

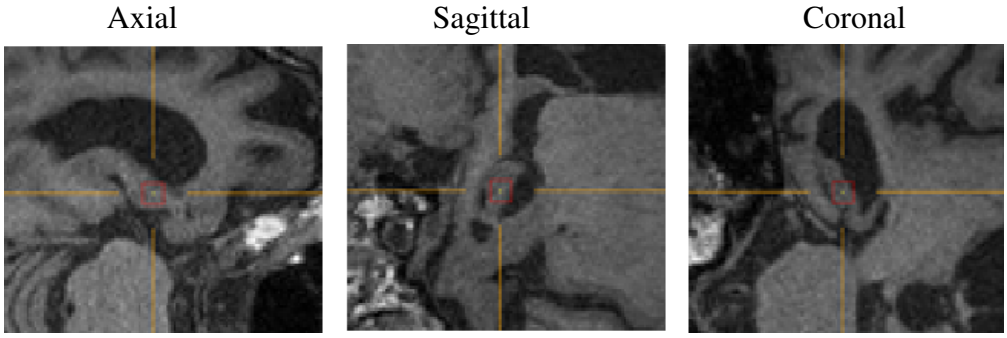


(a) Predicted hippocampus location

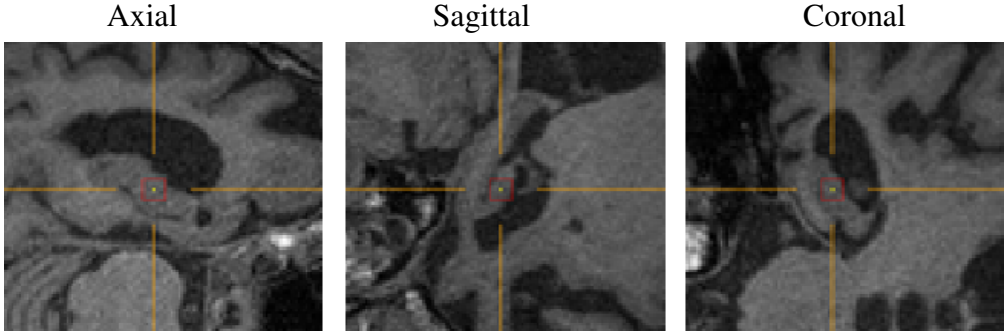


(b) Ground truth hippocampus location

Figure 15: GARD Dataset (BEST CASE): Left hippocampus. Minimum RMS error is 0.3062 mm.

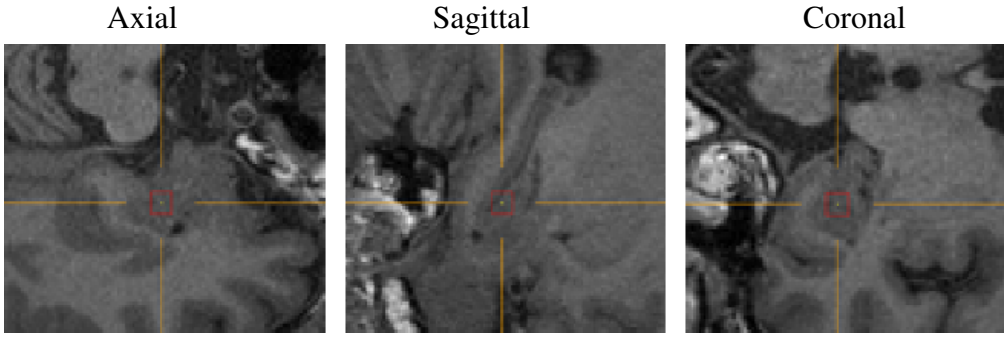


(a) Predicted hippocampus location

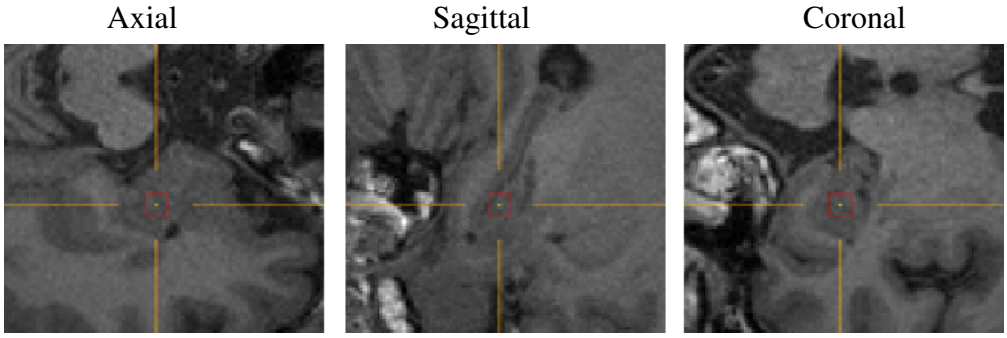


(b) Ground truth hippocampus location

Figure 16: GARD Dataset (WORST CASE): Left hippocampus. Maximum RMS error is 7.0523 mm.

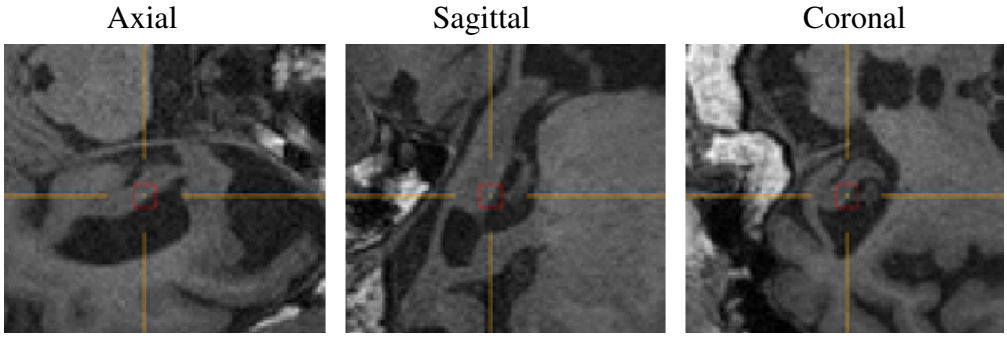


(a) Predicted hippocampus location

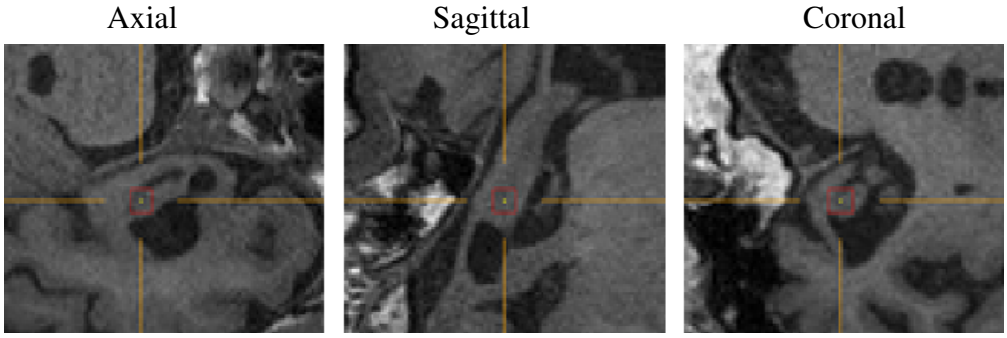


(b)Ground truth hippocampus location

Figure 17: GARD Dataset (BEST CASE): Right hippocampus. Minimum RMS error is 0.3546 mm.



(a) Predicted hippocampus location



(b) Ground truth hippocampus location

Figure 18: GARD Dataset (WORST CASE): Right hippocampus. Maximum RMS error is 5.8311 mm.

E. Discussion and Comparison

This study quantitatively assessed the automatic localization of ROI from MRI scan. The left and right hippocampi were localized as the representative ROI from

brain MRI scan. We have used the Hough voting strategy with the deep CNN to accomplish the automatic estimation of ROI. The Hough votes concern about a collection of evidences where each evidence refers to product set of various locations, aspects, scales in a Hough space and the peak of all Hough votes direct to an existence of an object. The deep CNN calculate the displacement vectors directing toward the ROI location based on the instances of Hough votes. Thus, the Hough-CNN is constructed in this study.

The ADNI MRI dataset of 343 MRI scans and GARD cohort dataset of 326 MRI scans were used to train, validate and test the proposed approach. After training all the GH-CNN and LH-CNN models, we have used them to construct the two-stage Hough-CNN. The constructed ensemble model were used to quantitatively estimate the left and right hippocampi's location in an MRI scan. Then, we have compared the result with the manual delineation to calculate the root mean square error (Euclidean distance) from the predicted voxel location to the target voxel location.

Achuthan et al.[1] estimated the root mean square error (Euclidean distance) of 3.5 mm from the predicted hippocampus location to the ground truth location. Our proposed approach has an acceptable average root mean square error of 2.24 mm from the approximated hippocampal position to the ground truth hippocampal location. A comparative analysis between the proposed method and other existing methods is shown in Table 14.

We have further analyzed our estimated hippocampal location by comparing with manual delineation using various statistical approach to show whether the proposed method can be a replacement of the gold standard manual approach. The statistical analysis showed that our proposed automatic delineation of brain ROI can be used as proxy.

Table 14: The comparative analysis between proposed method and other existing methods

Methods	Localized Organ's Name	Average RMS Error (mm)	
A. Achuthen et. al.[1]	Hippocampus (40-datasets [50])	3.5	
W.A. AI et. al.[15]	Aortic valve (CT)	2.04	
Y. Zheng et. al. [51]	Aortic valve (CT)	2.11	
M. Elattar et. al.[52]	Aortic valve (CT)	2.65	
Our proposed method	Left hippocampus (ADNI (56))	2.32	2.24
	Right hippocampus (ADNI (56))	2.04	
	Left hippocampus (GARD(65))	2.33	
	Right hippocampus(GARD(65))	2.25	

VI. CONCLUSION

In this thesis, an automatic method has been proposed to localize the ROIs from an MRI scan using two-stage Hough-CNN. The Hough-CNN was formed by the amalgamation of the Hough voting strategy with a deep CNN. To calculate the displacement vector from the target ROI, a deep CNN extracts the feature maps from input sample patches to vote for the center of the sample. We introduced a two-stage ensemble-based architecture to learn the global and local feature maps independently from the target MRI. The ensemble GH-CNN model has three different CNN models and they learn the global information from the training data to predict the global displacement vectors to lead the models to the target ROI using extracted global patches from the whole MRI scan. Similarly, the ensemble LH-CNN model has also three independent CNN models, where each model learns about the distinguishable local feature maps from extracted local patches to predict the local displacement vectors which lead to the exact location of the target ROI, i.e., hippocampus.

In the test phase, the global model predicts the global displacement vectors which will aggregated with patches centers to determine global location of the hippocampus. Then, those global position were used to generate the local patches and passed through the LH-CNN to predict the local displacement vectors. After that, the averaged the predicted displacement vector were estimated to add up with the patches centers to estimate the exact location of target ROI.

We considered in this study to localize the left and right hippocampi from MRI scan. The left and right hippocampi were localized from ADNI dataset of 56 MRI scans and the GARD cohort dataset of 65 MRI scans.The root mean square error were used to see the ensemble model performances on new data.

The estimated average rms error was 2.3187 mm and 2.0440 mm for the left and right hippocampi of 56 MRI scans of ADNI dataset, respectively. Similarly, for GARD dataset of 65 MRI scans, the average rms error was 2.3267 mm and 2.2519 mm for the left and right hippocampi, respectively.

The statistical analysis showed that the proposed approach is highly correlated with the manually delineated hippocampus location. The computed Pearson correlation coefficient strongly suggests that the proposed approach can be a substitute of manual localization. Bland-Altman study also lead to the same conclusion. Therefore, the proposed approach can be a substitute of the gold standard manual delineation of brain ROI from MRI scan.

PUBLICATIONS

A. Journals

1. A. Basher, K. Y. Choi, J. J. Lee, *et al.*, "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. [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2920005.

B. Conferences

1. A. Basher, S. Ahmed, and H. Y. Jung, "Ensembles of 3-d cnn models for pure volume measurement of hippocampi in smri", *The 8th International Conference on Smart Media and Application*, 2019.

2. A. Basher and H. Y. Jung, "Hippocampus localization using hough-cnn", *The 4th International Conference on Next Generation Computing (http://www.icngc.org/), Ba Ria- Vung Tau University, Vung Tau City, Vietnam.*, 2018.

3. A. Basher, A. N. R. Reza, S. Ahmed, *et al.*, "A survey on residual network evolution", *Korean Institute of Next Generation Computing, Spring Conference*, 2018.

REFERENCES

[1] A. Achuthan, M. Rajeswari, and W. M. S. Jalaluddin, "Hippocampus localization guided by coherent point drift registration using assembled point set", in *Hybrid Artificial Intelligent Systems - 8th International Conference, HAIS 2013, Salamanca, Spain, September 11-13, 2013. Proceedings*, J. Pan, M. M. Polycarpou, M. Wozniak, A. C. P. L. F. de Carvalho, H. Quintián-Pardo, and E. Corchado, Eds., ser. Lecture Notes in Computer Science, vol. 8073, Springer, 2013, pp. 92–102, ISBN: 978-3-642-40845-8. DOI: 10.1007/978-3-642-40846-5_10. [Online]. Available: https://doi.org/10.1007/978-3-642-40846-5_10.

[2] M. Siadat, H. Soltanian-Zadeh, and K. V. Elisevich, "Knowledge-based localization of hippocampus in human brain MRI", *Comp. in Bio. and Med.*, vol. 37, no. 9, pp. 1342–1360, 2007. DOI: 10.1016/j.compbiomed.2006.12.010. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2006.12.010.

[3] S. Zhao, D. Zhang, X. Song, and W. Tan, "Segmentation of hippocampus in MRI images based on the improved level set", in *4th International Symposium on Computational Intelligence and Design, ISCID 2011, Hangzhou, China, October 28-30, 2011, 2 Volumes*, IEEE Computer Society, 2011, pp. 123–126, ISBN: 978-1-4577-1085-8. DOI: 10.1109/ISCID.2011.39. [Online]. Available: https://doi.org/10.1109/ISCID.2011.39.

[4] M. Hajiesmaeili and M. Amirfakhrian, "A new approach to locate the hippocampus nest in brain mr images", in *2017 3rd International*

Conference on Pattern Recognition and Image Analysis (IPRIA), IEEE, 2017, pp. 140–145.

[5] 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.*, M. De Marsico, M. A. T. Figueiredo, and A. L. N. Fred, Eds., SciTePress, 2015, pp. 355–362, ISBN: 978-989-758-077-2.

[6] D. Gamberger, B. Zenko, A. Mitelpunkt, N. Shachar, and N. Lavrac, "Clusters of male and female alzheimer's disease patients in the alzheimer's disease neuroimaging initiative (ADNI) database", *Brain Informatics*, vol. 3, no. 3, pp. 169–179, 2016. DOI: 10.1007/s40708-016-0035-5. [Online]. Available: https://doi.org/10.1007/s40708-016-0035-5.

[7] J. L. Winterburn, J. C. Pruessner, S. Chavez, M. M. Schira, N. J. Lobaugh, A. N. Voineskos, and M. M. Chakravarty, "A novel in vivo atlas of human hippocampal subfields using high-resolution 3 T magnetic resonance imaging", *NeuroImage*, vol. 74, pp. 254–265, 2013. DOI: 10.1016/j.neuroimage.2013.02.003. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2013.02.003.

[8] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, and O. Colliot, "Fully automatic hippocampus segmentation and classification in alzheimer's disease and mild cognitive impairment applied on data from adni", *Hippocampus*, vol. 19, no. 6, pp. 579–587, 2009.

[9] M. Suzuki, H. Hagino, S. Nohara, S.-Y. Zhou, Y. Kawasaki, T. Takahashi, M. Matsui, H. Seto, T. Ono, and M. Kurachi, "Male-specific volume expansion of the human hippocampus during adolescence", *Cerebral Cortex*, vol. 15, no. 2, pp. 187–193, 2004.

[10] J. Gall, A. Yao, N. Razavi, L. J. V. Gool, and V. S. Lempitsky, "Hough forests for object detection, tracking, and action recognition", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2188–2202, 2011. DOI: 10.1109/TPAMI.2011.70. [Online]. Available: https://doi.org/10.1109/TPAMI.2011.70.

[11] T. D. Do, L. Vu, V. H. Nguyen, and H. Kim, "Full weighting hough forests for object detection", in *11th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2014, Seoul, South Korea, August 26-29, 2014*, IEEE Computer Society, 2014, pp. 253–258, ISBN: 978-1-4799-4871-0. DOI: 10.1109/AVSS.2014.6918677. [Online]. Available: https://doi.org/10.1109/AVSS.2014.6918677.

[12] G. Riegler, D. Ferstl, M. Rüther, and H. Bischof, "Hough networks for head pose estimation and facial feature localization", in *British Machine Vision Conference, BMVC 2014, Nottingham, UK, September 1-5, 2014*, M. F. Valstar, A. P. French, and T. P. Pridmore, Eds., BMVA Press, 2014. [Online]. Available: http://www.bmva.org/bmvc/2014/papers/paper039/index.html.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, pp. 436–44, May 2015. DOI: 10.1038/nature14539.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] W. A. Al, H. Y. Jung, I. D. Yun, Y. Jang, H.-B. Park, and H.-J. Chang, "Automatic aortic valve landmark localization in coronary ct angiography using colonial walk", *PLOS ONE*, vol. 13, no. 7, pp. 1–23, Jul. 2018. DOI: 10.1371/journal.pone.0200317. [Online]. Available: https://doi.org/10.1371/journal.pone.0200317.

[16] M. K. Abd-Ellah, A. I. Awad, A. A. Khalaf, and H. F. Hamed, "Two-phase multi-model automatic brain tumour diagnosis system from magnetic resonance images using convolutional neural networks", *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, p. 97, 2018.

[17] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability", *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[18] X. Lu and S. Luo, "The application of watersnakes algorithm in segmentation of the hippocampus from brain mr image", in *International Conference on Medical Imaging and Informatics*, Springer, 2007, pp. 277–286.

[19] O. Colliot, O. Camara, and I. Bloch, "Integration of fuzzy spatial relations in deformable models—application to brain mri segmentation", *Pattern recognition*, vol. 39, no. 8, pp. 1401–1414, 2006.

[20] O. Nempont, J. Atif, E. Angelini, and I. Bloch, "Combining radiometric and spatial structural information in a new metric for minimal surface segmentation", in *Biennial International Conference on Information Processing in Medical Imaging*, Springer, 2007, pp. 283–295.

[21] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images", *Computer methods and programs in biomedicine*, vol. 104, no. 3, e158–e177, 2011.

[22] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation", *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.

[23] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3d medical image segmentation: A review", *Medical image analysis*, vol. 13, no. 4, pp. 543–563, 2009.

[24] I. B. Malone, K. K. Leung, S. Clegg, J. Barnes, J. L. Whitwell, J. Ashburner, N. C. Fox, and G. R. Ridgway, "Accurate automatic estimation of total intracranial volume: A nuisance variable with less nuisance", *NeuroImage*, vol. 104, pp. 366–372, 2015. DOI: 10.1016/j.neuroimage.2014.09.034. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2014.09.034.

[25] J. L. Winterburn, J. C. Pruessner, S. Chavez, M. M. Schira, N. J. Lobaugh, A. N. Voineskos, and M. M. Chakravarty, "A novel in vivo atlas of human hippocampal subfields using high-resolution 3 t magnetic resonance imaging", *Neuroimage*, vol. 74, pp. 254–265, 2013.

[26] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson, "A bayesian model of shape and appearance for subcortical brain segmentation", *NeuroImage*, vol. 56, no. 3, pp. 907–922, 2011. DOI: 10.1016/j.neuroimage.2011.02.046. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2011.02.046.

[27] N Sriraam, S Raghu, K. Tamanna, L. Narayan, M. Khanum, A. Hegde, and A. B. Kumar, "Automated epileptic seizures detection using multi-features and multilayer perceptron neural network", *Brain informatics*, vol. 5, no. 2, p. 10, 2018.

[28] J. Dolz, C. Desrosiers, and I. B. Ayed, "3d fully convolutional networks for subcortical segmentation in MRI: A large-scale study", *NeuroImage*, vol. 170, pp. 456–470, 2018. DOI: 10.1016/j.neuroimage.2017.04.039. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2017.04.039.

[29] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes", *Pattern recognition*, vol. 13, no. 2, pp. 111–122, 1981.

[30] A. Yao, J. Gall, and L. Van Gool, "A hough transform-based voting framework for action recognition", in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2010, pp. 2061–2068.

[31] H. Ruppertshofen, C. Lorenz, G. Rose, and H. Schramm, "Discriminative generalized hough transform for object localization in medical images", *International journal of computer assisted radiology and surgery*, vol. 8, no. 4, pp. 593–606, 2013.

[32] O. Barinova, V. Lempitsky, and P. Kholi, "On detection of multiple object instances using hough transforms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1773–1784, 2012.

[33] M. Godec, P. M. Roth, and H. Bischof, "Hough-based tracking of non-rigid objects", *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1245–1256, 2013.

[34] A. Tran and A. Manzanera, "Fast growing hough forest as a stable model for object detection", in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, 2016, pp. 1–6.

[35] C. Henderson and E. Izquierdo, "Rethinking random hough forests for video database indexing and pattern search", *Computational Visual Media*, vol. 2, no. 2, pp. 143–152, 2016.

[36] T. D. Do, L. Vu, V. H. Nguyen, and H. Kim, "Full weighting hough forests for object detection", in *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2014, pp. 253–258.

[37] G. Riegler, D. Ferstl, M. Rüther, and H. Bischof, "Hough networks for head pose estimation and facial feature localization", *Journal of Computer Vision*, vol. 101, no. 3, pp. 437–458, 2013.

[38] A. Tejani, R. Kouskouridas, A. Doumanoglou, D. Tang, and T.-K. Kim, "Latent-class hough forests for 6 dof object pose estimation", *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 1, pp. 119–132, 2017.

[39] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, *et al.*, "Hough-cnn: Deep learning for segmentation of deep brain regions in mri and ultrasound", *Computer Vision and Image Understanding*, vol. 164, pp. 92–102, 2017.

[40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", in *Proceedings of the IEEE*, 1998, pp. 2278–2324.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.*, P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., 2012, pp. 1106–1114. [Online]. Available: `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks`.

[42] 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*, F. R. Bach and D. M. Blei, Eds., ser. JMLR Workshop and Conference Proceedings, vol. 37, JMLR.org, 2015, pp. 448–456. [Online]. Available: `http://proceedings.mlr.press/v37/ioffe15.html`.

[43] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", in *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa,*

Israel, J. Fürnkranz and T. Joachims, Eds., Omnipress, 2010, pp. 807–814. [Online]. Available: https://icml.cc/Conferences/2010/papers/432.pdf.

[44] 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.

[45] N. V. Shree and T. Kumar, "Identification and classification of brain tumor mri images with feature extraction using dwt and probabilistic neural network", *Brain informatics*, vol. 5, no. 1, pp. 23–30, 2018.

[46] K Somasundaram and T Genish, "The extraction of hippocampus from mri of human brain using morphological and image binarization techniques", in *2014 International Conference on electronics and communication systems (ICECS)*, IEEE, 2014, pp. 1–5.

[47] D. J. Dittman, T. M. Khoshgoftaar, and A. Napolitano, "Selecting the appropriate ensemble learning approach for balanced bioinformatics data", in *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2015, Hollywood, Florida, USA, May 18-20, 2015.*, I. Russell and W. Eberle, Eds., AAAI Press, 2015, pp. 329–334. [Online]. Available: http://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS15/paper/view/10366.

[48] 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*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980.

[49] J. M. Bland and D. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement”, *The lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[50] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga, “Construction of a 3d probabilistic atlas of human cortical structures”, *Neuroimage*, vol. 39, no. 3, pp. 1064–1080, 2008.

[51] Y. Zheng, M. John, R. Liao, A. Nottling, J. Boese, J. Kempfert, T. Walther, G. Brockmann, and D. Comaniciu, “Automatic aorta segmentation and valve landmark detection in c-arm ct for transcatheter aortic valve implantation”, *IEEE transactions on medical imaging*, vol. 31, no. 12, pp. 2307–2321, 2012.

[52] M. Elattar, E. Wiegerinck, F. van Kesteren, L. Dubois, N. Planken, E. Vanbavel, J. Baan, and H. Marquering, “Automatic aortic root landmark detection in cta images for preprocedural planning of transcatheter aortic valve implantation”, *The international journal of cardiovascular imaging*, vol. 32, no. 3, pp. 501–511, 2016.

ACKNOWLEDGEMENTS

First, I would love to express my heartiest thanks and profound indebtedness to my advisor, Professor Jung Ho Yub, for his endless patience, scholarly guidance and care during the entire program. His continuous encouragement, constructive criticism and constant supervision have dragged the best out of me. Without his guidance and supports, I could not have finished this thesis work.

In addition, I would like to thank the committee members, Professor Wooyeol Choi, and Professor Shin Seokjoo, for their critical comments and valuable thoughts on my work, which helped me to improve the final version of my thesis. I am highly grateful to them.

I am also grateful to the department of Computer Engineering. I have been offered a very good environment to continue my study and research. I am very happy to express my gratitude to the international office, graduate office, and especially Mr. Ji Hongbum for his cordial help throughout the semesters.

I am grateful to my family, friends and colleagues for their continuous support. Especially, I want to express my humble respect and love to my parents. I also want to thank my younger brother, Md. Mominul Islam, for his cordial help and support since I have started to live in abroad. I want to express my gratitude to Md. Amirul Alam Kanak, Amanie Al-Hussain, Alamgir Tapas, Md. Mamun Bhuiyan, Medha Ranjan Barman, Md. Ashraful Alam and Partha Mondal. I want to thank Md. Arafat Habib, Yeasir Araft, Masud An-Nur Islam Fahim, Md. Atikul Islam, and Izaz Ahmed for their cordial help. I am also grateful to other fellow Bangladeshi students in Chosun University.

Lastly, I am grateful to Almighty for all the hardships and beautiful moments given to me to test my patient.